Reinforcement-Based Decision Making in Corticostriatal Circuits: Mutual Constraints by Neurocomputational and Diffusion Models

Roger Ratcliff
ratcliff.22@osu.edu
Department of Psychology, The Ohio State University, Columbus,
OH 43210, U.S.A.

Michael J. Frank
Michael_Frank@Brown.edu
Department of Cognitive, Linguistic and Psychological Sciences,
Brown Institute for Brain Science, Brown University, Providence RI 02912, U.S.A.

In this letter, we examine the computational mechanisms of reinforcement-based decision making. We bridge the gap across multiple levels of analysis, from neural models of corticostriatal circuits—the basal ganglia (BG) model (Frank, 2005, 2006) to simpler but mathematically tractable diffusion models of two-choice decision making. Specifically, we generated simulated data from the BG model and fit the diffusion model (Ratcliff, 1978) to it. The standard diffusion model fits underestimated response times under conditions of high response and reinforcement conflict. Follow-up fits showed good fits to the data both by increasing nondecision time and by raising decision thresholds as a function of conflict and by allowing this threshold to collapse with time. This profile captures the role and dynamics of the subthalamic nucleus in BG circuitry, and as such, parametric modulations of projection strengths from this nucleus were associated with parametric increases in decision boundary and its modulation by conflict. We then present data from a human reinforcement learning experiment involving decisions with low- and high-reinforcement conflict. Again, the standard model failed to fit the data, but we found that two variants similar to those that fit the BG model data fit the experimental data, thereby providing a convergence of theoretical accounts of complex interactive decision-making mechanisms consistent with available data. This work also demonstrates how to make modest modifications to diffusion models to summarize core computations of the BG model. The result is a better fit and understanding of reinforcement-based choice data than that which would have occurred with either model alone.

1 Introduction

Common models of two-choice decision making assume that noisy evidence is accumulated and the decision is made when the process reaches one of two decision criteria (also referred to as boundaries or decision thresholds). The process either accumulates evidence in a single accumulator (the standard diffusion model) in which the process accumulates either positive evidence for one choice relative to the other choice or it accumulates evidence separately in two separate accumulators, one for each choice. Ratcliff and Smith (2004) reviewed the classes of evidence accumulation models and found that only those that assumed that evidence accumulation could be represented as diffusion processes (whether one accumulator or two accumulators) were successful in accounting for the qualitative patterns of results found in two-choice tasks (see Bogacz, Brown, Moehlis, Holmes, & Cohen, 2006; Ratcliff, 2006; Ratcliff & McKoon, 2008; Ratcliff, Thapar, Smith, & McKoon, 2005; Usher & McClelland, 2001).

These models have been widely applied to topics in perceptual processing and memory and used to examine the effects of a number of variables such as age, sleep deprivation, and aphasia. Other models in the class of diffusion models have been applied to decision making (Busemeyer & Townsend, 1993; Roe, Busemeyer, & Townsend, 2001), value-based decision making (Milosavljevic, Malmaud, Huth, Koch, & Rangel, 2010), and simple reaction time (Ratcliff & Van Dongen, 2011; Smith, 1995). Relationships have also been discovered between neurophysiological measures and decision-making models in humans (Donaldson, Wheeler, & Peterson, 2010; Forstmann et al., 2008; Philiastides, Ratcliff, & Sajda, 2006; Ratcliff, Philiastides, & Sajda, 2009; Wheeler, Petersen, Nelson, Ploran, & Velanova, 2008).

The models of behavioral decision making also provide good fits to data from animal experiments. Moreover, predictions from the models have shown a tight connection with the behavior of populations of single cells that appear to implement the processes of evidence accumulation to criterion (Gold & Shadlen, 2001, 2007; Mazurek, Roitman, Ditterich, & Shadlen, 2003; Ratcliff, Cherian, & Segraves, 2003; Ratcliff, Hasegawa, Hasegawa, Smith, & Segraves, 2007; Smith & Ratcliff, 2004). However, there is no formal model of the processes that control or regulate the decision making process.

The class of diffusion decision models has not yet been quantitatively fit to data from reinforcement learning paradigms. For these paradigms, there are biologically realistic models that acknowledge a key role of more complex circuits that link frontal cortex with basal ganglia (BG), which play a role in both the learning and decision-making processes. In this letter, we focus on one particular instantiation of BG models, that of Frank (2005,

2006), which we refer to in the remainder of the letter as "the BG model" solely for brevity. We discuss related BG models in the discussion. Our aim is to examine the relationship between diffusion decision models and BG models by first treating the BG model as a subject producing behavioral data and fitting its response proportions and response time (RT) distributions with the diffusion model, and, second, by fitting human data in a reinforcement learning paradigm that has been extensively used to test and refine the BG models. Our goal was to understand BG model mechanisms in terms of a diffusion process and, conversely, to refine diffusion models based on constraints from empirical and biological data in reinforcement tasks as informed by the BG model.

In contrast to simple cortical models of decision making, action selection in the BG model involves the entire cortico-basal ganglia circuit. In the BG models, candidate actions are generated by frontal cortical units in response to sensory input, and the BG gates the facilitation of the most adaptive actions while suppressing competing actions of lesser value. The gating process itself involves a series of inhibitory and disinhibitory connections between the striatum—the input segment of the basal ganglia involved in learning stimulus and action reward probabilities—and BG output nuclei, the thalamus, and back up to motor cortex (Frank, 2005, 2006; see also Gurney, Prescott, & Redgrave, 2001a, 2001b). These dynamics are additionally regulated by a mechanism for detecting cortical response or reinforcement conflict. When conflict is detected, the subthalamic nucleus (STN), a key node in BG circuitry, exhibits increased activation and temporarily delays the BG gating process, preventing impulsive choice (Frank, 2006; see below for a more detailed description). But how does this map into to constructs such as decision criteria?

Given the complexity and nonlinearity of these dynamics of neuronal activity in the BG models, unlike the diffusion models, it has not been possible to derive closed-formed solutions that can make precise predictions about the entire distribution of RTs for more reinforced versus less reinforced choices. Furthermore, while these models have been used to develop a theoretical framework for how neural circuitry gives rise to cognitive computations and how these computations might change as a function of biological manipulations (e.g., disease, pharmacology, genetics, and deep brain stimulation), they are not amenable to quantitative fits to the behavior of individual subjects. Thus, prior attempts to model individual subject choices in reinforcement learning tasks have employed algorithmic models summarizing some of the key computational features embedded in the neural circuits (Frank, Moustafa, Haughey, Curran, & Hutchison, 2007; Frank, Doll, Oas-Terpstra, & Moreno, 2009; Doll, Jacobs, Sanfey, & Frank, 2009; Cavanagh et al., 2010; Frank & Badre, 2011). While those models accounted for the learning process and genetic and neuroimaging predictors thereof, this study attempts to model the dynamics of the decision-making

process, including the full RT distributions for the different optimal and suboptimal choices.

The remainder of this letter is organized as follows. First, we elaborate the basic features of the neural model of corticostriatal circuitry in learning and decision making. Next, we use the diffusion model to quantitatively fit accuracy and RT distribution data generated by the BG model, with parametric modulations of biological parameters (dopamine levels and STN strength) to show how these alter diffusion parameters as a function of decision conflict. We then describe a probabilistic reinforcement learning experiment with healthy human subjects in which we collected data suitable for modeling with the diffusion model. We show that while the diffusion model provides quantitative fits to subject choices, it fails to account for the substantially slowed RTs observed during high-conflict choices. Based on theoretical considerations of the BG model and related data, we modified the diffusion model to summarize the function of conflict-based decision making and show that this modification is sufficient to improve model fits to the data. We finally argue for the utility of this approach in which the diffusion model can be used as a meeting point between the BG model and data from the reinforcement learning paradigm.

2 Basal Ganglia Model

The BG model is primarily intended to address the learning and decision-making functions of corticostriatal circuitry. It has been applied in various instances to simulate changes in decision making resulting from manipulations of this BG system (for review, Doll & Frank, 2009). The model is also constrained by data at the lower neural level of analysis: for example, Frank (2006) simulated patterns of neural data in healthy individuals (Magill, Sharott, Bevan, Brown, & Bolam, 2004) and pathological states associated with Parkinson's disease (Levy, Hutchison, Lozano, & Dostrovsky, 2000). Furthermore, by virtue of interactions with different areas of frontal cortex, the BG can participate in a wide range of cognitive functions at different levels of abstraction. Specifically, although we focus on a single BG circuit in decision making, such networks can be cascaded such that one loop makes higher order cognitive decisions and provides contextual input to the lower-level motor circuit for response selection (Frank & Badre, 2011).

Figure 1 shows the basic circuitry in the canonical BG model representing a single frontal basal ganglia circuit (here, in the preSMA motor loop). (A more complete description of the implemented model is presented in the appendix.) Like the cortical models of decision making, the BG models assume that a response is made once one of a population of cortical response
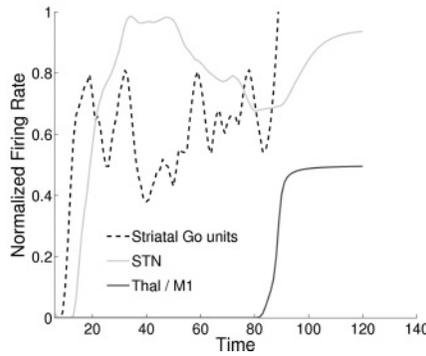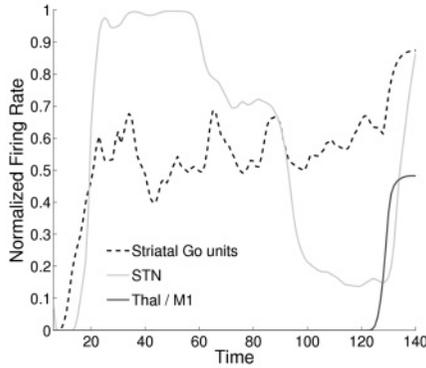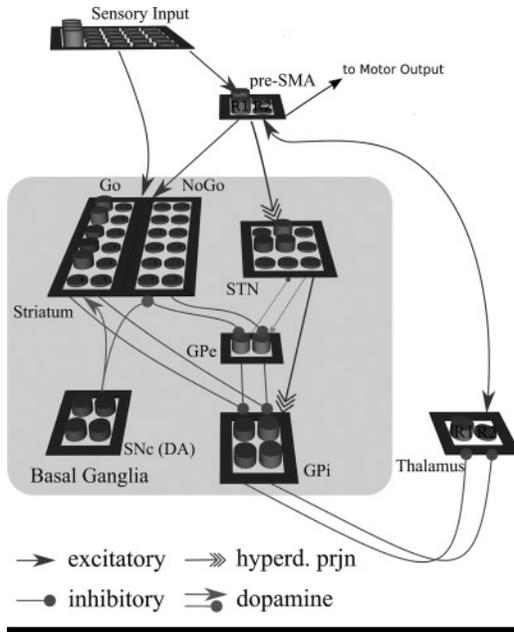
units exceeds some threshold activation. There is also lateral inhibition between the cortical response units, which are leaky and noisy. Although the cortical response units can integrate evidence from the stimulus input, their activities are dynamically regulated (gated) by BG circuitry. Specifically, in addition to receiving excitatory input from (high-level) sensory representations, cortical response units also receive strong excitatory input from the motor thalamus. However, under baseline conditions, thalamus units are all inhibited by the output of the BG, the internal segment of the globus pallidus (GPi), in which neurons are tonically active and send inhibitory projections to the thalamus. Thus, under these baseline conditions, the BG

Figure 1: (Top) The BG model. The striatum receives sensory stimulus input and candidate motor actions from preSMA. Go neurons in the direct pathway to GPi act to facilitate the associated response by inhibiting the corresponding column of GPi, disinhibiting the thalamus. The resulting boost of thalamocortical activity is sufficient to preferentially drive a given set of preSMA response units, which then inhibit their competitors via lateral inhibition (not shown). NoGo neurons in the indirect pathway compete with Go activity by sending their outputs to the GPe, which then inhibits GPi, preventing the BG from selecting a response. The STN receives input along the hyperdirect pathway from preSMA and exerts a temporary global NoGo function by sending diffuse excitatory projections to the GPi, preventing Go signals from inhibiting the GPi and therefore from disinhibiting the thalamus. This is particularly evident when multiple responses are coactivated in preSMA (in the snapshot, there is low conflict and low STN activity). Dopamine neurons in the SNc modulate striatal activity, exciting Go neurons via simulated D1 receptors while inhibiting NoGo neurons via D2 receptors. preSMA: presupplementary motor area; GPi: globus pallidus internal segment; GPe globus pallidus external segment; SNc: substantia nigra pars compacta; STN: subthalamic nucleus. (Bottom) Normalized mean firing rate of STN and all Go units coding for the response that is ultimately selected, and corresponding thalamus units, in two single trials. Thal. is activity in the thalamic units, and activity in these is tracked by motor units M1. Note initially that thalamic units are inhibited, and preSMA activity is noisy (not shown). Striatal Go unit firing increases with time, modulated by noisy activity in preSMA for a corresponding response, thus leading to on "accumulation" of value. STN activity surges early in the trial and then declines, after which the Go units can disinhibit the thalamus and a response is executed. In the top panel, STN activity declines substantially before this can occur, whereas in the bottom panel, there is sufficiently high Go activity to elicit a response earlier in time. Thus, STN activity prevents early premature responding by inducing a delay (initial rise) and thereafter acts as a dynamic collapsing decision threshold. (The figure shown here is a grayscale version of a color figure presented in the online supplement, available at http://www.mitpressjournals.org/doi/suppl/10.1162/NECO_a_00270.)

model essentially reduces to a standard cortical model of decision making with leak and lateral inhibition (Usher & McClelland, 2001). Indeed, if there is sufficient evidence for one of the alternative responses (due to strong weights from the sensory input to one of the motor response units), then the decision process is similar to that described in many models, without requiring BG gating (see Frank & Claus, 2006; Frank, Scheres, & Sherman, 2007; Ashby, Ennis, & Spiering, 2007; see also the appendix). In contrast, with BG gating, the thalamic units for one of the responses become active, providing selective bottom-up input to one of the cortical response units such that this unit has a tremendous advantage over those coding for alternative responses, which are quickly suppressed due to lateral inhibition. This amounts to a dynamic nonlinear gating process, as described in more detail below.

In reinforcement learning tasks, the rules for determining which response is correct are not known in advance, and there is no overt information in the stimulus that conveys which response to emit. Instead, these associations have to be learned over trials by reinforcement. An extensive literature shows that the BG are critically involved in both the learning of reinforcement probabilities and the selection of actions based on reinforcement history (see Doll & Frank, 2009, and Wiecki & Frank, 2010, for reviews and detailed biological evidence). To appreciate how this process operates in the model, it is critical to consider how different pathways from the cortex to the BG can affect the GPi, which controls inhibition of the thalamus. As we shall see, there are three main pathways from the cortex though the BG to the GPi, termed the direct (Go), indirect (NoGo), and hyperdirect (Global NoGo) pathways.

Neurons in the striatum (the main input nucleus of the BG) receive inputs from both the sensory cortex (representing the current sensory state) and presupplementary motor area (preSMA, representing the candidate responses). Two main striatal cell populations ultimately control response selection. The Go population facilitates the selection of a particular response by inhibiting the corresponding units in GPi, thereby releasing the inhibition onto thalamus. This process is termed disinhibition (Chevalier & Deniau, 1990) and has a very strong influence on the decision process in cortical (preSMA) units. Indeed, thalamic disinhibition will allow active preSMA units to excite the corresponding thalamic unit (due to top-down excitatory projections from the cortex to the thalamus), which then reciprocally amplify the preSMA activity. At this point, activation in the cortical response unit ballistically accelerates as it inhibits its competitor, swiftly reaching motor threshold (see Figure 1b).

Counteracting the Go population is the NoGo population, which prevents the facilitation of responses. The main difference between Go and NoGo populations is that the NoGo cells exert their effects on GPi indirectly, sending their focused inhibitory projections to corresponding units in the external segment of the globus pallidus (GPe), which then inhibits

GPi.[1] The NoGo cells prevent specific thalamic columns of units in the thalamus from being disinhibited due to the additional inhibitory route between the NoGo cells and the GPi. Thus, separate Go and NoGo populations of units can selectively facilitate and suppress specific responses.

Dopamine (DA) units in the substantia nigra pars compacta (SNc) modulate the overall balance of activity, boosting corticostriatal Go activity while inhibiting corticostriatal NoGo activity. This occurs via separate excitatory and inhibitory projections from SNc dopamine units to Go and NoGo populations, simulating the differential effects of D1 and D2 dopamine receptors in these two populations (see Frank, 2005). (Note that dopamine does not activate an entire column of Go units, but instead acts as an excitatory current in these units that amplifies activity in those units receiving strong excitatory input from cortex). Thus, higher levels of simulated dopamine lead to relatively greater overall propensity for Go than NoGo activity and faster responding. Importantly, dopamine also modulates learning in the Go and NoGo populations, with dopamine bursts during rewarding outcomes promoting increased synaptic plasticity in active Go units. Conversely, dopamine dips during negative outcomes promote plasticity in NoGo units. In this way, the striatum learns the probability that a given response will be rewarded and that it will be punished. The relative balance of these quantities influences both the likelihood that this response will be executed and the speed with which it is executed (substantially more Go than NoGo activity will result in faster disinhibition of the thalamus and, hence, faster responses). These two factors—dopamine effects on activity and plasticity—allow the model to account for the effects of various dopamine manipulations on overall response speed, reinforcement learning, and the impact of reinforcement learning on response speed (Wiecki & Frank, 2010).

Finally, in addition to these Go and NoGo pathways, there is a third pathway from the cortex to BG output involving the subthalamic nucleus (STN). The STN receives input from cortical response units and sends diffuse excitatory projections directly to all GPi units. STN activity implements

---

[1]The classic indirect pathway refers to a more circuitous route from the striatum to GPe to STN to GPi, whereas in our models, we focus on the less indirect striatum-GPe-GPi route. It was not originally known that the GPe projects directly to the GPi, and, moreover, whereas GPe-GPi projections are focused, STN-GPi projects are diffuse. This means that the indirect pathway in our model can inhibit specific motor actions associated with the column, whereas the route through STN tends to globally inhibit all actions. Furthermore, the third hyperdirect pathway (cortex-STN-GPi) was not known when the classical model was developed in the 1980s; hence, the STN was considered part of the indirect pathway. We thus now consider the STN pathway to be functionally distinct, though the GPe-STN route still provides feedback inhibition to STN. Functionally, this feedback also yields greater conflict-related STN activity when the two decision options both have negative value: the greater NoGo activity in this case results in inhibition of GPe and thus disinhibition of STN.

a Global NoGo signal that excites the GPi and in turn inhibits the thalamus, thereby making it more difficult for striatal Go activity to facilitate any response. Thus, in contrast to the NoGo units, which act to suppress specific responses based on their negative values, the STN Global NoGo mechanism suppresses all responses for a period of time. Frank (2006) focused specifically on simulating STN contributions in regulating the dynamics of the decision process. During the initial response selection process, surges in STN activity are seen (in both the model and electrophysiological recordings; Magill et al., 2004), and more so under conditions of response conflict (Isoda & Hikosaka, 2008; Cavanagh et al., 2011). In the model, this temporary STN surge serves to prevent premature or impulsive responding. Coactivation of multiple cortical response units (an index of response conflict; cf. Botvinick, Braver, Barch, Carter, & Cohen, 2001) results in a larger STN surge and thereby makes it more difficult to facilitate a response early during the decision process. With time, however, this STN surge subsides (due to feedback inhibition from GPe and neural accommodation), at which point it is easier to select a response. Simulations showed that this delay in the decision process is adaptive when alternative responses have subtly different reinforcement values, as it enables the system to integrate noisy activity over longer periods to determine the best response (Frank, 2006). Moreover, this function is consistent with experimental data showing that disruption of STN function results in impulsive premature responding particularly under conditions of decision conflict (Frank, Samanta, Moustafa, & Sherman, 2007; Wylie et al., 2010; Cavanagh et al., 2011).

Given all of these dynamics, variability in RTs is influenced by the following factors. First, noise in the cortical response units translates into noise in the BG selection process. Activity in striatal Go units is proportional to the output activity of their corresponding cortical response units. Because cortical activity is noisy, striatal Go units integrate the value of the coded response across bouts of increasing and decreasing cortical activity. Figure 1b shows two examples of how striatal Go unit activity for the response that is ultimately selected shows an increasing trend throughout the trial, appearing to accumulate value evidence preferentially for the activated cortical response. (This process is analogous to recently reported data with the diffusion model in which good fits to reward-based decision making were obtained by assuming that the drift process is biased toward the decision option to which the subject is currently fixating; Krajbich, Armel, & Rangel, 2010). This accumulating striatal Go activity also resembles monkey striatal electrophysiological data in perceptual decision-making tasks (Ding & Gold, 2010). As noted earlier, the timing of BG selection is further influenced by relative Go to NoGo activity levels, which are themselves modulated by dopamine and past learning (i.e., if the activated cortical response is not adaptive, the NoGo cells will be active and it will be suppressed). Finally, early surges in STN activity, which are larger when multiple cortical response units are coactive, delay responding. Responses are disinhibited

only after STN activity declines; nevertheless, with sufficient Go activity, a response can be gated at an earlier point in this process, before STN activity has completely subsided (see Figure 1b, bottom). We will investigate how this STN modulation of response selection can be captured by a collapsing decision bound, together with a fixed delay in the onset of evidence accumulation.

Thus far, the BG model simulations and experiments have primarily focused on accuracy and learning. As we have alluded, in some applications, model predictions have been generated as to the effect of behavioral and biological manipulations on mean RT in decision making (Frank, Samanta et al., 2007; Frank, Scheres et al., 2007, Moustafa, Cohen, Sherman, & Frank, 2008). However, the model has not generated quantitative predictions for individual subjects and has not provided detailed predictions about RT distributions or relationship with accuracy (including speed-accuracy trade-offs). To begin to address this issue, we generated responses from the BG model and fit them with the diffusion model. (See Frank, 2006, and the appendix for detailed model equations and parameters.)

3 Generating Predictions from the BG Model

The BG model supports decision-making processes by simulating nonlinear dynamics among neural populations within multiple BG nodes, with links to neurophysiological and pharmacological data. While the model makes qualitative predictions about behavioral data resulting from biological manipulations, it might be disingenuous to attempt to quantitatively fit individual subject behavioral data by optimizing neural network model parameters. The model does not attempt to simulate the entire perceptual and motor output processes, and parameter fits would invariably require modifying available model parameters to account for these phenomena, but these would clearly be the wrong parameters. Moreover, there are simply too many degrees of freedom in neural models to allow free parameter search when only observing behavioral output (accuracy and RT distributions). Nevertheless, we believe it is of utmost importance to be able to link different levels of modeling analysis. If a model becomes more and more complex, it becomes difficult to understand in terms of its core functional principles. We thus generated behavioral data from the BG model, systematically varying a few of its relevant parameters, to observe the resulting effects on diffusion model parameters.

The core model parameters used here remain identical to those described previously (Frank, 2006, Moustafa et al., 2008). We generated a large number (5000) of simulated responses, focusing here on decisions based on a fixed set of synaptic weights to approximate decision making for a given learned association, without the added complexity of estimating how diffusion decision variables change during learning. (RTs in the model change

depending on stage of learning, the degree of positive versus negative associations, and so on; Moustafa et al., 2008; Wiecki, Riedinger, Meyerhofer, Schmidt, & Frank, 2009.) This is similar to fitting decisions made by subjects in the testing phase of the experiment described below (and not the learning phase).

To simulate varying degrees of conflict and positive and negative associations, we directly manipulated the following variables in order to precisely control them. We simulated two input stimuli, represented by four units in the sensory input layer. These units projected (with full connectivity) to both the striatum and preSMA. This allows, due to the random distribution of corticostriatal synaptic weights, differential striatal populations to encode the conjunction between the sensory state and the candidate action, which is important in learning environments. To code the correctness of the response given the stimulus, we added a unit projecting to one of the preSMA columns with a weight of 1.0. To simulate response conflict, we manipulated the weights from this input unit to the incorrect motor units. For low conflict, this weight was set to 0.7, whereas for high conflict, it was set to 0.9. (Other values can be used, but we found that these produced accuracy rates corresponding roughly to those observed in the experiment.) Gaussian noise is added to the membrane potential of preSMA response units, so that the firing rates of these units are not deterministic from one trial to the next and from one processing cycle to the next.

Because of the projections from cortical response units to the STN, the high-conflict condition (with greater overall cortical response unit activity) results in a larger and prolonged STN surge, and hence slowed responding (Frank, 2006). The degree of slowing is greater than would result from just lateral inhibition between the cortical response units as in standard accumulator models. We thus assessed the role of parametric changes in STN function on diffusion model parameters and their sensitivity to conflict. To do so, we varied the weight scale of the projections from the STN to the GPi, which acts as a constant multiplier on individual synaptic projection weights. The default was 0.55 (STN mid), and we varied this to reduce STN impact (0.4, STN low) and increase STN impact (0.7, STN high).

We also simulated the impact of changing dopamine (DA) levels on conflict-based decision making. In the model, dopamine units in the SNc project to Go and NoGo units along different projections that are excitatory and inhibitory (due to the differential expression of D1 and D2 receptors in the two striatal populations; see Frank, 2005). Thus, an increase in dopamine will favor Go unit activity over NoGo unit activity, and vice versa for a decrease in dopamine. These changes in dopamine allow us to simulate both chronic biological changes (e.g., due to Parkinson's disease or medications) and the values of the candidate choices. Thus, when dopamine levels are high, they emphasize Go unit activity, so if both candidate actions are strongly activated in cortex (i.e., conflict), the corresponding two

populations of Go units will be more active in a manner equivalent to that which occurs in a win-win decision between two responses that had both received a high probability of positive reinforcement (as in the empirical experiment reported by Frank, Samanta et al., 2007, and below). Conversely, low DA levels potentiate NoGo activity, and if cortical conflict is high, this is equivalent to simulating a high-conflict lose-lose decision. In our simulations, we therefore consider conditions in which DA levels are high (normalized SNc dopamine unit activity ~0.7) and low (normalized SNc unit activity < 0.5). For low-conflict win-lose conditions, these changes in dopamine activity will simply affect the relative excitability of the Go units coding for the more activated (winning) response.

Response times are measured as described previously (Frank, Scheres, et al., 2007; Moustafa et al., 2008; Wiecki et al., 2009). Specifically, time in the neural model is measured in terms of processing cycles in which membrane potentials and neural firing rates are updated (as a function of current inputs and subject to time constants limiting the rate with which activations can change). Because we are interested in examining the BG contributions to decision making, we measure RTs in terms of the number of processing cycles until a response is gated by the BG, that is, until a given thalamic unit exceeds a threshold level of normalized firing rate (arbitrarily set to 0.8) and one of the output motor units is at least 50% active. (When the thalamus is excited, preSMA units reach maximal activity almost immediately.) Using the thalamus ensures that we are always examining RT distributions influenced by the BG circuitry, and not responses that could in principle be generated by direct sensory-motor transformations. (In practice, this rarely occurs in the model unless the weights from sensory cortex to preSMA units sufficiently favor one response over the other, as would be the case in simple perceptual decisions but not reinforcement-based ones).

We first generated simulated data from the BG model and manipulated three key factors: the degree of response conflict, the levels of dopamine (emphasizing Go or NoGo activity levels), and the strength of STN-GPi weight projections (affecting the degree to which STN activity contributes to preventing responses from being gated prematurely). Next, we used the diffusion model to fit the simulated data as if they had been generated by a human subject for all combinations of the above factors. We linearly transformed RTs in network processing cycles to a realistic range in seconds by multiplying them by 10. This procedure allows us to estimate the diffusion model parameters that best correspond to the BG model's predictions and how they vary as a function of experimental condition: high versus low conflict, high versus low dopamine (simulating effects of value, e.g., win-win versus lose-lose), and three levels of STN-GPi projection strengths (low, mid, and high). This theoretical exercise is needed if one were to derive predictions from the BG model resulting from biological manipulation (e.g., dopamine medications, STN deep brain stimulation, or individual
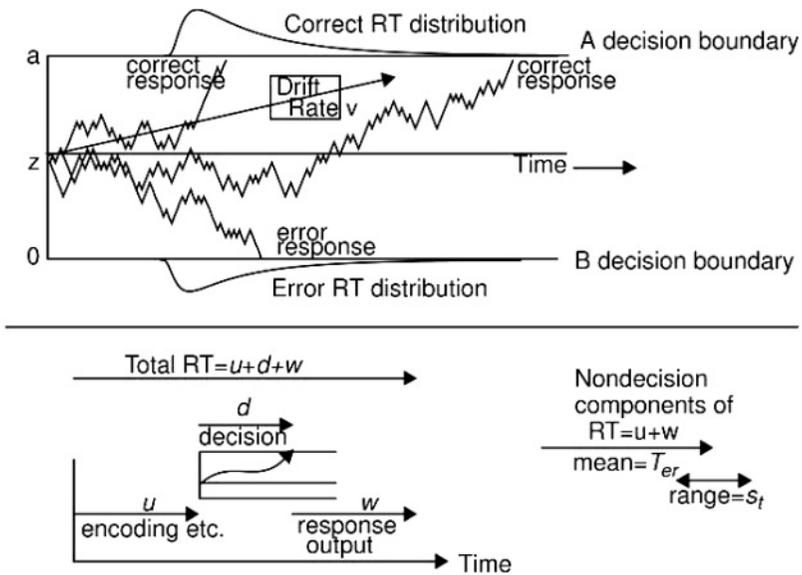
Figure 2: The diffusion model. (Top) Three simulated paths with drift rate v, boundary separation a, and starting point z. (Bottom) Encoding time (u), decision time (d), and response output (w) time. The nondecision component is the sum of u and w with mean $= T_{er}$ and with variability represented by a uniform distribution with range s_t.

differences in tract strengths) and to then test the model predictions as a function of these manipulations with a human experiment. For example, we hypothesize that transiently increased STN activity associated with decision conflict induces a change in diffusion model parameters, associated with an increase in decision threshold or delaying the onset of evidence accumulation. Thus, we predicted that conflict would induce a change in estimated diffusion parameters and that this conflict effect would be parametrically scaled by the strength of STN projections. We tested these and other predictions by varying the relevant BG model parameters and estimating their resultant effects by fitting the diffusion model to the outputs of the biological model.

4 The Diffusion Model

The diffusion model (Ratcliff, 1978) is a model of the cognitive processes involved in making simple two-choice decisions (see Figure 2). It separates the quality of evidence entering a decision from the decision criteria and from other, nondecision processes such as stimulus encoding and response

execution. The model applies only to relatively fast two-choice decisions (mean RTs typically less than about 1000 to 1500 ms) and only to decisions that are a single-stage decision process (as opposed to the multiple-stage processes that might be involved in, for example, reasoning tasks or card sorting tasks).

Decisions are made by a noisy process that accumulates information over time from a starting point z toward one of two response criteria, or boundaries: a and 0. When a boundary is reached, a response is initiated. The rate of accumulation of information is called the drift rate (v), and it is determined by the quality of the information extracted from the stimulus in perceptual tasks and the quality of match between the test item and memory in memory and lexical decision tasks. The nondecision components of processing such as encoding and response execution are combined into one component with mean T_{er}. Within-trial variability (noise) in the accumulation of information from the starting point toward the boundaries results in processes with the same mean drift rate terminating at different times (producing RT distributions) and sometimes at the wrong boundary (producing errors and associated RT distributions). It is assumed that components of processing vary from trial to trial. Across-trial variability in drift rate (normally distributed with SD η) and starting point (uniformly distributed with range s_z), in conjunction with boundary positions and drift rates, determines the relative speed of correct versus error responses. It is also assumed that the nondecision component varies across trials, uniformly distributed with range s_t (the precise form of this distribution is not critical because the nondecision time variability is much less than the decision time variability; thus, in the convolution of the two, the decision time distribution dominates). (For further details of the model, see Ratcliff & McKoon, 2008, for a review; Ratcliff & Smith, 2004, for comparisons among the different sequential sampling models; and Ratcliff & Tuerlinckx, 2002, for how to fit the model to data.)

Fits of the diffusion model to a large number of experiments have produced one-to-one mappings between experimental manipulations and model parameters. Changes in the quality of evidence entering the decision process are modeled by changes in drift rate. Speed-accuracy trade-offs are modeled by changing the distance between the boundaries of the decision process: wider boundaries require more information before a decision can be made, leading to more accurate and slower responses. Both of these manipulations, quality of evidence and speed-accuracy instructions, produce changes in both accuracy and RTs, including changes in the spreads and locations of the RT distributions across conditions, for both correct and error responses. Changes in all aspects of the data (accuracy and the spread and location of RT distributions for correct and error responses) are handled by changes only in drift rates when the difficulty of the stimulus information is manipulated and only in distance between the boundaries if speed and accuracy instructions are manipulated. In other words, a change in one

parameter accounts for changes entailing many degrees of freedom in the data.

Ratcliff and Tuerlinckx (2002) carried out a number of simulations that showed that the model is identifiable. Simulated data were generated, and the model was then fit to the simulated data. With large numbers of observations, the original parameters were recovered with little bias. But with smaller numbers of observations, there were some biases. This showed that the model is identifiable in that a change in boundary separation cannot be mimicked by a change in drift rate, for example. Ratcliff (2002) also showed that the model is quite constrained. He generated simulated data from a number of different assumptions and showed that the model could not fit those patterns. The main aspect of the simulated data that the model could not fit was the behavior of RT distributions. They had to be of just the right shape, and they had to change in just the right ways across conditions. Thus, we asked whether changes in conflict and BG model parameters would lead to identifiable changes in diffusion model parameters.

Figure 3 (bottom) presents a heat map of simulated processes in the diffusion model for parameter values that correspond to the low-conflict condition in fits to the simulated BG data (for low STN and high dopamine/Go conditions). As Ratcliff (1988, 2006) described, the evolution of paths is to move the average toward the boundary to which the process is drifting. This rapidly produces an almost stationary distribution, and then, as processes exit the diffusion process, this distribution gradually collapses. An example of this distribution is shown to the right of the heat map. The peak of this distribution can be seen to correspond to the peak heat from time 0.2 and upward, which illustrates the stationary distribution of processes described. We plotted a similar heat map for the BG model, this time plotting the summed and normalized difference between populations of striatal Go unit activities coding for the two responses. The overall dynamic is similar, but with the distribution collapsing more swiftly as responses exceed threshold.

5 Fitting the BG Model Accuracy and RT Distributions with the Diffusion Model

Simulated data from the BG model were generated with parameter values that qualitatively match the experimental reinforcement conflict design in Frank, Samanta et al. (2007) and Cavanagh et al. (2011). The simulated data had three factors manipulated. We parametrically varied the strength of the projections from STN to GPi (STN strength). This manipulation allows us to investigate to what degree STN influence on BG output affects decision parameters and can also approximate the effects of STN manipulations (e.g., by deep brain stimulation) or individual differences in engagement of the hyperdirect pathway (e.g., due to differential tract strengths or STN excitability to preSMA inputs). We therefore estimated different values of boundary
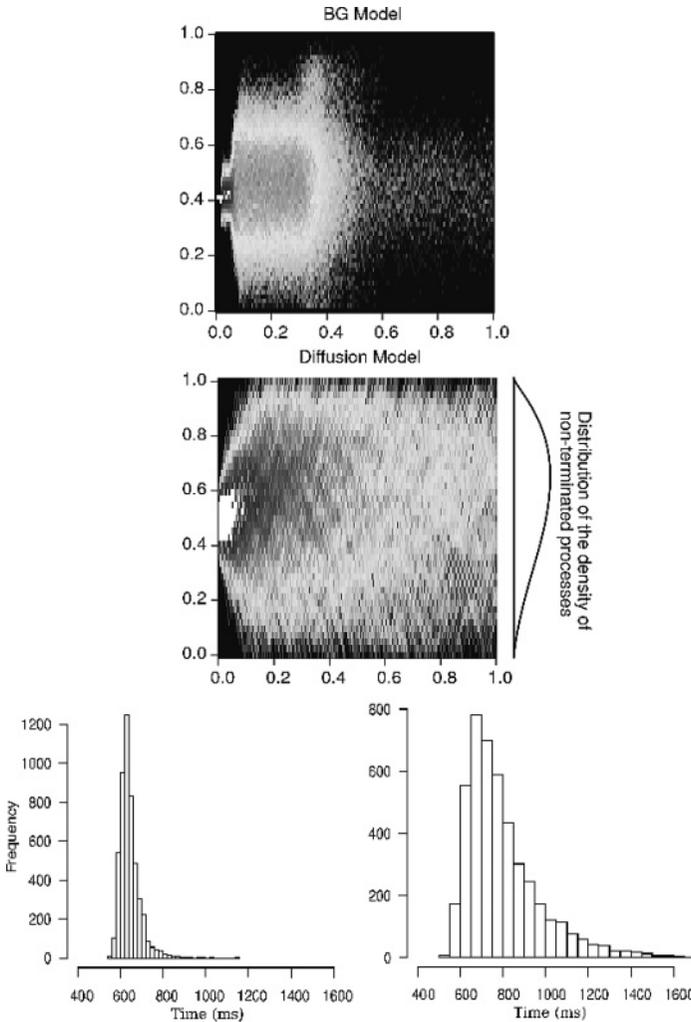
Figure 3: Heat maps of simulated paths for the diffusion model and the BG model (difference between summed and normalized firing rates for striatal Go units coding for the two responses). Red corresponds to a high path density, blue to a low path density. For the diffusion model, the distribution to the right corresponds to the asymptotic distribution of path positions after about 0.2 units on the *x*-axis. (Bottom). Example RT histograms generated from the BG model for low- and high-conflict conditions. Distributions are skewed and show approximately the same shape as human RT data and of the diffusion model, and hence the corresponding diffusion model parameters that produce these distributions can be estimated. (The figure shown here is a grayscale version of a color figure presented in the online supplement, available at http://www.mitpressjournals.org/doi/suppl/10.1162/NECO_a_00270.)

separation and nondecision time for each level of STN strength. The other factors, degree of response conflict and levels of dopamine (Go/NoGo activity), were manipulated within each of these levels of STN strengths (e.g., to simulate different task conditions involving levels of conflict and reward value).

The resulting simulated data are summarized in Table 1, showing response proportions for the more rewarded choice and the 0.1, 0.5, and 0.9 RT quantiles for the more and less rewarded choices. The RT distributions from the BG model are right-skewed and look typical of RT data. The right skew can be seen from the quantiles because the distance between the 0.9 and 0.5 quantiles is greater than the distance between the 0.1 and 0.5 quantiles. The effect of increasing conflict was to reduce accuracy by 15% and disperse the RT distribution, delaying the 0.1 quantile RT by between 0 and 20 ms and the 0.9 quantile by between 40 and 120 ms. Increased STN strength led to more skewed RT distributions, delaying the 0.1 quantile by 60 to 80 ms and the 0.9 quantile by 150 to 200 ms, and also modestly improved accuracy (by about 1–2%). In general, responses for less rewarded choices are a little slower than those for more rewarded choices.

These findings are largely consistent with those of Frank (2006), showing that removing the STN altogether from the model led to premature responding and impaired accuracy in high-conflict conditions. They are also consistent with electrophysiological data showing increased STN activity during response conflict and accordingly increased RTs (Isoda & Hikosaka, 2008; Cavanagh et al., 2011). In contrast to STN effects, low dopamine (more NoGo activity) relative to high dopamine (more Go activity) did not affect accuracy but selectively delayed the leading edge of the RT distribution and not the tail (i.e., the 0.1 quantile RT was slowed by 20 to 40 ms, for low to high STN strength, but subsequent quantiles were delayed less, with the 0.9 quantile RT completely unchanged).

It is important to note here that some of the effects, such as a change in only the leading edge of the RT distribution, are not normally seen with standard decision-making manipulations, such as speed-accuracy, or difficulty. However, these effects are observed with biological manipulations such as deep brain stimulation, dopamine manipulations, and conflict.

We fit these simulated data with two versions of the following assumptions. To simulate different levels of stimulus quality, we allowed drift rates to vary as a function of conflict. However, we also asked whether, over and above changes in drift, conflict is associated with an increase in decision bound, due to STN contributions, as posited informally (Frank, 2006). Recall that this model suggests that during high-conflict decisions, a transient increase in STN activity induces a global NoGo process that temporarily prevents responding or makes it more difficult to facilitate any response. That is, if the STN is sufficiently active so that no amount of striatal Go activity will gate a response, this would be captured by a delay in the onset of decision process. In contrast, if the STN surge simply makes it more

Table 1: Simulated Data from the BG Model.

Factors Manipulated			More Reinforced Choice				Less Reinforced Choice		
STN	Conflict	Go/NoGo	Proportion	0.1 Quantile RT	0.5 Quantile RT	0.9 Quantile RT	0.1 Quantile RT	0.5 Quantile RT	0.9 Quantile RT
Low	High	Go	0.622	393.0	484.0	601.0	410.0	500.0	612.0
		NoGo	0.623	434.0	494.0	601.0	441.0	504.0	612.0
	Low	Go	0.768	391.0	473.0	563.0	407.0	492.0	593.0
		NoGo	0.768	431.0	483.0	562.0	443.0	500.0	581.0
Medium	High	Go	0.630	430.0	523.0	662.0	443.0	532.0	680.5
		NoGo	0.631	454.0	531.0	664.0	464.0	540.0	682.0
	Low	Go	0.777	421.0	503.0	612.0	442.0	520.0	633.0
		NoGo	0.773	453.0	514.0	613.0	462.0	524.0	622.5
High	High	Go	0.631	471.0	580.0	814.0	474.0	590.0	823.0
		NoGo	0.633	490.0	590.0	814.0	494.0	592.0	821.0
	Low	Go	0.785	454.0	542.0	694.0	472.0	561.0	704.0
		NoGo	0.785	480.0	551.0	692.0	492.5	563.5	703.5

difficult to gate a response but which could still occur with sufficient striatal Go activity, this would correspond to an increased decision threshold (boundary). Finally, a combination of both effects is possible: there might be a period during which STN effects are so strong as to delay all responses, followed by a collapsing bound reflecting the decline in STN activity. We thus allowed diffusion model parameters representing boundary settings and nondecision time to differ as a function of STN strength. Finally, we allowed dopamine levels to also influence the nondecision time, given the known role of dopamine in facilitating motor response execution. In sum, this combination of parameter fits allows us to test whether dopamine affects decision- and nondecision-related processes, whether decision conflict can be captured solely by changes in drift rate, or whether a change in boundary is also needed, and whether STN strength modulates all of the above effects.

We tested the above assumptions in two sets of fits. In the first set (the static model), the boundary setting and nondecision time variables remain static throughout the trial but can change across conditions. In the second set of fits (the dynamic model), we assume that the decision bounds collapse exponentially with time (Frazier & Yu, 2008; Ditterich, 2006; Churchland, Kiani, & Shadlen, 2008; Viviani, 1979) from a single value to asymptotic values as in the static model. This choice was motivated by the notion that STN activity, posited to affect the decision bound as a function of conflict, shows an initial surge during response selection but then subsides with time to allow a response to be gated. We assumed an exponential decay from initial level to asymptote with a time constant of decay that was estimated to be 250 ms.

This collapsing-bound assumption is difficult to implement in the diffusion model because there are no exact solutions for the RT distribution with collapsing boundaries. We therefore used a random walk approximation to the diffusion process (Tuerlinckx, Maris, Ratcliff, & De Boeck, 2001) to generate accuracy values and quantile RTs for 2000 simulated trials per experimental condition. The step size in the simulation was 1 ms. The diffusion model was fit to the simulated data (and experimental data later) by minimizing a chi square statistic in a two-step process: using a Markov chain Monte Carlo method to obtain parameter values near the best fit and then using a standard SIMPLEX minimization routine to find the best-fitting parameter values. The SIMPLEX routine was restarted seven times using the parameter values from the prior fit. The data entered into the minimization routine for each experimental condition were the 0.1, 0.3, 0.5, 0.7, 0.9 quantile RTs for correct and error responses (more reinforced choices and less reinforced choices) and the corresponding response proportions. The quantile RTs and the diffusion model were used to generate the predicted cumulative probability of a response occurring by that quantile RT. Subtracting the cumulative probabilities for each successive quantile from the next higher quantile gives the proportion of responses between adjacent

quantiles. For the chi square computation, these are the expected proportions, to be compared to the observed proportions of responses between the quantiles (i.e., the proportions between 0, 0.1, 0.3, 0.5, 0.7, 0.9, and 1.0, which are 0.1, 0.2, 0.2, 0.2, 0.2, and 0.1). The proportions are multiplied by the number of observations in the condition to give observed (O) and expected (E) frequencies and summing over $(O - E)^2/E$ for all conditions gives a single chi square value to be minimized (see Ratcliff & Tuerlinckx, 2002, for a full description of the method).

We minimized the chi square measure described above to fit the static and dynamic models to the simulated data (see Ratcliff & Tuerlinckx, 2002). Because the units are in ten's of milliseconds (i.e., very granular) and because the statistical properties of the model generating the simulated data are not known, we cannot assign significance levels to the fits. However, we can use the values to assess relative goodness of fit.

We also compared diffusion model predictions for the standard constant boundary model from this simulation method with exact predictions using the diffusion model equations (Ratcliff & Tuerlinckx, 2002), and they produced almost exact matches. For all conditions in the dynamic model, a common decay time constant was used, and a common initial level of the boundary separation was used (after preliminary fits showed similar values when these were allowed to vary with STN level).

Figure 4 shows the fits of the model to the response proportion and the 0.1, 0.5, and 0.9 quantile RTs. The static model fits the RT quantiles a little better than the dynamic model, but the dynamic model fits the accuracy values better than the static model. Because these are simulated with a very large number of observations (about 5000 per condition), the variabilities in the quantiles and accuracy values are much lower than would be seen in human data. Thus, although there are small consistent misses, the overall fits are certainly as good as is obtained when fitting human data (cf. Ratcliff, Thapar, & McKoon, 2010 and the experiment presented below).

The chi square for the model with different static values for nondecision component and boundary separation was 2031, and for the model with collapsing boundaries, it was 1806. We also fit a (standard) model with one value of the nondecision time and constant boundaries, and the chi square value was 6592. Thus, the numerical fit of the simple standard model is three times worse than the numerical fits of the other two models, and the static and dynamic fit about as well as each other, with a modest (but not necessarily meaningful) advantage for the dynamic model. Our results support the notion that the effects of dopamine and conflict on RT distributions of the BG model can be captured by a diffusion model in which conflict either transiently or statically increases the decision threshold as a function of STN strength, and dopamine speeds response execution. Although the fits do not clearly distinguish between static and dynamic bounds, it is notable that the collapsing bounds model more closely corresponds to the known underlying profile of STN activity in the BG model, posited to affect
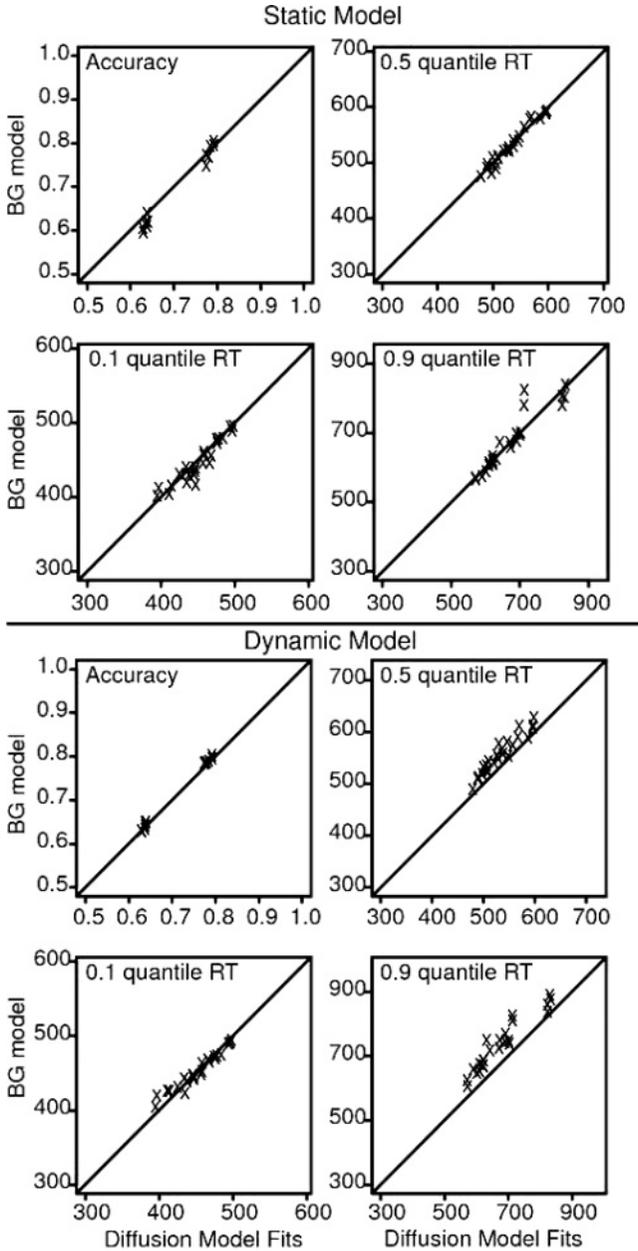
Figure 4: Plots of response proportion (accuracy) and the 0.1, 0.5, and 0.9 quantile RTs from the BG model simulations and the fitted diffusion model predictions.
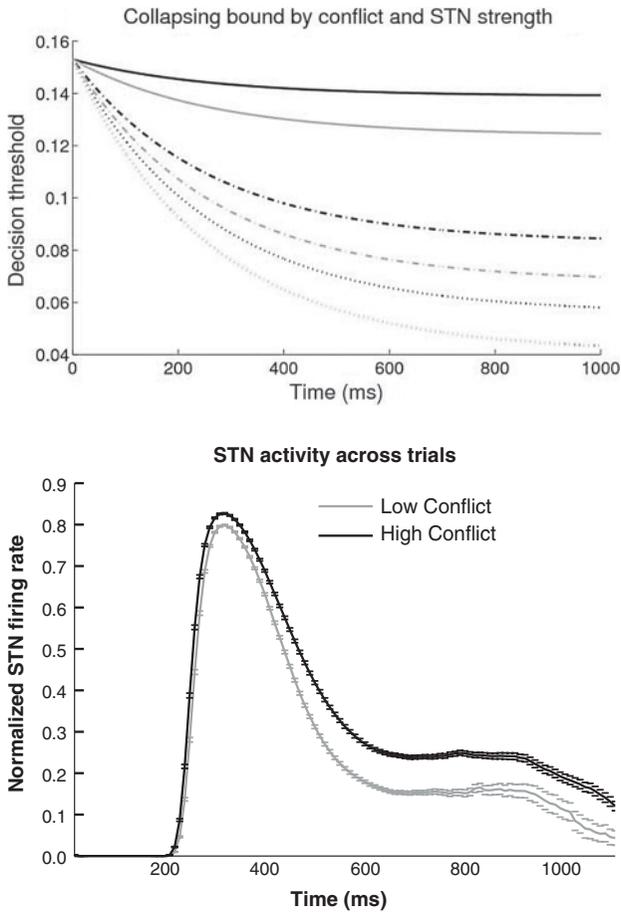
Figure 5: (Top) Plots of the decaying boundaries as a function of conflict and STN strength. The gray lines are for low conflict and the black lines for high conflict. Solid, dashed, and dotted lines are for fits to high, medium, and low STN strength, respectively. (Bottom) Plots of normalized firing rates across all STN units as a function of time, averaged across trials, for low and high conflict, respectively. Error bars reflect SEM. Note that the best-fit decision boundaries in the top panel match the profile of STN activity, where there is little difference between initial heights of the STN surge but diverging trajectories across conflict conditions with time. The effect on threshold is greater with higher STN strength.

decision bounds (see Figure 5). Indeed, although the diffusion model fits had access only to the RT distributions (see Figure 5, bottom), the best-fitting trajectories of the decision bounds closely reflected the internal temporal dynamics of the decrease in STN activity with time.

Table 2: Parameters from Diffusion Model Fits to Simulated Data from the BG Model.

Model	a Baseline	T_{er} Baseline	η	s_z	s_t	v_1 High Conflict	v_2 Low Conflict	τ
Static model	.078	.420	.349	.041	.107	.121	.337	
Dynamic model	.056	.379	.183	.016	.121	.077	.206	.249
	a_1	a_2	a_3	T_{er1}	T_{er2}	T_{er3}	a_4	$a_{initial}$
Static model	.011	.032	−.007	.016	.033	.026	−.010	
Dynamic model	.027	.056	−.015	.015	.033	.020		.153

Notes: a-baseline and T_{er}-baseline are the values of the boundary and nondecision time for go, high-conflict, and low STN. v_1 and v_2 are drift rates, η is across-trial SD in drift rate, s_z is the across-trial range starting point, s_t is the across trial range in nondecision time. a_1 and T_{er1} are the increments in boundary and nondecision time from low to medium STN, a_2 and T_{er2} are the increments in boundary and nondecision time from medium to high STN, a_3 is the increment from high to low conflict (the negative value means it is actually decremented), T_{er3} is the increment from Go to NoGo. For the static model, a_4 is the increment for NoGo (the negative value means it is actually decremented), and for the dynamic model, τ is the time constant of decay in the collapsing boundaries, and $a_{initial}$ is the initial value of the collapsing boundary.

Analysis of fitted model parameters reinforces the above interpretations. Plots of the best-fit collapsing bounds as a function of time for the different levels of STN strength and different conflict levels are presented in Figure 5, and parameter estimates are shown in Table 2. Overall, STN strength parametrically influences the decision bound in both models (parameters a_1 and a_2 in the table are both positive). Further, conflict modulates the bound within each STN level (parameter a_3, change from high to low conflict, is negative). These effects of conflict are over and above changes in drift: drift rate is lower for high conflict (v_1) than low conflict (v_2), as expected due to the different levels of evidence simulated in these conditions (see Table 2). Also in the table, nondecision time is modulated by relative NoGo activity (low dopamine–greater NoGo–slower response execution), as captured by a positive value of parameter T_{er3}. Nondecision time also parametrically increases with STN strength in both models (T_{er1} and T_{er2}), suggesting that STN activity induces a delay in the onset of evidence accumulation. Finally, follow-up simulations showed that STN strength selectively modulated boundary and nondecision time, and not drift rate or other parameters. Specifically, we tested a version of the collapsing-bound model in which all parameters were estimated separately for each level of STN strength. In these simulations, drift rates for both low and high conflict were estimated to be virtually identical across levels of STN strengths (low conflict: 0.13, 0.12, and 0.14 for low, mid-, and high-STN; high conflict: 0.31, 0.33, 0.31). In contrast, boundaries and nondecision times again increased parametrically (for low conflict and high dopamine, the boundary increased from 0.066 to

0.079 to 0.103 across levels of STN and T_{er} increased from 0.40 to 0.42 to 0.45). Because this model is more complex (having nearly three times as many parameters), we focus our analysis on the model accounting for differences in STN strength only by changes in bound and nondecision time.

Next, we present data from a reinforcement conflict experiment in young, healthy subjects, similar to that used in Parkinson's patients originally motivated by the BG model (Frank, Samanta et al. 2007; Cavanagh et al., 2011). We then apply the three diffusion model variants to determine whether they fit the data in a way that is consistent with the way in which the diffusion model had to be altered to fit data simulated from the BG model.

6 Experiment

The experiment replicates the probabilistic selection reinforcement learning task (Frank, Seeberger, & O'Reilly, 2004; Frank, Samanta et al., 2007) with college-age subjects at Ohio State University. Experimental procedures are described in detail below. The main difference between this experiment and prior versions of this task is that many more subjects were tested and many more trials were collected in the test phase following learning in order to provide enough data to examine RT distributions.

6.1 Subjects. Thirty normal healthy undergraduate students from Ohio State University and in surrounding areas participated in the experiments. All subjects were paid $12 for their participation in one 45-minute session.

6.2 Method. Pairs of letters were presented to the subject, and the subject's task was to choose one of them. The letters were presented on the screen side by side. Subjects had to respond with the / key to choose the letter on the right and the z key to choose the letter on the left. The letter pair remained on the screen until a response was made. The letters used were dissimilar consonants (QF, NB, and XT), but we refer to them here as AB, CD, and EF (for ease of presentation and comparison to prior studies with this task in which Japanese Hiragana characters were used as stimuli). In the training phase, for the AB pair, A was reinforced 80% of the time by providing a correct message 80% of the time when A was chosen and an error message the other 20% of the time. For the CD pair, C was reinforced 70% of the time, and for the EF pair, E was reinforced 60% of the time. The reinforcement probabilities of the alternative stimuli (B, D, F) are complementary $(1 - p)$ to those for the ones described (A, C, E). The letter pairs were presented in random order and random screen location (e.g., AB or BA). Feedback was presented for 300 ms followed by a 100 ms blank screen before the next pair was presented. The training phase of the experiment consisted of 360 of these trials (equivalent to the maximum of 6 blocks of 60 trials used in prior studies, although here this was broken down into 4 blocks of 90 trials). Subjects were able to take a break between blocks.

In order to examine RT distributions, the test phase was considerably longer than that typically used, consisting of 800 trials; 80% of the trials (640) used exactly the same test structure as the training phase. The remaining 20% of the trials consisted of 10 novel combinations of stimuli that had not been presented during the training phase. Each of these letter pairs was presented 15 times in the test phase. There were 100 trials in a block in the test phase.

Because there were more test trials, we provided feedback for these trials (in contrast to the standard task in which no feedback is applied during test). The reinforcement probabilities were chosen to be approximately consistent with the individual letter reinforcement probabilities, thereby maintaining the reinforcement hierarchy. Specifically, the reinforcement probabilities of the first letter of each novel pair were AC .7, AE .7, CE .6, AF .9, BD .4, BF .4, DF .4, EB .6, AD .7, and CB .7. These can be roughly divided into three classes according to the amount of reinforcement conflict engendered (Frank, 2006; Frank, Samanta et al., 2007). In high-conflict win-win choices, both letters had been positively reinforced (AC, AE, and CE); in high-conflict lose-lose, both letters had been associated with negative reinforcement (BD, BF, and DF); and in low-conflict test pairs, one stimulus was positive and the other negative (AF, EB, AD, and CD). In the data analyses, the data from these 10 conditions were grouped into the three classes, and, together with the three conditions that had been learned in the training phase (AB, CD, EF), this provided six conditions for model fitting. Thus, the grouped repaired conditions had only about 20% of the number of observations as the test conditions that were used in training (AB, CD, and EF). Because we focus here on decision making based on already learned reinforcement probabilities, only the data from the 800 test trials are presented, not the data from the 360 training trials.

Previously test phases in similar experiments usually had only 60 or 120 trials. If we used 800 trials without feedback, we would likely no longer observe conflict effects on RT as subjects begin to respond simply by stimulus-response habit. We therefore opted to continually reinforce choices in the test phase. Our assumption for fitting test trials was that almost all the learning had already occurred and the feedback just maintained the reward probabilities. Over the course of test, accuracy improved only modestly, by 1.5%, and RTs decreased by 60 ms from the first half to the second half. However, most of the decrease was in the tail of the RT distribution; the leading edge decreased by about 25 ms. From experience, averaging these together (first half versus second half) will produce parameter values that are about the average of the separate fits. We did not do this because the quality of the fits with these few data points will increase variability and reduce the quality of fits overall and reliability of parameter values a great deal.

We also included a procedure to prevent subjects from responding too fast without processing the stimulus. When a test pair was presented, if the

response was shorter than 280 ms, a message was presented indicating that the response was too fast. This message remained on the screen for 900 ms. In the data analyses, responses shorter than 280 ms and longer than 5000 ms were eliminated (this resulted in elimination of about 2% of the data).

6.3 Results.

6.3.1 Response Proportion and Quantile RTs. Table 3 shows response proportions (accuracy in choosing the optimal stimulus) and five quantile RTs for the three trained conditions and the three groups of recombined conditions. The response proportions for the optimal stimulus in low-conflict test pairs (AF, EB, AD, and CD) was about the same as the average of that for the three pairs used in training (AB, CD, and EF). Both the high-conflict win-win (AC, AE, and CE) and lose-lose (BD, BF, DF) pairs had choice proportions with lower accuracy (54% and 63% accuracy, respectively).

Median RTs (the 0.5 quantiles in Table 3) showed an increase in median RT for the high-conflict lose-lose condition relative to the win-win (and other) condition. The difference was 244 ms, but about 80 ms of this difference was due to the results from two subjects, as is shown in the second-to-bottom row of Table 3.

This shows that negative conflict delays processing more than positive conflict, a feature also found in previous data and BG models of conflict (see supplemental materials of Frank, Samanta et al., 2007, and Cavanagh et al., 2011). The speed of less rewarded choices relative to more rewarded choice is mixed. For individual subjects, sometimes they are faster and sometimes slower. In the training pairs (AB, CD, EF), the high reinforced response is faster than the low reinforced choice. These patterns of RTs for the two choices across individuals have been shown to be readily accommodated by variability in drift rate and starting point in the diffusion model (for a summary, see Ratcliff & McKoon, 2008).

For lose-lose (high-conflict), RTs are longer than for low-conflict trials, with a delay in the leading edge and a longer tail. This shift in the RT distributions is a target for modeling in the next section (cf. Ratcliff & Smith, 2010). In contrast, there is almost no delay in the leading edge for the positive win-win conflict pairs.

7 Diffusion Model Fits

The diffusion model was fit to the data from individual subjects. We fit three versions of the model. The first was the standard model with a single boundary and nondecision time across conditions, but still allowing drift rates to differ. The second allowed a different value of the nondecision time (a delay in the onset of evidence accumulation) for the lose-lose high-conflict condition. The third assumed initially raised but collapsing boundaries for the lose-lose condition with constant boundaries for the other conditions.

Table 3: Response Proportions and Quantile RTs for Data from the Experiment.

	Condition	Proportion	More Reinforced Choice					Less Reinforced Choice				
			.1Q	.3Q	.5Q	.7Q	.9Q	.1Q	.3Q	.5Q	.7Q	.9Q
Data	AB	0.855	437	515	605	744	1061	461	550	659	826	1070
	CD	0.867	441	516	592	703	1008	459	541	623	749	999
	EF	0.773	431	507	587	704	991	449	529	626	756	1040
	AC,AE,CE (win/win)	0.540	461	546	640	814	1240	444	527	627	787	1103
	BD,BF,DF (lose/lose)	0.427	572	686	884	1100	1614	553	697	874	1095	1500
	AF,EB,AD,CD (win/lose)	0.838	448	533	631	765	1137	517	676	791	928	1184
	BD,BF,DF (lose/lose) (2 subjects)	0.300	1425	1514	1979	2273	2941	1185	1581	1903	2222	2607
	BD,BF,DF (lose/lose) (28 subjects)	0.436	511	627	805	1016	1519	508	634	801	1014	1421
Collapsing-boundary model predictions	AB	0.849	390	479	569	711	1078	441	561	693	904	1415
	CD	0.862	390	479	570	712	1090	448	567	703	924	1468
	EF	0.769	397	492	590	747	1147	441	561	699	912	1462
	AC,AE,CE (win/win)	0.540	426	551	694	925	1475	443	568	714	946	1504
	BD,BF,DF (lose/lose)	0.427	551	696	854	1097	1671	543	676	825	1051	1597
	AF,EB,AD,CD (win/lose)	0.832	398	494	595	757	1170	445	567	704	926	1470

(Recall that the fits to the BG model showed both increases in nondecision time and boundary as a function of STN strength and conflict.)

We show the results in two ways. First, we show the mean response proportions and quantile RTs for the predictions of the best-fitting collapsing boundary model in Table 3. These are to be compared with the mean data values also shown in Table 3. The model with two nondecision time parameters shows almost identical fits to those of the collapsing boundary model. In contrast, the standard model with one nondecision time parameter and constant boundaries provides a large misfit, predicting quantile RTs for the lose-lose conflict condition almost the same as for the win-win conflict condition, unlike the data (see lines 4 and 5 in Table 3). However, in this standard model, the response proportions match the data.

The second display of the data shows plots of predicted and actual accuracy values against each other and of the predicted 0.1, 0.5, and 0.9 quantile RTs against each other for each of the 30 individual subjects for the collapsing boundary model. Because the training conditions have about five times the amount of data as the re-paired conditions, we would expect them to fit better because they are less noisy and weighted more. The data from the training conditions in the test phase are shown as x's in Figure 6, and the repaired conditions are shown as o's. Most of the larger misfits are from the re-paired conditions. The model with two values of the nondecision time shows quite similar plots with no deviations more extreme than those in Figure 6.

These results show that allowing only the drift rate to change is not sufficient to account for slowing in the lose-lose conflict condition, but that the same two effects needed to capture the simulated data from the BG model (adding a constant delay or assuming raised and collapsing boundaries) provide a better (adequate) fit.

The diffusion model parameters for the three models are shown in Tables 4 and 5. Across models, the parameters not used to fit the lose-lose condition are remarkably similar (these contain 94% of the data). In contrast, for the model with two nondecision times, the mean nondecision time for the lose-lose condition is over 100 ms longer than that for the other conditions. For the collapsing-boundary model, the initial boundary is over twice as large as the asymptotic boundary, and the decay constant is almost 300 ms.

The average chi square goodness-of-fit statistic for the standard model was 70.5 with 54 df, the value for the model with two nondecision time parameters was 64.0 with 53 df, and the value for the model with decaying boundaries was 62.2 with 52 df. The standard model was nested in the other two models, and the chi square values between the standard model and the other two were significantly different. For the standard model, the fits for 16 of 30 subjects were significant (indicating deviations from expected distributions predicted by the model); for the two nondecision time models, 8 fits were significant; and for the collapsing-boundary model,
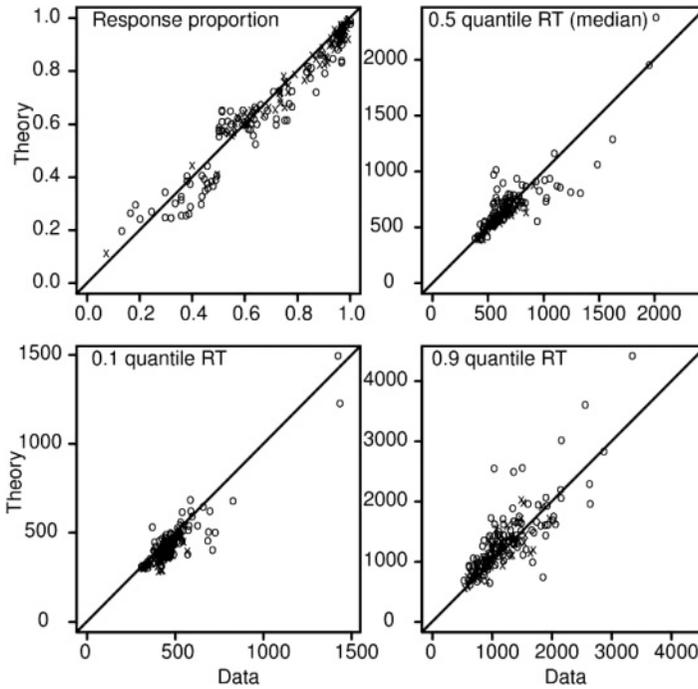
Figure 6: Plots of response proportion (accuracy) and the 0.1, 0.5, and 0.9 quantile RTs for the diffusion model predictions (with collapsing boundary) against the data. The x's are for the conditions used in training at test and the o's are for the re-paired conditions (the latter have five times less data than the former and so are more variable).

Table 4: Model Parameters for Fits to Data from the Experiment.

Model	a	T_{er}	T_{er2}	η	s_z	s_t	$a_{initial}$	Decay Const.	Chi Square
Standard	0.176	307		0.173	0.012	159			70.5
2T_{er}'s	0.175	314	422	0.191	0.012	212			64. 0
Collapse	0.175	312		0.185	0.013	180	0.370	295	62. 2

9 were significant. Thus, more than two-thirds of the subjects were well fit by the alternative models, whereas only half of them were reasonably fit by the standard model.

Although the differences in the chi square values were not large for the standard versus the other two models, the lose-lose condition (the one causing the miss) was only about 6% of the total data, so the misfit in this was large enough to produce over a 10% change in the goodness of fit. It is

Table 5: Drift Rates for Fits to Data from the Experiment.

Model	AB	CD	EF	Positive Conflict AC, AE, CE	No Conflict Re-Paired AF, EB, AD, CD	Negative Conflict BD, BF, DF
Standard	0.206	0.204	0.164	0.029	0.167	−0.021
2 T_{er}'s	0.221	0.222	0.176	0.024	0.179	−0.023
Collapse	0.218	0.215	0.168	0.020	0.180	−0.022

also worth noting that almost half the difference in chi square between the standard model and the other two models was due to relatively poor fits for the standard model to the data from just two subjects (e.g., 3.0 out of the 8.3 difference between the standard model and the collapsing-boundary model). But even for the rest of the subjects, the discrepancy was reliable. To illustrate the quality of the fits, the bottom half of Table 3 shows averages across subjects of the values of predicted response proportion and quantile RTs for the model with collapsing boundaries.

In sum, modeling the lose-lose conflict condition by either inducing a delay in onset of the decision process or allowing a temporary increase in decision boundaries produced almost the same quality of fit. Both of these assumptions are consistent with the fits of the diffusion model to the BG model.

8 Discussion

These findings demonstrate the benefit of examining multiple levels of modeling analysis, combining the best features of biologically plausible neural circuits of decision making with the theoretical grounding afforded by the diffusion model. Diffusion model fits to both empirical and simulated BG model data in the domain of reinforcement conflict–based decision making provided a unitary explanation of the effects of decision conflict. In particular, reasonable fits to the high-conflict data could be achieved by either inducing a delay in the onset of the decision process or increasing the initial decision boundary and allowing this boundary to collapse with time.

In the BG model, the decision conflict is detected by the coactivation of multiple competing motor units. Standard cortical models of decision making (Mazurek et al., 2003; Ratcliff et al., 2007; Usher & McClelland, 2001) predict that due to either lateral inhibition among cortical response units or lower drift rate, this conflict would induce slowed RTs. However, this mechanism (which, as noted above, is also present in the cortical units of the BG model) is not sufficient to account for the additional slowing observed in the high-conflict condition. In the BG model, as in the data, responses are particularly slowed in high-conflict lose-lose conditions. The slowing is observed

for two primary reasons. First, conflict detected in cortical response units activates the STN via the hyperdirect pathway. This STN surge excites the output of the BG, which inhibits the thalamus and prevents any response from being facilitated until the STN activity subsides (Frank, 2006). We showed here that parametric changes in STN strength are associated with parametric increases in decision bound and its sensitivity to conflict. This effect is exacerbated in lose-lose conflict conditions due to strong striatal NoGo unit activity (Wiecki et al., 2009) which slows RTs by adding negative evidence to each response. It also leads to greater STN recruitment (due to less feedback inhibition from GPe), thereby magnifying conflict effects.

Thus conflict and negative value conspire to produce slowed decisions under conflict (see Frank, Samanta, et al., 2007). This could be interpreted as a delay in the onset of evidence accumulation (if the STN surge is sufficient to prevent Go activity from affecting BG output). Or, alternatively (as proposed in Frank, 2006), it could be interpreted as a transient increase in decision boundaries, which then collapse as the STN activity subsides (if the STN surge only makes it more difficult for Go activity to facilitate a response). In contrast, for the positive win-win decisions, the STN activity primarily acts to compensate for impulsive speeded responding that would otherwise occur for two highly valenced decision options, due to the overabundance of striatal Go activity in this condition. Indeed, without an intact STN, a race model might be more appropriate than a diffusion model to fit the positive-conflict (win-win) decision trials. Supporting this interpretation, deep brain stimulation disrupts normal STN function and leads to win-win decisions that are even faster than low-conflict trials (Frank, Samanta, et al., 2007). Moreover, recent electrophysiological field potential recordings revealed increased activity in both mediofrontal cortex and STN as a function of conflict in the same low-frequency bands and in same time period following stimulus presentation (Cavanagh et al., 2011). That study also used diffusion model fits to estimate the degree to which variations in mediofrontal activity related to variations in decision thresholds. Notably, during high- but not low-conflict decisions, increases in mediofrontal activity were associated with increases in estimated decision thresholds. Moreover, this relationship between mediofrontal conflict and decision threshold was reversed when STN function was disrupted with deep brain stimulation (Cavanagh et al., 2011). This result provides converging evidence for the claim, explored via computational simulations here, that the interaction between cortical conflict and STN results in an adjustment of decision threshold—and not drift rates or other parameters.

The diffusion model is a particular instance of a class of sequential sampling models of decision making. We do not claim that it is the only such model that could relate to BG circuitry or the current experimental data. Indeed, similar conclusions would likely be drawn if relating the BG model to other related models, such as the linear ballistic accumulator (LBA) model (Brown & Heathcote, 2008) and the leaky competing accumulator (LCA)

model (Usher & McClelland, 2001). Specifically, these other models would have to be altered in similar ways to capture both the experimental data and the BG model simulated data. Although models differ in their other details, all of them involve a single decision boundary and nondecision value, which would need to be altered to account for the pattern of accuracy and RT distributions in the high-conflict versus low-conflict conditions (or as a function of STN strength). We found that drift rate alone was not sufficient to account for the differences between conflict conditions because accuracy was about the same for the two conditions, but there was a large shift in the RT distribution. This pattern cannot be modeled by a simple change in drift rate.

How do these results relate to other accounts of decision threshold variation in fronto-BG circuitry? Forstmann et al. (2008, 2010) showed that the change of decision threshold (estimated with either the diffusion model or LBA) due to speed versus accuracy instructions is accompanied by changes in functional connectivity between preSMA and striatum (rather than STN). However, as Bogacz, Wagenmakers, Forstmann, and Nieuwenhuis, (2010) noted, this may reflect a shift from the default tendency to focus on accuracy to a controlled objective to speed up, which may result from increasing the baseline activation of the response units in preSMA and striatum. Such an effect would result in a lowered effective decision threshold, which is formally equivalent to maintaining a fixed threshold but adding a constant amount of evidence to both accumulators so that they begin closer to threshold. Indeed, other neural models of BG circuitry have shown similar effects by increasing synaptic connection strengths from cortex to striatum (Lo & Wang, 2006), which would be equivalent to increasing weights from cortex to Go units in our model and therefore making it easier to disinhibit a response. We have focused here on an alternative mechanism in the STN route, which dynamically alters the amount of evidence required and is engaged when slowing down due to conflict or errors, or shifting from a prepotent response to a controlled one. Other neuroimaging and neurophysiological data support this role for the STN hyperdirect pathway (Aron, Behrens, Smith, Frank, & Poldrack, 2007; Isoda & Hikosaka, 2008; Fleming, Thomas, Dolan, 2010; Jahfari et al., 2011).

In principle, our goal to relate the diffusion model to corticostriatal circuitry shares much in common with the model of Bogacz and Gurney (2007). Indeed, they proposed that the BG precisely implemented the diffusion model (for two alternative choices) or the multiple sequential probability ratio test (MSPRT) for more than two alternatives. However, the approach taken was to derive what functional form the "neurons" in this model would have to take in order for the relationship with the diffusion model to hold. As a result, the model could not have produced anything other than what it was designed to do, and (understandably) a number of simplifying assumptions had to be made for this to be the case. For example, each striatal unit exactly copies the activity signaled by the cortical input (implying a linear transfer function), omits the indirect pathway altogether, and does not

allow for transient surges in STN activity at the onset of the decision process that decay. Nevertheless, on the surface, the Bogacz model makes a similar prediction about the role of STN in responding to conflict. Indeed, the STN conflict computations in their model are critical, implementing a diffusion process such that evidence for the winning accumulator must be sufficiently greater than the conflict term in order for a response to be selected. However, the form of the conflict function in Bogacz models is actually quite different from that assumed here: it is proportional to the sum of the evidence across all accumulators. This means that a decision in which one option has high value and another has low value (e.g., win-lose) would be predicted to have greater conflict (and increased STN activity) relative to a decision in which both options have low values (lose-lose). Notably, their model does not predict that the decision boundary, if estimated with a diffusion model, is further increased with any type of conflict. The diffusion model has already taken into account the conflict by accumulating evidence in the form of a difference signal. Indeed, because the Bogacz model is formally equivalent to the diffusion model for two alternatives, it could not account for the data in high-conflict lose-lose conditions without additional modifications (Note that we make no claim about optimality here, but see below.)

In contrast, our approach has been to explore how the dynamics of distributed neuronal activity may support learning and decision-making processes without attempting to derive closed-formed solutions (due to in-herent nonlinear complexity). Nevertheless, by fitting the diffusion model to the output of the BG model, we showed that reasonable fits to the simu-lated data can be obtained by assuming a delay in processing or collapsing boundaries in the lose-lose conditions. The same assumptions allow the diffusion model to fit human experimental data well. These results show how levels of dopamine and conflict affect diffusion model parameters.

It is currently unclear whether the effects of reinforcement conflict on decision bound that we observe here would extend to other kinds of conflict paradigms. For example, in the Eriksen paradigm (Eriksen & Eriksen, 1974) subjects must identify a target item that is surrounded (flanked) by items that, on conflict trials, suggest the other response to the target. An example of this is an arrow or angle bracket signaling the target as a response to the right (>) but with flankers that suggest a response to the left (<<><<). White, Brown, and Ratcliff (in press) and White, Ratcliff, and Starns (in press) found that models that assumed that drift rate gradually changed over time as processing focused more and more on the target provided the best account of processing in this task. However, it is also possible that the same STN conflict mechanism is involved to transiently raise the bound in other situations. For example, consider a Simon task in which subjects have to make a left response to a left arrow that appears on the right side of the screen. Initially, the right response is captured by the location of the stimulus, but after a period of time, the subject can detect conflict between this response and the rule. Under this situation, the STN

may reflect temporary response conflict to prevent the initial prepotent response from reaching threshold. Indeed, this situation is analogous to that observed in monkey studies in which the STN increases its firing rate when automatic responses conflict with controlled responses, leading to a delay in the RT distribution (Isoda & Hikosaka, 2008). Moreover, STN DBS elicits premature impulsive responding in these conflict conditions just as it does in the reinforcement conflict tasks (Wylie et al., 2010).

Our results also have theoretical implications for decision-making and diffusion processes in general. In almost all applications, the diffusion model assumes that the decision process turns on with a constant drift rate. However, there are two commonsense possibilities in which drift is nonstationary that are worth discussing. First, it could be that the rate of drift changes with time, such that it ramps gradually up to some constant value. It turns out that the original model with constant drift from the onset of the decision process actually mimics the gradual ramp. Ratcliff (2002) generated simulated data with the assumption that the drift rate ramped up over a 50 ms interval followed by a constant drift rate. When the constant drift model was fit to the simulated data, it fit well with three changes in the parameter values, in particular, the nondecision time (T_{er}) was delayed by 25 ms, and variability in both nondecision time (s_t) and starting point (s_z) (Laming, 1968) increased. Therefore, even if in many situations the drift rate ramps up, this can almost be perfectly mimicked by a constant drift model.

The second case, more proximally related to the current project, involves a zero drift for some duration followed by an abrupt increase to a constant drift rate. Such an assumption might seem a natural way to fit the lose-lose conditions in the reinforcement learning paradigm, as these conditions might simply delay the availability of discriminative information even though the decision process had begun. However, a delay in onset of discriminative information will not produce the correct predictions for RT distributions. In the experiment, there was a shift in the leading edge of the RT distribution for the lose-lose condition. This shift is well accommodated by an increase in the nondecision time or a large increase in the initial decision bound followed by decay or both. But such a shift is not well captured by zero drift because within trials, noise causes processes to terminate and hence would elicit only slight delays in the leading edge. Indeed, to accommodate the delay in the leading edge with zero drift, variability in the accumulation process would need to be reduced to near zero to avoid processes prematurely terminating (Laming, 1968; Smith & Ratcliff, 2009; see also Ratcliff & Smith, 2010). The BG model offers a proposal about which circuits control the onset of the decision process. We therefore feel that successful integration of the BG model with the diffusion model might begin to allow these questions to be addressed.

Others have explored collapsing boundaries in diffusion models. For example, Frazier and Yu (2008) have shown that the collapsing boundary is optimal for a diffusion process if a time deadline is imposed, but they

do not examine reinforcement conflict paradigms. It is possible that when faced with multiple seemingly negative options, it is helpful to initially avoid whichever responses are considered, but then eventually to reduce the boundary so that one is not subject to decision paralysis. Indeed, while there was no deadline imposed in the test phase, subjects knew that they must respond at some point in order to advance to the next trial and eventually finish the experiment. Similarly, we have proposed that the STN is involved in the phenomenon of the paradox of choice, in which choices are avoided or deferred in the face of conflict, but that the reduction in STN activity with time is helpful to reduce decision thresholds such that one is not subject to complete decision paralysis.

The contribution of this research is to show how data from the conflict conditions in the reinforcement learning paradigm cannot be fit in the traditional way with the diffusion model (and this applies to other models of this class). We presented two simple modifications, a delay in processing or collapsing decision bounds, and showed how these allow the model to fit. These modifications were motivated by the BG model which implements equivalent mechanisms to account for conflict in reinforcement learning. We also showed that the output from the BG model using parameter values typical of those used in other applications was well fit by the diffusion model (with the modifications for the conflict conditions). These results show how a simple computational model can inform a more complicated but realistic neural model designed to account for the control processes as well as decision processes.

Appendix: BG Model Implementational Details

For animated video captures of model dynamics during response selection and learning, see http://ski.clps.brown.edu/BGmodel_movies.html. The BG model can be obtained by e-mailing michael_frank@brown.edu. Several demonstrations of reinforcement learning processes are available for download at http://ski.clps.brown.edu/BG_Projects.

The model is implemented with the emergent neural simulation software package (Aisa, Mingus, & O'Reilly, 2008), adapted to simulate the anatomical and physiological properties of the BG circuitry in reinforcement learning and decision making (Frank, 2006). This framework uses point neurons with excitatory, inhibitory, and leak conductances contributing to an integrated membrane potential, which is then thresholded and transformed to produce a rate code output communicated to other units. In the BG model, discrete spiking can also be used and produces similar results for decision making (but requires additional considerations to function in learning environments).

Model parameters remain unchanged from several prior simulations and are listed in Frank (2006). The equations that below are written in

general form; parameters vary according to physiological properties of different BG nuclei. For example, GPi/GPe units are tonically active in the absence of synaptic input, whereas striatal units fire only with convergent excitatory synaptic input from sensory input and preSMA. The model neuron parameters below are adjusted to capture these properties as described in Frank (2006).

The membrane potential V_m is updated as a function of ionic conductances g with reversal (driving) potentials E according to the following differential equation:

$$C_m \frac{dV_m}{dt} = g_e(t)\bar{g}_e(E_e - V_m) +$$
$$g_i(t)\bar{g}_i(E_i - V_m) +$$
$$g_l(t)\bar{g}_l(E_l - V_m) +$$
$$g_a(t)\bar{g}_a(E_a - V_m), \tag{A.1}$$

where C_m is the membrane capacitance and determines the time constant with which the voltage can change, and subscripts e, l, i, and a refer to excitatory, leak, inhibitory, and accommodation channels, respectively (accommodation is included in STN, which allows the surge of STN activity to decline). The reversal or equilibrium potentials E_c determine the driving force of each of the channels, whereby E_e is greater than the resting potential and E_l and E_i are typically less than resting potential (with the exception of tonically active neurons in GPi and GPe, where leak drives current into the neuron; Frank, 2006). Following electrophysiological convention, the overall conductance for each channel c is decomposed into a time-varying component $g_c(t)$ computed as a function of the dynamic state of the network and a constant $\bar{g}_c$ that controls the relative influence of the different conductances. The equilibrium potential can be written in a simplified form by setting the excitatory driving potential (E_e) to 1 and the leak and inhibitory driving potentials (E_l and E_i) of 0,

$$V_m^\infty = \frac{g_e\bar{g}_e}{g_e\bar{g}_e + g_l\bar{g}_l + g_i\bar{g}_i}, \tag{A.2}$$

which shows that the neuron is computing a balance between excitation and the opposing forces of leak and inhibition. This equilibrium form of the equation can be understood in terms of a Bayesian decision-making framework, whereby the neuron evaluates whether the excitatory evidence for the "hypothesis" it is detecting according to its synaptic weights is sufficiently greater than the evidence against that hypothesis. In the preSMA, gaussian noise ($\mu = 0$, $\sigma = 0.002$) is added to the membrane potential of each unit, producing temporal variability in the extent to which each candidate response is activated before one of them is gated by the BG. Accommodation currents in the STN build up with time-integrated activity.

The excitatory net input/conductance $g_e(t)$ is computed as the proportion of open excitatory channels as a function of sending activations times the weight values:

$$g_e(t) = \langle x_i w_{ij} \rangle = \frac{1}{n} \sum_i x_i w_{ij}. \tag{A.3}$$

For units with inhibitory inputs from other layers (the red projections in Figure 1), predominant in the basal ganglia, the inhibitory conductance is computed similarly, whereby $g_i(t)$ varies as a function of the sum of the synaptic inputs. Dopamine also adds an inhibitory current to the NoGo units, simulating effects of D2 receptors. (See below for a simplified implementation of within-layer lateral inhibition.) Leak is a constant.

Activation communicated to other cells (y_j) is a thresholded (Θ) sigmoidal function of the membrane potential with gain parameter γ:

$$y_j(t) = \frac{1}{\left(1 + \frac{1}{\gamma [V_m(t) - \Theta]_+}\right)}, \tag{A.4}$$

where $[x]_+$ is a threshold function that returns 0 if $x < 0$ and x if $X > 0$. Note that if it returns 0, we assume $y_j(t) = 0$, to avoid dividing by 0. As it is, this function has a very sharp threshold, which interferes with graded learning mechanisms (e.g., gradient descent). To produce a less discontinuous deterministic function with a softer threshold, the function is convolved with a gaussian noise kernel ($\mu = 0$, $\sigma = .005$), which reflects the intrinsic processing noise of biological neurons:

$$y_j^*(x) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma} e^{-z^2/(2\sigma^2)} y_j(z - x) dz, \tag{A.5}$$

where x represents the $[V_m(t) - \Theta]_+$ value and $y_j^*(x)$ is the noise-convolved activation for that value.

A.1 Inhibition Within and Between Layers. Inhibition between layers (i.e., for GABAergic projections from striatum to GPi/GPe, GPe to GPi/STN, and GPi to thalamus) is achieved via simple unit inhibition, where the inhibitory current g_i for the unit is determined from the net input of the sending unit in the same way as described for g_e (see above).

For within-layer lateral inhibition (used in striatum and preSMA), Leabra uses a kWTA (k-winner-takes-all) function to achieve inhibitory competition among units within each layer (area). The kWTA function computes a uniform level of inhibitory current for all units in the layer, such that the $k + 1$th most excited unit within a layer is generally below its firing threshold, while the kth is typically above threshold. Activation dynamics

similar to those produced by the kWTA function have been shown to result from simulated inhibitory interneurons that project both feedforward and feedback inhibition, and indeed other versions of the BG model use explicit populations of striatal inhibitory interneurons (Wiecki et al., 2009). Thus, the kWTA function provides a computationally effective and efficient approximation to biologically plausible inhibitory dynamics.

kWTA is computed via a uniform level of inhibitory current for all units in the layer as follows:

$$g_i = g_{k+1}^{\Theta} + q\big(g_k^{\Theta} - g_{k+1}^{\Theta}\big), \tag{A.6}$$

where $0 < q < 1$ (.25 default used here) is a parameter for setting the inhibition between the upper bound of g_k^{Θ} and the lower bound of g_{k+1}^{Θ}. These boundary inhibition values are computed as a function of the level of inhibition necessary to keep a unit right at threshold:

$$g_i^{\Theta} = \frac{g_e^* \bar{g}_e (E_e - \Theta) + g_l \bar{g}_l (E_l - \Theta)}{\Theta - E_i}, \tag{A.7}$$

where g_e^* is the excitatory net input.

Two versions of kWTA functions are typically used in Leabra. In the kWTA function used in the striatum, g_k^{Θ} and g_{k+1}^{Θ} are set to the threshold inhibition value for the kth and $k+1$th most excited units, respectively. Thus, the inhibition is placed to allow k units to be above threshold and the remainder below threshold.

The preSMA uses the average-based kWTA version, g_k^{Θ} is the average g_i^{Θ} value for the top k most excited units, and g_{k+1}^{Θ} is the average of g_i^{Θ} for the remaining $n - k$ units. This version allows more flexibility in the actual number of units active depending on the nature of the activation distribution in the layer and the value of the q parameter (which is set to default value of .6). This flexibility is necessary for the premotor units to have differential levels of activity during settling (depending on whether a single response has been facilitated), and also allows greater activity in high-conflict trials.

A.2 Connectivity. The connectivity of the BG network is critical, and is thus summarized here (see Frank, 2005, 2006, for details and references). Unless stated otherwise, projections depicted in Figure 1 are fully connected (that is all units from the source region target the destination region, with a randomly initialized synaptic weight matrix). However the units in preSMA, striatum, GPi, GPe, and thalamus are all organized with columnar structure. Units in the first column of preSMA represent one response and project to a single column of each of the Go and NoGo units in the striatum, which in turn project to the corresponding columns in GPi/GPe and the thalamus. Each thalamic unit is reciprocally connected with the associated

column in preSMA. This connectivity is similar to that described by anatomical studies, in which the same cortical region that projects to the striatum is modulated by the output through the BG circuitry and Thalamus.

In contrast to the focused connectivity in the striatal Go and NoGo pathways representing separate cortical responses, the projections from the cortex to the STN, and from STN to GPi, are fully connected, representing the diffused projections in this hyperdirect pathway that support a Global NoGo function.

Dopamine units in the SNc project to the entire striatum, but with different projections to encode the effects of D1 receptors in Go neurons and D2 receptors in NoGo neurons. Specifically, dopamine influences Go unit activity levels by contributing an excitatory current in Go units and an inhibitory current in NoGo units, matching the differential effects of dopamine on postsynaptic activity associated with these receptors.

Thus, increases in firing of SNc dopamine units promote active Go units to become more active (or more excitable to cortical input) and NoGo units to become less active, and vice-versa for decreases in dopamine). However, the particular set of units affected by dopamine is determined by those receiving excitatory input from sensory cortex and preSMA. Thus, dopamine modulates this activity, thereby affecting the relative balance of Go versus NoGo activity in those units activated by cortex. Given that one of the preSMA responses is more strongly activated than the other (subject to noise), the corresponding striatal Go and NoGo units for this response will also be preferentially active. In other applications, these corticostriatal synaptic strengths are learned, such that it is possible for each preSMA response to be equally active but for the striatum to nevertheless show preferential selection of one of them based on reward history. However, as noted in the text, here we consider the case that learning has already occurred and the correct response is preferentially active due to bias on the preSMA units.

A.3 Parametrically Modulating Scaling of STN-GPi Projections. In the main text we explored the effects of modulating the relative contributions of the STN by changing the strength of the overall projections from STN to GPi. Specifically, the equation for computing excitatory conductances (see equation A.3) is subject to additional scaling of overall projections so that some "count" more than others (meant to reflect the fact that neurons from some areas synapse close to or farther from cell bodies and thereby have more or less influence on postsynaptic potentials than neurons from other areas, given the same synaptic weights). Specifically, excitatory conductances g_{e_k} arising as a function of different projections k are scaled as

$$g_{e_k} = \frac{r_k}{\sum_p r_p} \langle x_i w_{ij} \rangle_k,$$ (A.8)

where r_k reflects the relative scaling of the projection k normalized by the sum of the scalings across all projections p. In the text we modified r_k from STN to GPi units from low (0.4) to mid (0.55, default) to high (0.7). This then alters the contributions of the STN to GPi membrane potentials relative to other GPi inputs (striatum, GPe).

A4 Learning. While learning is not relevant for the current simulations (we refer interested readers to prior papers for mathematical details), a core function of this BG model to learn stimulus-response-reinforcement probabilities as a function of dopaminergic reward prediction error signals. Furthermore, as the striatum becomes more and more likely to facilitate the most rewarding responses, the input to preSMA synaptic strengths evolves to reflect the prior probability that a given response had been selected in the past given the stimulus. Thus, with extended training, the correct stimulus-response links are learned directly from sensory cortex to preSMA. At this stage, the cortex can be said to "select" the action, with the striatum simply facilitating it, with differential speed depending on, for example, dopamine levels. In the current simulations, we simply hand-coded the stimulus-response strengths from sensory cortex to preSMA in order to precisely manipulate the level of conflict.

Acknowledgments

References

Aisa, B., Mingus, B., & O'Reilly, R. C. (2008). The emergent neural modeling system. *Neural Networks*, *21*, 1146–1152.

Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted MRI and functional MRI. *Journal of Neuroscience*, *27*, 3743–3752.

Ashby, F. G., Ennis, J. M., & Spiering, B. J. (2007). A neurobiological theory of automaticity in perceptual categorization. *Psychological Review*, *114*, 632–656.

Bogacz, R., Brown, E., Moehlis, J., Holmes, P., & Cohen, J. D. (2006). The physics of optimal decision making: A formal analysis of models of performance in two-alternative forced choice tasks. *PsychologicalReview*, *113*, 700–765.

Bogacz, R., & Gurney, K. (2007). The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Computation*, *19*, 442–477.

Bogacz, R., Wagenmakers, E-J., Forstmann, B. U., & Nieuwenhuis, S. (2010). The neural basis of the speed-accuracy tradeoff. *Trends in Neuroscience*, *33*, 10–16.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict monitoring and cognitive control. *Psychological Review*, *108*, 624–652.

Brown, S. D., & Heathcote, A. J. (2008). The simplest complete model of choice response time: Linear ballistic accumulation. *Cognitive Psychology*, *57*, 153–178.

Busemeyer, J. R., & Townsend, J. T. (1993). Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment. *Psychological Review*, *100*, 432–459.

Cavanagh, J. F., Frank, M. J., Klein, T. J., & Allen, J. J. B. (2010). Frontal theta links prediction errors to behavioral adaptation in reinforcement learning. *NeuroImage*, *49*, 3198–3209.

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C., Norton, T., Sherman, S. J., et al. (2011). Subthalamic stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience*, *14*, 1462–1467.

Chevalier, G., & Deniau, J. M. (1990). Disinhibition as a basic process in the expression of striatal function. *Trends in Neuroscience*, *13*, 277–280.

Churchland, A. K., Kiani, R., & Shadlen, M. N. (2008). Decision-making with multiple alternatives. *Nature Neuroscience*, *11*, 693–702.

Ding, L., & Gold, J. I. (2010). Caudate encodes multiple computations for perceptual decisions. *Journal of Neuroscience*, *30*, 15747–15759.

Ditterich, J. (2006). Stochastic models of decisions about motion direction: Behavior and physiology. *Neural Networks*, *19*, 981–1012.

Doll, B. B., & Frank, M. J. (2009). The basal ganglia in reward and decision making: computational models and empirical studies. In J. Dreher & L. Tremblay (Eds.), *Handbook of reward and decision making* (pp. 399–425). Orlando, FL: Academic Press.

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank , M. J. (2009). Instructional control of reinforcement learning: A behavioral and neurocomputational investigation. *Brain Research*, *1299*, 74–94.

Donaldson, D. I., Wheeler, M. E., & Petersen, S. E. (2010). Remember the source: Dissociating frontal and parietal contributions to episodic memory. *Journal of Cognitive Neuroscience*, *22*, 377–391.

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception and Psychophysics*, *16*, 143–149.

Fleming, S. M., Thomas, C. L., & Dolan, R. J. (2010). Overcoming status quo bias in the human brain. *Proceedings of the National Academy of Sciences*, *107*, 6005–6009.

Forstmann, B. U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E. J., et al. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences*, *107*, 15916–15920.

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., et al. (2008). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences*, *105*, 17538–17542.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *Journal of Cognitive Neuroscience*, *17*, 51–72.

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks*, *19*, 1120–1136.

Frank, M. J., & Badre, D. (2011). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex*. doi. 10.1093/cercor/bhr114.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making and reversal. *Psychological Review*, *113*, 300–326.

Frank, M. J., Doll, B. B., Oas-Terpstra, J., & Moreno, F. (2009). Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. *Nature Neuroscience*, *12*, 1062–1068.

Frank, M. J., Moustafa, A. A., Haughey, H., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences*, *104*, 16311–16316.

Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007). Hold your horses: Impulsivity, deep brain stimulation and medication in Parkinsonism. *Science*, *318*, 1309–1312.

Frank, M. J., Scheres, A., & Sherman, S. J. (2007). Understanding decision making deficits in neurological conditions: Insights from models of natural action selection. *Philosophical Transactions of the Royal Society-B*, *362*, 1641–1654.

Frank, M. J., Seeberger, L., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science*, *306*, 1940–1943.

Frazier, P., & Yu, A. J. (2008). Sequential hypothesis testing under stochastic deadlines. In J. C. Platt, D. Koller, Y. Singer, & S. Roweis (Eds.), *Advances in Neural Information Processing Systems, 20* (pp. 465–472). Cambridge, MA: MIT Press.

Gold, J. I., & Shadlen, M. N. (2001). Neural computations that underlie decisions about sensory stimuli. *Trends in Cognitive Science*, *5*, 10–16.

Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535–574.

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

Gurney, K. N., Prescott, T. J., & Redgrave, P. (2001b). A computational model of action selection in the basal ganglia II: Analysis and simulation of behaviour. *Biological Cybernetics*, *84*, 411–423.

Isoda, M., & Hikosaka, O. (2008). A neural correlate of motivational conflict in the superior colliculus of the macaque. *Journal of Neurophysiology*, *100*, 1332–1342.

Jahfari, S., Waldorp, L., van den Wildenberg, W. P., Scholte, H. S., Ridderinkhof, K. R., & Forstmann, B. U. (2011). Effective connectivity reveals important roles for both the hyperdirect (fronto-subthalamic) and the indirect (fronto-striatal-pallidal) fronto-basal ganglia pathways during response inhibition. *Journal of Neuroscience*, *31*, 6891–6899.

Krajbich, I., Armel, C., & Rangel, A. (2010). Visual fixations and the computation and comparison of value in simple choice. *Nature Neuroscience*, *13*, 1292–1298.

Laming, D.R.J. (1968). *Information theory of choice reaction time*. New York: Wiley.

Levy, R., Hutchison, W. D., Lozano, A. M., & Dostrovsky, J. O. (2000). High-frequency synchronization of neuronal activity in the subthalamic nucleus of Parkinsonian patients with limb tremor. *Journal of Neuroscience*, *20*, 7766–7775.

Lo, C-C., & Wang, X-J. (2006). Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nature Neuroscience*, *9*, 956–963.

Magill, P. J., Sharott, A., Bevan, M. D., Brown, P., & Bolam, J. P. (2004). Synchronous unit activity and local field potentials evoked in the subthalamic nucleus by cortical stimulation. *Journal of Neurophysiology*, *92*, 700–714.

Mazurek, M. E., Roitman, J. D., Ditterich, J., & Shadlen, M. N. (2003). A role for neural integrators in perceptual decision-making. *Cerebral Cortex*, *13*, 1257–1269.

Milosavljevic, M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction times of value-based choice under high and low time pressure. *Judgment and Decision Making*, *5*, 437–449.

Moustafa, A. A., Cohen, M. X., Sherman, S. J., & Frank, M. J. (2008). A role for dopamine in temporal decision making and reward maximization in Parkinsonism. *Journal of Neuroscience*, *28*, 12294–12304.

Philiastides, M. G., Ratcliff, R., & Sajda, P. (2006). Neural representation of task difficulty and decision making during perceptual categorization: A timing diagram. *Journal of Neuroscience*, *26*, 8965–8975.

Ratcliff, R. (1978). A theory of memory retrieval. *Psychological Review*, *85*, 59–108.

Ratcliff, R. (1988). Continuous versus discrete information processing: Modeling the accumulation of partial information. *Psychological Review*, *95*, 238–255.

Ratcliff, R. (2002). A diffusion model account of reaction time and accuracy in a two choice brightness discrimination task: Fitting real data and failing to fit fake but plausible data. *Psychonomic Bulletin and Review*, *9*, 278–291.

Ratcliff, R. (2006). Modeling response signal and response time data. *Cognitive Psychology*, *53*, 195–237.

Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of simple two-choice decisions. *Journal of Neurophysiology*, *90*, 1392–1407.

Ratcliff, R., Hasegawa, Y. T., Hasegawa, Y. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, *97*, 1756–1774.

Ratcliff, R., & McKoon, G. (2008). The diffusion decision model: Theory and data for two-choice decision tasks. *Neural Computation*, *20*, 873–922.

Ratcliff, R., Philiastides, M. G., & Sajda, P. (2009). Quality of evidence for perceptual decision making is indexed by trial-to-trial variability of the EEG. *Proceedings of the National Academy of Sciences*, *106*, 6539–6544.

Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333–367.

Ratcliff, R., & Smith, P. L. (2010). Perceptual discrimination in static and dynamic noise: The temporal relation between perceptual encoding and decision making. *Journal of Experimental Psychology: General*, *139*, 70–94.

Ratcliff, R., Thapar, A., Smith, P. L., & McKoon, G. (2005). Aging and response times: A comparison of sequential sampling models. In J. Duncan, P. McLeod, & L. Phillips (Eds.), *Measuring the mind: Speed, control, and age* (pp. 3–32), New York: Oxford University Press.

Ratcliff, R., Thapar, A., & McKoon, G. (2010). Individual differences, aging, and IQ in two-choice tasks. *Cognitive Psychology*, *60*, 127–157.

Ratcliff, R., & Tuerlinckx, F. (2002). Estimating the parameters of the diffusion model: Approaches to dealing with contaminant reaction times and parameter variability. *Psychonomic Bulletin and Review*, *9*, 438–481.

Ratcliff, R., & Van Dongen, H. P. A. (2011). A diffusion model for one-choice reaction time tasks and the cognitive effects of sleep deprivation. *Proceedings of the National Academy of Sciences*, *108*, 11285–11290.

Roe, R. M., Busemeyer, J. R., & Townsend, J. T. (2001). Multialternative decision field theory: A dynamic connectionist model of decision-making. *Psychological Review*, *108*, 370–392.

Smith, P. L. (1995). Psychophysically principled models of visual simple reaction time. *Psychological Review*, *102*, 567–591.

Smith, P. L., & Ratcliff, R. (2004). The psychology and neurobiology of simple decisions. *Trends in Neuroscience*, *27*, 161–168.

Smith, P. L., & Ratcliff, R. (2009). An integrated theory of attention and decision making in visual signal detection. *Psychological Review*, *116*, 283–317.

Tuerlinckx, F., Maris, E., Ratcliff, R., & De Boeck, P. (2001). A comparison of four methods for simulating the diffusion process. *Behavior Research Methods, Instruments, and Computers*, *33*, 443–456.

Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550–592.

Viviani, P. (1979). A diffusion model for discrimination of temporal numerosity. *Journal of Mathematical Psychology*, *19*, 108–136.

Wheeler, M. E., Petersen, S. E., Nelson, S. M., Ploran, E. J., & Velanova, K. (2008). Dissociating early and late error signals in perceptual recognition. *Journal of Cognitive Neuroscience*, *20*, 2211–2225.

White, C. N., Brown, S., & Ratcliff, R. (in press). A test of Bayesian observer models of processing in the Eriksen flanker task. *Journal of Experimental Psychology: Human Perception and Performance*.

White, C. N., Ratcliff, R., & Starns, J. J. (in press). Diffusion models of the flanker task: Discrete versus gradual attentional selection. *Cognitive Psychology*.

Wiecki, T. V., & Frank, M. J. (2010). Neurocomputational models of motor and cognitive deficits in Parkinson's disease. *Progress in Brain Research*, *183*, 275–297.

Wiecki, T. V., Riedinger, K., Meyerhofer, A., Schmidt, W., & Frank, M. J. (2009). A neurocomputational account of catalepsy sensitization induced by D2-receptor-blockade in rats: Context-dependency, extinction, and renewal. *Psychopharmacology*, *204*, 265–277.

Wylie, S. A., Ridderinkhof, K. R., Elias, W. J., Frysinger, R. C., Bashore, T. R., Downs, K. E., et al. (2010). Subthalamic nucleus stimulation influences expression and suppression of impulsive behaviour in Parkinson's disease. *Brain*, *133*, 3611–3624.