A Computational Model of Inhibitory Control in Frontal Cortex and Basal Ganglia

Thomas V. Wiecki and Michael J. Frank
Brown University

Planning and executing volitional actions in the face of conflicting habitual responses is a critical aspect of human behavior. At the core of the interplay between these 2 control systems lies an override mechanism that can suppress the habitual action selection process and allow executive control to take over. Here, we construct a neural circuit model informed by behavioral and electrophysiological data collected on various response inhibition paradigms. This model extends a well-established model of action selection in the basal ganglia by including a frontal executive control network that integrates information about sensory input and task rules to facilitate well-informed decision making via the oculomotor system. Our simulations of the anti-saccade, Simon, and saccade-override tasks ensue in conflict between a prepotent and controlled response that causes the network to pause action selection via projections to the subthalamic nucleus. Our model reproduces key behavioral and electrophysiological patterns and their sensitivity to lesions and pharmacological manipulations. Finally, we show how this network can be extended to include the inferior frontal cortex to simulate key qualitative patterns of global response inhibition demands as required in the stop-signal task.

Keywords: response inhibition, neural network model, cognitive control, prefrontal cortex, basal ganglia

Supplemental materials: http://dx.doi.org/10.1037/a0031542.supp

Before you act, listen. Before you react, think. Before you spend, earn. Before your criticize, wait.

—Ernest Hemingway

This quote by Ernest Hemingway highlights our basic tendency to act impulsively while reminding us that sometimes it is advisable to inhibit these prepotent response biases and act more thoughtful. Recent scientific advancements have shed light on the neural and cognitive mechanisms that implement inhibitory control of prepotent response biases (Andrés, 2003; Aron, 2007; Logan, 1985; Miyake et al., 2000; Munoz & Everling, 2004;

Thomas V. Wiecki and Michael J. Frank, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University.

Correspondence concerning this article should be addressed to Michael J. Frank, Department of Cognitive, Linguistic, and Psychological Sciences, Brown University, Box 1821, 190 Thayer Street, Providence, RI 02912-1821. E-mail: Michael_Frank@brown.edu or thomas_wiecki@brown.edu

Stuphorn & Schall, 2006). As part of this effort, a multitude of tasks exist to study response inhibition empirically. Among the tasks thought to require selective response inhibition are the anti-saccade task, the Simon task, and the saccade-override task. Each of these tasks induces a prepotent response bias that sometimes needs to be overridden with a controlled response based on executive control. For example, the anti-saccade task requires subjects to saccade in the opposite direction of an appearing stimulus. The Simon task requires subjects to respond according to an arbitrary stimulus-response rule (e.g., respond left or right depending on stimulus color), but where the stimulus is presented on one side of the screen, inducing a prepotent response bias to that side. In congruent trials, the stimulus is presented on the same side as the correct response indicated by the rule, whereas on incongruent trials it is on the opposite side. Finally, the saccade-override task (Isoda & Hikosaka, 2007) requires subjects to saccade in the direction of a stimulus of a particular color for several repetitions in a row. On so-called switch-trials, the instruction cue indicates that the other colored stimulus is now the target, so that the participant has to override the initial planned response and switch to the other one. While critical differences exist, all of these tasks require subjects to inhibit a prepotent response and replace it with a different response. In contrast, while also requiring response inhibition, the well-studied stop-signal task does not require subsequent initiation of an active response but only outright inhibition of the planned response (Verbruggen & Logan, 2008).

Electrophysiological and functional imaging data implicate key nodes in frontostriatal circuitry as being active during response inhibition and executive control. At the cortical level, these include the right inferior frontal gyrus (rIFG; Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003; Chambers et al., 2007; Sakagami et

al., 2001; Xue, Aron, & Poldrack, 2008), the dorsolateral prefrontal cortex (DLPFC; Funahashi, Chafee, & Goldman-Rakic, 1993; Johnston & Everling, 2006; Wegener, Johnston, & Everling, 2008), the supplementary eye fields (SEF; Schlag-Rey, Amador, Sanchez, & Schlag, 1997), the presupplementory motor area (pre-SMA; Aron, Behrens, Smith, Frank, & Poldrack, 2007; Congdon, Constable, Lesch, & Canli, 2009; Isoda & Hikosaka, 2007), and the frontal eye fields (FEF; Munoz & Everling, 2004). At the subcortical level, the striatum (Ford & Everling, 2009; Watanabe & Munoz, 2011; Zandbelt & Vink, 2010), the subthalamic nucleus (STN; Aron, Behrens, et al., 2007; Aron & Poldrack, 2006; Eagle et al., 2008; Hikosaka & Isoda, 2008; Isoda & Hikosaka, 2008), and the superior colliculus are involved. Manipulations that disrupt processing in either frontal or subcortical areas cause deficits in response inhibition (Chambers et al., 2007; Ray et al., 2009; Verbruggen, Aron, Stevens, & Chambers, 2010). Moreover, response inhibition deficits are commonly observed in a wide range of psychiatric patients with frontostriatal dysregulation, including attention-deficit/hyperactivity disorder (ADHD; Nigg, 2001; Oosterlaan, Logan, & Sergeant, 1998; Schachar & Logan, 1990), obsessive compulsive disorder (OCD; Chamberlain, Fineberg, Blackwell, Robbins, & Sahakian, 2006; Menzies et al., 2007; Morein-Zamir, Fineberg, Robbins, & Sahakian, 2010; Penadés et al., 2007), schizophrenia (SZ; Badcock, Michie, Johnson, & Combrinck, 2002; Bellgrove et al., 2006; Huddy et al., 2009), Parkinson's disease (PD; van Koningsbruggen, Pender, Machado, & Rafal, 2009), and substance abuse disorders (Monterosso, Aron, Cordova, Xu, & London, 2005; Nigg et al., 2006).

Together, the above data suggest that intact functioning of the entire fronto-basal ganglia network is required to support response inhibition. However, it is far from clear that the underlying source of these deficits is the same. Inhibitory control is a very dynamic process, influenced by different interacting cognitive variables and neuromodulatory systems. Thus, response inhibition can be impacted by not only dysfunctional stopping *per se,* but can also be influenced by changes in motivational state (Leotti & Wager, 2010), attentional saliency (Morein-Zamir & Kingstone, 2006), maintenance and retrieval of task rules (Hutton & Ettinger, 2006; Nieuwenhuis, Broerse, Nielen, & de Jong, 2004; Reuter & Kathmann, 2004; Roberts, Hager, & Heron, 1994), and separable modulations of selective versus global inhibition mechanisms (Aron, 2011), to name a few. Although electrophysiological recording studies demonstrate neuronal populations that differentiate between successful and unsuccessful stopping (Isoda & Hikosaka, 2007, 2008), or inhibition of prepotent responses in favor of controlled responses (Ford & Everling, 2009; Watanabe & Munoz, 2009), there is at present no coherent framework integrating all of these findings into a single model attempting to account for patterns of electrophysiological data, or selective disruptions of component parts and their effects on behavior.

The point of departure for our neural model builds on existing theorizing and data regarding the differential roles of the three main pathways linking frontal cortex with the basal ganglia (BG), often referred to as the direct, indirect, and hyperdirect pathways. According to this framework, the cortico-striatal direct "Go" and indirect "NoGo" pathways together implement a selective gating mechanism by computing the evidence for facilitating or suppressing each of the candidate motor actions identified by frontal cortex. Dopamine plays a critical role in this model by differentially modulating the activity levels in the two striatal populations, affecting both learning and choice. During rewards and punishments, phasic bursts and dips in dopamine neurons convey reward prediction errors (Montague, Dayan, & Sejnowski, 1996) that transiently amplify Go or NoGo activity states, and therefore activity-dependent plasticity. In this manner, these striatal populations learn the positive and negative evidence for each cortical action (Frank, 2005). More chronic increases in tonic dopamine levels also directly affect choice by shifting the overall balance of activity toward the Go pathway over the NoGo pathway, thereby emphasizing learned positive relative to negative associations and speeding responding (and vice-versa for tonic decreases in dopamine). Many of this model's predictions have been validated with behavioral studies involving dopaminergic manipulations and functional imaging in humans and monkeys (e.g., Frank, Seeberger, & O'Reilly, 2004; Jocham, Klein, & Ullsperger, 2011; Nakamura & Hikosaka, 2006; Palminteri et al., 2009; Voon et al., 2010), and synaptic plasticity and opto-genetic and genetic engineering studies in rodents (Hikida, Kimura, Wada, Funabiki, & Nakanishi, 2010; Kravitz et al., 2010; Kravitz, Tye, & Kreitzer, 2012; Shen, Flajolet, Greengard, & Surmeier, 2008).

Note that in the above model, responses are *selectively* facilitated or suppressed via separate striatal Go and NoGo populations modulating the selection of particular cortical actions. However, more recent models have also incorporated the third hyperdirect pathway from frontal cortex to the STN to BG output. Communication along this pathway provides a *global* and dynamic regulation of the gating threshold, by transiently suppressing the gating of all responses when there is conflict between alternative actions (Frank, 2006; Ratcliff & Frank, 2012). Empirical studies using STN manipulations (Cavanagh et al., 2011; Frank, Samanta, Moustafa, & Sherman, 2007; Wylie et al., 2010), direct recordings (Cavanagh et al., 2011; Isoda & Hikosaka, 2008; Zaghloul et al., 2012), and fMRI/DTI (Aron, Behrens, et al., 2007) have similarly supported this notion.

Nevertheless, the existing BG model cannot handle situations in which an initial prepotent response is activated but then needs to be suppressed (either altogether, or in favor of a more controlled response)—situations typically studied under the rubric of "response inhibition." Here, we extend the model by incorporating additional cortical regions that facilitate executive control and can inhibit and override the more habitual response selection mechanism. We consider dynamics of the prepotent response process, the subsequent detection that this response needs to be inhibited, and the inhibition process itself—and how all of these factors are modulated by biological and cognitive variables. We consider electrophysiological data in various frontal (DLPFC, FEF, pre-SMA, ACC) and basal ganglia (striatum, STN) regions that are well captured by the model, and how these are linked to changes in behavior in a functional way.

Neural models are complex, in that they involve a number of parameters interacting to produce nonlinear effects on dynamics and behavior. There is also a risk of overfitting that could result from adjusting parameters to precisely match electrophysiological data from one experiment, which may make it difficult to precisely capture electrophysiological (or behavioral) data from a different experiment. Thus, our aim was instead to capture qualitative patterns of data in both electrophysiology at multiple levels of cortical and subcortical network, and of the effects of their ma-

nipulation on behavior, with a single set of parameters.[1] In other work (Wiecki & Frank, 2010a), we show that systematic variations in neural model parameters are related in a lawful, monotonic fashion to more computational level parameters in a modified drift diffusion framework, providing a principled understanding and falsifiable experimental predictions. Moreover, despite the qualitative nature of model fits, we nevertheless aim to distinguish our model from others in the literature based on general principles independent of particular parameterizations. Toward this goal we extracted a set of qualitative behavioral and neurocognitive benchmark results (listed in the Results section) which we use to assess the validity of our model and compare to other models.

As noted above, despite surface features suggesting a single integrated response inhibition network, there are actually multiple dynamic components that can affect inhibition. Our contribution in this article is to formalize these separable neural processes and to explore their interactive dynamics. To summarize and preview the core aspects of our work:

• We present a neural network model of the three main frontal-BG pathways supporting prepotent action selection, inhibitory control, conflict-induced slowing, and volitional action generation.

• We show that behavioral changes in a range of tasks dependent on these basic processes can result from alterations in brain connectivity and state and provide testable predictions for effects of distinct brain disorders.

• Selective response inhibition involves global conflict-induced slowing via the hyperdirect pathway, raising the effective decision threshold to prevent prepotent responding, followed by DLPFC induction of striatal NoGo activity to inhibit the planned prepotent response. Subsequently, the DLPFC provides top-down facilitation onto striatal Go populations encoding the controlled response.

• Response selection and inhibition are further regulated by neuromodulatory influences including dopamine linked to changes in motivational and attentional state. Dopamine reflects potential reward values and facilitates Go actions. In addition, our model suggests that while selective response inhibition is influenced by tonic levels of dopamine (DA), global response inhibition is not.

• Our model is challenged in its ability to overcome prepotent responses and evaluated by its ability to reproduce key qualitative patterns reported in the literature, including:

– Behavioral RT distribution patterns in selective response inhibition tasks.
– Electrophysiological activity patterns of the FEF (Everling & Munoz, 2000), pre-SMA (Hikosaka & Isoda, 2008), the STN (Isoda & Hikosaka, 2008), striatum (Watanabe & Munoz, 2009), superior colliculus (SC; Paré & Hanes, 2003; Pouget et al., 2011), and scalp recordings (Yeung, Botvinick, & Cohen, 2004).
– Psychiatric, developmental, lesion, and pharmacological manipulations of frontal function and DA modulations.

• We show that when our model is extended to include the rIFG it can recover key electrophysiological and behavioral data from the stop-signal task literature.

In sum, this approach provides a mechanistic account of a major facet of cognitive control and executive functioning, which we hope will allow for a richer understanding of the relationship between behavioral, imaging, and patient findings.

Neural Network Model

We first introduce the neural circuit model of interacting dynamics among multiple frontal and basal ganglia nodes and their modulations by dopamine. We then describe how we vary model parameters to capture biological and cognitive manipulations.

Overview

The model is implemented in the Emergent software (Aisa, Mingus, & O'Reilly, 2008) with the neuronal parameters adjusted to approximate known physiological properties of the different areas (Frank, 2005, 2006). The simulated neurons use a rate-code approximation of a leaky integrate-and-fire neuron (henceforth referred to as units) with specific channel conductances (excitatory, inhibitory, and leak). Multiple units (simulated neurons) are grouped together into layers which correspond to distinct anatomical regions of the brain. Units within each layer project to those in downstream areas, and in some cases, when supported by the anatomy, there are bidirectional projections (e.g., bottom-up superior colliculus projection to cortex as well as top-down projections from cortex to colliculus). We summarize the general functionality of the model here to foster an intuitive understanding; implementational and mathematical details can be found in the Appendix. While a single set of core parameters (i.e., integration dynamics and overall connection strength between layers) is used to simulate various electrophysiological and behavioral data in the intact state, each reported simulation is tested on 8 networks with randomly initialized weights between individual neurons. The model can be downloaded from our online-repository http://ski.clps.brown.edu/BG_Projects.

The model represents an extension of our established model of the BG (Frank, 2005, 2006; Wiecki & Frank, 2010b). Because the extended model involves multiple components, we progressively introduce each part, beginning with its core and then describing how each new component contributes additional functionality.

Basic Basal Ganglia Model

The architecture of the core model is similar to Frank (2006). While the original model simulated manual motor responses, our model features a slightly adapted architecture in accordance to the neuroanatomy and physiology underlying rapid eye-movements (i.e., saccades) as reviewed in Hikosaka (2007) and Munoz and Everling (2004). Stimuli are presented to the network in the input layer, corresponding to high level sensory cortical representations. An arbitrary number of motor responses can be simulated, but here

---

[1] By *qualitative*, we mean that we do not attempt to quantitatively fit the precise shape of firing of any given cell type, but we do aim to show that a given population of cells increases or decreases firing rate at a particular point in time relative to some task event or to some estimated cognitive process. For example, for an area to be involved in inhibition, it must show increased activity prior to the time it takes to inhibit a response. Or in striatum, particular cell populations are active related to biasing the prepotent response, suppressing that response, and then activating the controlled response—our model recapitulates this qualitative pattern.

we include a model with just two candidate responses. The input layer projects directly to the cortical response units in the frontal eye fields (FEF) which implements action planning and monitoring and projects to the superior colliculus (SC), which acts as an output for saccade generation (Sparks, 2002). The SC consists of two units coding for a leftward and a rightward directed saccade. If the firing rate of one unit crosses a threshold, the corresponding saccade is initiated (Everling, Dorris, Klein, & Munoz, 1999). The time it takes an SC unit to cross its threshold from trial onset is taken as the network's response time (RT). Stimulus-response mappings can be prepotently biased by changing projection strengths (i.e., weights) so that certain input patterns preferentially activate a set of FEF response units more than the alternative response units. (These sensory-motor cortical weights can also be learned from experience, such that they come to reflect the prior probability of selecting a particular response given the sensory stimulus [Frank, 2006].) In fact, with only these three structures our model would only be capable of prepotent, inflexible responding.

By itself, FEF activation is not sufficiently strong to initiate saccade generation because the SC is under tonic inhibition from the BG output nucleus: the substantia nigra pars reticulata (SNr), whose neurons fire at high tonic rates. However, the tonic SNr–SC inhibition is removed following activation of corresponding direct (Go) pathway striatal units, which inhibit the SNr, and therefore disinhibit the SC (Goldberg, Farries, & Fee, 2012; Hikosaka, 1989; Hikosaka, Takikawa, & Kawagoe, 2000). The indirect pathway acts in opposition to the direct pathway by further exciting the SNr (indirectly, via inhibitory projections to the globus pallidus [GP] which inhibits the SNr). Thus, direct pathway activity results in gating of a saccade (i.e., Go) while indirect pathway activity prevents gating (i.e., NoGo). Striking evidence for this classical model was recently presented by optogenetic stimulation selectively of direct or indirect pathways cells, showing inhibition or excitation of SNr, respectively, and resulting in increased or decreased movement (Kravitz et al., 2010).

The Go and NoGo striatal populations include multiple units that code for the positive and negative evidence in favor of the FEF candidate actions given the sensory input context. Relative activity of the striatal pathways is modulated by dopaminergic innervation from the substantia nigra pars compacta (SNc) due to differential simulated D1 and D2 receptors present in the two pathways. In particular, dopamine further excites active Go units while inhibiting NoGo units. These effects on activity also produce changes in activity-dependent plasticity, allowing corticostriatal synaptic strength in the Go population to increase following phasic dopamine bursts during rewarding events, and those in the NoGo population to decrease (and vice-versa for negative events; Frank, 2005). For simplicity, in the present model, we omit learning because the paradigms we simulate do not involve learning, and focus on associations that have already been learned. However, it is now well-known that striatal unit activity is modulated by the reward value of the candidate action, such that rewarding saccades are more likely to be disinhibited (Hikosaka, Nakamura, & Nakahara, 2006).

Bottom-up projections from SC to FEF allow action-planning to be modulated according to direct and indirect pathway activity (Sommer & Wurtz, 2002, 2004a, 2004b, 2006). This effectively forms a closed loop in which FEF modulates the striatum which,

via gating through SNr and SC, in turn modulates the FEF. Loosely, FEF considers the candidate responses and "asks" the BG if the corresponding action should be gated or not. Thus, with these structures the model can selectively gate responses modulated by DA.

In addition to the above gating dynamics, the overall threshold for gating is controlled by the ease with which the SNr units are inhibited by the striatal Go units. The STN sends diffuse excitatory projections to the SNr (Parent & Hazrati, 1995), and therefore when STN units are active they increase the gating threshold for all responses, effectively contributing a "global NoGo" signal (Frank, 2006; Ratcliff & Frank, 2012). The STN does not however, act as a static increase in threshold. Rather, the STN receives input directly from frontal cortex, and becomes more active when there is response conflict (or choice entropy) during the early response selection process. In the current model, conflict is computed explicitly by the dorsal anterior cingulate cortex (dACC), which detects when multiple competing FEF response units are activated concurrently, and in turn activates the STN to make it more difficult to gate any response until this conflict is resolved. The full computational role of dACC is far from resolved and likely to be more complex than conflict detection and control (see, e.g., Alexander & Brown, 2011; Botvinick, Cohen, & Carter, 2004; Holroyd & Coles, 2002; Kolling, Behrens, Mars, & Rushworth, 2012). Nevertheless, alternative accounts of dACC function (Kolling et al., 2012) are entirely compatible with our model (an issue we return to in the Discussion section), but for convenience we label the computation as "conflict."

Frontal Pathway Model

Volitional response selection.   So far our model is able to select/gate responses and slow down gating when an alternative response appears to have some value relative to the initial planned action.

However, selective response inhibition tasks (SRITs) require executive control: integration of the sensory state together with the task rule to not only inhibit the prepotent response but replace it with a volitional one. Such rule-based processing is effortful and time-consuming, and hence the controlled response process lags that of the initial fast response capture. Based on a variety of evidence, we ascribe the rule-based representations to the dorsolateral prefrontal cortex (DLPFC; e.g., Chambers, Garavan, & Bellgrove, 2009; Miller & Cohen, 2001). This structure is involved in the active maintenance of stimulus-response rule representations (Brass, Derrfuss, Forstmann, & von Cramon, 2005; Derrfuss, Brass, Neumann, & von Cramon, 2005; Derrfuss, Brass, & von Cramon, 2004), is necessary for correct anti-saccade trials (Funahashi et al., 1993; Johnston & Everling, 2006; Wegener et al., 2008), and is involved in selective response inhibition (Garavan, Hester, Murphy, Fassbender, & Kelly, 2006; Simmonds, Pekar, & Mostofsky, 2008) and response selection (Braver, Barch, Gray, Molfese, & Snyder, 2001; Rowe, Friston, Frackowiak, & Passingham, 2002). Moreover, SEF (Schlag-Rey et al., 1997) and pre-SMA (Isoda & Hikosaka, 2007; Ridderinkhof, Forstmann, Wylie, Burle, & van den Wildenberg, 2011) are also critically involved in correct SRIT performance.

We consequently added an abstract executive control layer to summarize the DLPFC, SEF, and pre-SMA complex (in the future

referred to as DLPFC). This layer selects FEF responses and biases BG gating according to task rules (see Figure 1). Although not explicitly represented separately in the model architecture, we conceptualize the individual contribution of DLPFC as rule encoding and abstract action selection, whereas SEF and pre-SMA are transforming this abstract action representation into concrete motor-actions (Curtis & D'Esposito, 2003; Schlag-Rey et al., 1997; Schlag-Rey & Schlag, 1984). In turn, these planned motor actions can influence the selected response in FEF and bias gating via projections to striatal Go and NoGo neurons (Munoz & Everling, 2004).

Figure 1. Box-and-arrow view of the neural network model. The sensory input layer projects to the frontal eye fields (FEF), striatum, and executive control (i.e., dorsolateral prefrontal cortex [DLPFC], supplementary eye fields [SEF], and presupplementory motor area [pre-SMA]). Via direct projections to FEF (i.e., cortico-cortical pathway), stimulus-response-mappings can become ingrained (habitualized). FEF has excitatory projections to the superior colliculus (SC) output layer that executes saccades once a threshold is crossed. However, under baseline conditions, SC is inhibited by tonically active substantia nigra pars reticulata (SNr) units. Thus, for SC units to become excited, they have to be disinhibited via striatal direct pathway Go unit activation and subsequent inhibition of corresponding SNr units. Conversely, responses can be selectively suppressed by striatal NoGo activity, via indirect inhibitory projections from striatum to globus pallidus (GP) and then to SNr. Coactivation of mutually incompatible FEF response units leads to dorsal anterior cingulate cortex (dACC) activity (conflict or entropy in choices), which activates the subthalamic nucleus (STN). This STN surge makes it more difficult to gate a response until the conflict is resolved, via excitatory projections to SNr, effectively raising the gating threshold. Striatum is innervated by dopamine (DA) from substantia nigra pars compacta (SNc), which amplifies Go relative to NoGo activity in proportion to reward value and allows the system to learn which actions to gate and which to suppress. The instruction layer represents abstract task rule cues (e.g., anti-saccade trial). The DLPFC integrates the task cue together with the sensory input (i.e., stimulus location) to initiate a controlled response corresponding to task rules, by activating the appropriate column of units in FEF and striatum. Cong = congruent; Incong = incongruent; R1 = Response 1; R2 = Response 2; ACC = anterior cingulate cortex; GPe = external segment of the globus pallidus.

Anatomical and functional studies demonstrate projections from both DLPFC to SEF and pre-SMA (Lu, Preston, & Strick, 1994; Y. Wang, Isoda, Matsuzaka, Shima, & Tanji, 2005) and to striatum to affect response gating (Doll, Jacobs, Sanfey, & Frank, 2009; Frank & Badre, 2012; Haber, 2003), and from SEF to FEF (Huerta, Krubitzer, & Kaas, 1987). We explore how these projections impact dynamics of response selection. But how does the executive controller in our model "know" which rule to activate? We do not address here how these rule representations arise via learning, which is the focus of other PFC-BG modeling studies (see Collins & Frank, 2012; Frank & Badre, 2012; Rougier, Noelle, Braver, Cohen, & O'Reilly, 2005). Instead, we simulate the state of the network after learning by simply including an Instruction layer as a second input layer to the model encoding task condition (e.g., anti-saccade trial). In case of the anti-saccade task, the sensory input layer encodes the direction of the visual stimulus and the instruction layer encodes whether the network should perform a pro- or anti-saccade. The DLPFC complex then integrates these two inputs and activates a (pre-specified) rule unit that (i) projects to the correct FEF response units supporting the anti-saccade, (ii) activates striatal NoGo units to prevent gating of the active pre-potent pro-saccade response, and (iii) activates striatal Go units encoding the controlled anti-saccade.

Critically, DLPFC units are relatively slow to activate the appropriate rule unit. This is due to the need to formulate a conjunctive rule representation between the visual location of the stimulus and the task instruction (either one of these is not sufficient to determine the correct response, and indeed, each individual input provides evidence for multiple potential rules). Time constants of membrane potential updating is reduced to support this integration, which also is intended to approximate slower time course of rule retrieval and subsequent computation to determine the correct action (via interactions with pre-SMA and SEF). Moreover, we include considerable inter-trial noise in DLPFC activation dynamics so that executive control is available earlier on some trials while later on others. The slowed integration and the increase of inter-trial noise in executive control are necessary for the model to capture the quantitative benchmark results (demonstrated below). Moreover, the slower controlled processing is also a core feature of classical dual process models of cognition (e.g., Sloman, 1996) and the increased noise accords with the general statistical observation that longer latencies are typically associated with greater variability.

Competition between the two response selection mechanisms. As outlined above, our model features two response selection mechanisms: (i) a fast, prepotent mechanism driven by a biased projection from sensory input to FEF; and (ii) a slow, volitional mechanism that originates in the DLPFC which integrates instruction and sensory input to select and gate the correct response. Importantly, the volitional mechanism is slower but stronger than the prepotent one. If, due to noise in the speed of integration, executive control is slower on some trials, it might be too late to activate the correct rule representation before the prepotent response is gated. In contrast, when the executive controller is faster, it activates the alternative FEF response, leading to conflict-induced slowing, and then actively suppresses the prepotent response via projections to striatal NoGo units encoding the pro-saccade. This conceptualization can be regarded as a biologically plausible implementation of the cognitive activation-suppression model (Ridderinkhof, 2002; Ridderinkhof, van den
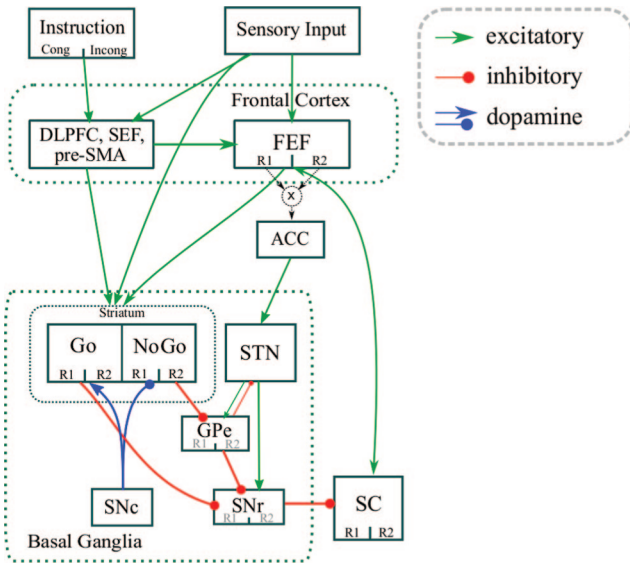
Wildenberg, Segalowitz, & Carter, 2004). Note however that our implementation involves two suppression mechanisms, one in which conflict results in global threshold adjustment, and another in which the prepotent response is selectively inhibited.

Modulations. To test the influence of different biological manipulations on executive control paradigms, we modify various parameters in the network model. Here, we list the different modulations and their implementation.

• *Prepotency:* To simulate differences in the strength of the prepotent response capture of an appearing stimulus (e.g., the prosaccade stimulus), we modulate the projection strength between sensory input to the dominant response units in FEF and striatum.

• *Speed of DLPFC:* To simulate efficacy of prefrontal function, we modulate the speed of DLPFC integration by adjusting the time constant of membrane potential updating in these units. Faster updating implies proactive control.

• *Connectivity of DLPFC:* To simulate differences in intra-cortical connectivity, we modulate the DLPFC → FEF projection strength.

• *Speed–accuracy trade-off:* To simulate strategic adjustments in the speed–accuracy trade-off, we modulate the connection strength between frontal cortex and striatum (Forstmann et al., 2010). In particular, when speed is emphasized, the FEF more effectively activates striatal Go units so that it is easier to reach gating threshold. In contrast, accuracy adjustments are reflected in increased STN baseline ultimately increasing the response gating threshold.

• *STN impact:* STN contributions are simulated by manipulating the relative synaptic strengths from STN to SNr, effectively changing the amount of STN activity required to prevent BG gating (Cavanagh et al., 2011; Ratcliff & Frank, 2012).

• *Tonic DA:* Pharmacological and disease modulations of DA levels are simulated by either decreasing (e.g., PD) or increasing (e.g., SZ) tonic DA activity, which in turn modulates relative activity of Go versus NoGo units.

Selective Response Inhibition

Method. As summarized earlier, all SRITs have a common task structure. (i) A prepotent response bias is induced by priming an action. In the anti-saccade task, this is a result of the appearance of a stimulus that initiates a "visual grasping reflex" (Hess, Bürgi, & Bucher, 1946); in the Simon task, this is the result of placing the target stimuli on either side of the screen, initiating a response capture (Ridderinkhof, 2002); in the saccade-overriding task this is the result of repeated responding to the same colored stimulus which renders this response habitual. (ii) In congruent trials, the correct response is the same as the prepotently biased one. (iii) In incongruent trials, the correct response is incompatible with the prepotently biased response, and subjects can use executive control to suppress the initially predominant action in favor of the task-appropriate one.

We implemented this common task structure as follows in our neural network model (see Figure 2; alternative task implementations that accommodate the differences between the tasks lead to similar patterns, so we simplified in order to use a single task representation of this basic process, but nevertheless simulate patterns of data evident in specific tasks below). Two stimulus positions, left and right, were encoded in the input layer as two distinct columns of activated units. The prepotent bias toward an appearing target was hard-coded by strong weights from each input stimulus to corresponding response units in FEF. This prepotent weight facilitates fast responding for

congruent trials, but biases responding in the erroneous direction for incongruent trials. The DLPFC layer integrates sensory input and instruction input to activate a conjunctive rule unit encoding the unique combination of sensory and instruction input, which then projects to the associated correct response unit in FEF. Each of the four DLPFC units project to the appropriate FEF response unit. Note that weights from the DLPFC to the FEF are stronger than the prepotent bias connection from the input layer to the FEF so that the DLPFC would eventually override an erroneous prepotent response. (The same functionality could be achieved by simply allowing DLPFC units to reach a higher firing rate or to engage a larger population of units, instead of adjusting the weights.) In addition, DLPFC units activate corresponding Go and NoGo units in the striatum (e.g., in an anti-saccade trial, Go units coding for the correct response and NoGo units coding for the incorrect response get activated by top-down PFC input).

Results. We identified a set of key behavioral and neurophysiological qualitative patterns across SRITs that form desiderata for our model to capture:

1. Incongruent trials are associated with higher error rates than congruent trials (e.g., Isoda & Hikosaka, 2008; McDowell et al., 2002; Reilly, Harris, Keshavan, & Sweeney, 2006).

2. Reaction times (RTs) are faster for errors than correct trials (e.g., Isoda & Hikosaka, 2008; McDowell et al., 2002; Reilly et al., 2006).

3. Strategic adjustments in the speed–accuracy trade-off (via changes in decision threshold) modulates functional connection strength between frontal cortex and striatum (Forstmann et al., 2010). Similarly, STN activity is associated with modulations of the decision threshold (Cavanagh et al., 2011; Ratcliff & Frank, 2012).

4. Various psychiatric diseases associated with frontostriatal catecholamine dysregulation lead to increased error rates and speeded responses (e.g., Harris, Reilly, Keshavan, & Sweeney, 2006; McDowell et al., 2002; Reilly et al., 2006; Reilly, Harris, Khine, Keshavan, & Sweeney, 2007).

5. Early activation of prepotent motor response, for example, in EMG measurements (Burle, Possamaï, Vidal, Bonnet, & Hasbroucq, 2002).

6. At least four different types of activation dynamics in FEF neurons during correct and error incongruent trials (Everling & Munoz, 2000). Specifically, neurons coding for the erroneous (i.e., prepotent) response are fast to activate and their activity is greater on error trials than correct trials. In contrast, neurons coding for the correct (i.e., controlled) response are slower to activate and their activity is reduced and delayed on error trials. (See Figure 6c for the quantitative data that forms the basis of this qualitative pattern.)

7. At least four different types of striatal neurons with dissociable dynamics and direction selectivity in congruent and incongruent trials (Ford & Everling, 2009; Watanabe & Munoz, 2009). Specifically, (i) during pro-saccades, distinct
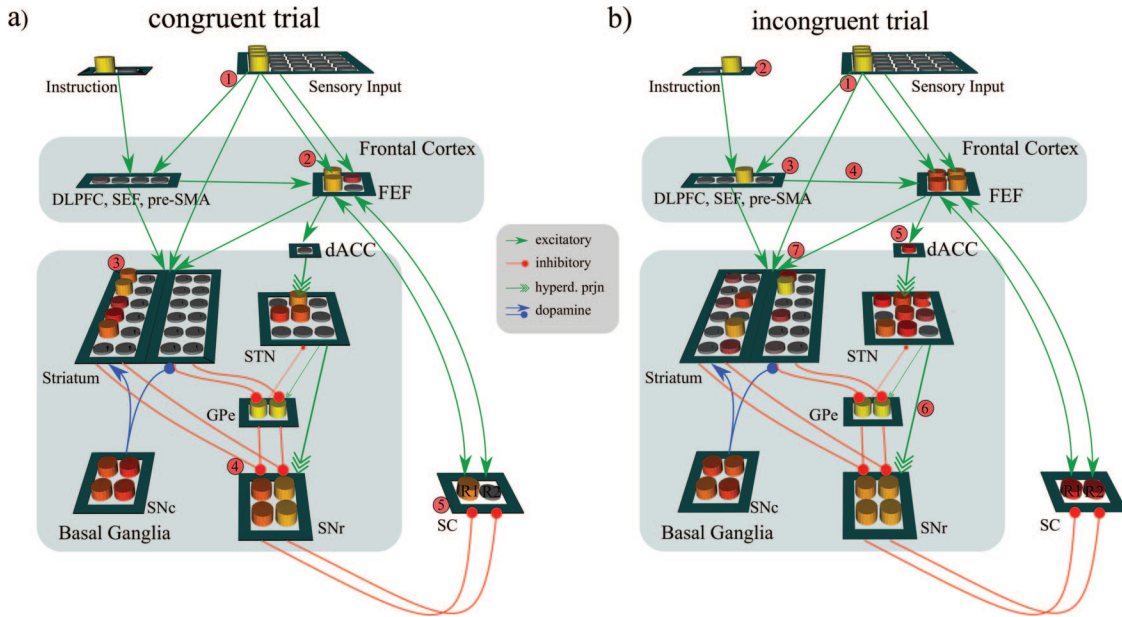
Figure 2. Neural network model in different task conditions. (a) Pro-saccade condition. (1) Left stimulus is presented in input layer; (2) Prepotent weights bias left response coding units in frontal eye fields (FEF); (3) Left response Go gating neurons in striatum are activated; (4) Left response coding units in substantia nigra pars reticulata (SNr) are inhibited; (5) The left response unit in superior colliculus (SC) is disinhibited and, due to recurrent excitatory projections with FEF, is excited and the action is executed. (b) Anti-saccade condition. The activity pattern early in the trial (i.e., before dorsolateral prefrontal cortex [DLPFC] comes online) is similar to that in the pro-saccade condition. (1) Left stimulus is presented in input layer activating prepotent left response in FEF; (2) The unit coding for the anti-saccade condition is externally activated in instruction layer; (3) DLPFC integrates sensory and instruction input according to task rules and activates *right* coding units in FEF together with *right* Go gating units and *left* NoGo units in striatum; (4) In FEF, right coding units are activated due to DLPFC input in addition to the prepotent left coding units already active; (5) Dorsal anterior cingulate cortex (dACC) detects co-activation of multiple FEF action plans and activates (6) hyperdirect pathway to excite subthalamic nucleus (STN) and SNr, globally preventing gating until conflict is resolved. Eventually, stronger controlled DLPFC activation of the right coding FEF response results in gating of the correct anti-saccade (7). In some trials, DLPFC activation is too late, and the prepotent left saccade will have already crossed threshold, resulting in an error. SEF = supplementary eye fields; pre-SMA = presupplementory motor area; hyperd. prjn = hyperdirect projection; GPe = external segment of the globus pallidus; SNc = substantia nigra pars compacta.

neural populations code for facilitation of the correct response and suppression of the alternative; (ii) during anti-saccade trials, (iia) neurons coding for facilitation of the *incorrect* prepotent response initially become active but return to baseline when (iib) neurons coding for the suppression of that response become active together with (iic) neurons coding for facilitation of the correct controlled response (see Figure 9b).

8. Neurons forming part of the hyperdirect pathway from frontal cortex (pre-SMA, dACC) to the STN show increased activity (i) *before* correct incongruent responses and (ii) *after* incorrect incongruent responses, but (iii) baseline activity during congruent responses (Isoda & Hikosaka, 2007, 2008; Yeung et al., 2004; Zaghloul et al., 2012). This pattern of activity co-occurs with delayed but more accurate incongruent responding.

In the following, we demonstrate how our model reproduces these qualitative patterns.

Behavior. As expected, intact networks make considerably more errors on incongruent trials (error rate of 15%) compared to perfect performance in congruent trials (error rate close to 0%, not shown), thereby capturing Qualitative Pattern 1.

Further, networks in general have longer response times (RTs) in incongruent trials (see Figure 3b), thus capturing Qualitative Pattern 2. Incongruent trials are slower for two reasons: (i) it takes time for executive control (DLPFC) computations due to the requirement to integrate two sources of input to activate the associated rule; and (ii) once activated, the controlled response conflicts with the prepotent response, leading to STN activation and associated increases in BG gating threshold.

Additional analysis revealed that incongruent error trials are associated with faster RTs compared to correctly performed incongruent trials (see Figure 4). In our model, errors are made when the faster prepotent action reaches threshold before the inhibitory process can cancel it. This mechanism allows the model to capture Qualitative Patterns 2 and 3.
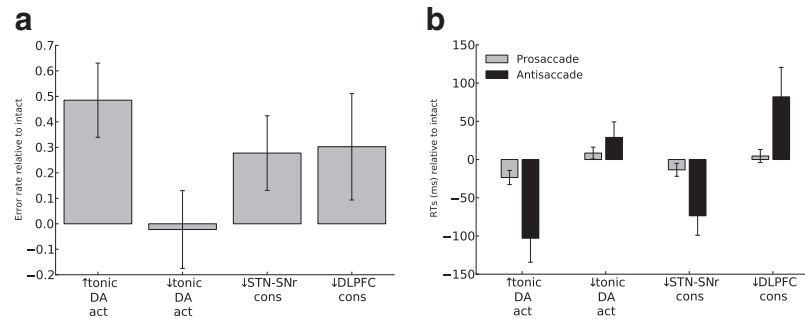
Figure 3. (a) Error rates in incongruent trials ± *SEM* relative to intact networks for different neural manipulations. Networks make more errors with increased tonic dopamine (DA) levels, or subthalamic nucleus (STN) dysfunction, compared to intact networks. (b) Response times (RTs) ± *SEM* relative to intact networks, for pro- and anti-saccade trials as a function of neural manipulations. For more analysis, see the main text. SNr = substantia nigra pars reticulate; DLPFC = dorsolateral prefrontal cortex.

We next investigated how these behavioral patterns were affected by manipulations (see Figure 3a). Incongruent error rates were most exaggerated with increased tonic DA levels, and by disrupted STN function to simulate deep brain stimulation. The effect of increased striatal DA on incongruent error rates captures corresponding patterns (see Qualitative Pattern 4) observed in non-medicated schizophrenia patients, who have elevated striatal DA (e.g., Harris et al., 2006; McDowell et al., 2002; Reilly et al., 2006, 2007). Tonic DA elevations are associated with speeded responding in both congruent and incongruent trials, due to shifted balance toward the Go pathway facilitating response gating. This same mechanism explains the increased anti-saccade error rate. Conversely, decreased tonic DA leads to slowed responding due to increased excitability of the indirect NoGo pathway. The model also predicts that STN dysfunction produces increased error rates, due to an inability to raise the threshold required for striatal facilitation of prepotent responses. Indeed, STN-DBS induces impulsive (fast but inaccurate) responding in SRITs (Wylie et al., 2010).

Finally, we tested in more detail how systematic parametric changes in a biological variable affect RT and accuracy. Figure 5a shows how RT distributions change under different settings of FEF → striatum connection strength. Figure 5b shows quantitatively how increases in FEF → striatum connectivity leads to faster RT and decreased accuracy (Qualitative Pattern 3). Loosely, increasing FEF connection strength onto Go-units in the direct pathway leads to faster gating of responses. Conversely, increases in STN → SNr connectivity lead to slower RT and improved accuracy (see Figure 5c). The reason for both of these effects is that they differentially modulate SNr activity. Recall that the SNr tonically inhibits the SC, unless it is itself inhibited by the striatal direct pathway. Hence, any modulation of the ease with which SNr units are inhibited—either via stronger connections from cortex onto Go units, or by increasing the SNr via the STN—will change the threshold required for the BG to gate an action. Indeed, Ratcliff and Frank (2012) and Lo and Wang (2006) have shown that these two mechanisms are
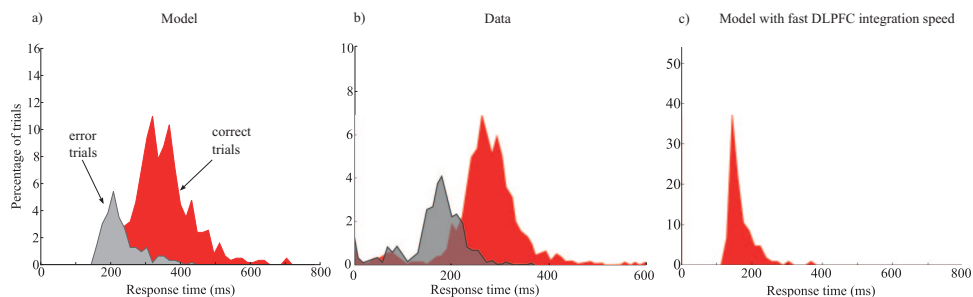

Figure 4. (a) Response time (RT) histogram for correct and erroneous incongruent trials in the model. Error RT distributions were shifted to the left due to prepotent response capture. This pattern is exaggerated with increased tonic dopamine (DA) due to lowered effective gating threshold. (b) RT histograms of a monkey during the switch-task (data based on Isoda & Hikosaka, 2008). In blocks of trials, monkeys are continuously rewarded following saccades to one of two targets. On so-called "switch-trials," a cue indicates that the monkey should perform a saccade to the opposite target, requiring the monkey to inhibit his planned saccade and perform a saccade to the opposite direction. As in the model, errors are associated with shorter reaction time. (c) Reaction time distribution of an alternative model with fast dorsolateral prefrontal cortex (DLPFC) integration speeds. Correct trials are in red, and errors are in gray (not present). This model cannot account for the behavioral pattern of errors and RTs as a function of congruency, in contrast to models with slowed DLPFC integration (Panel a).
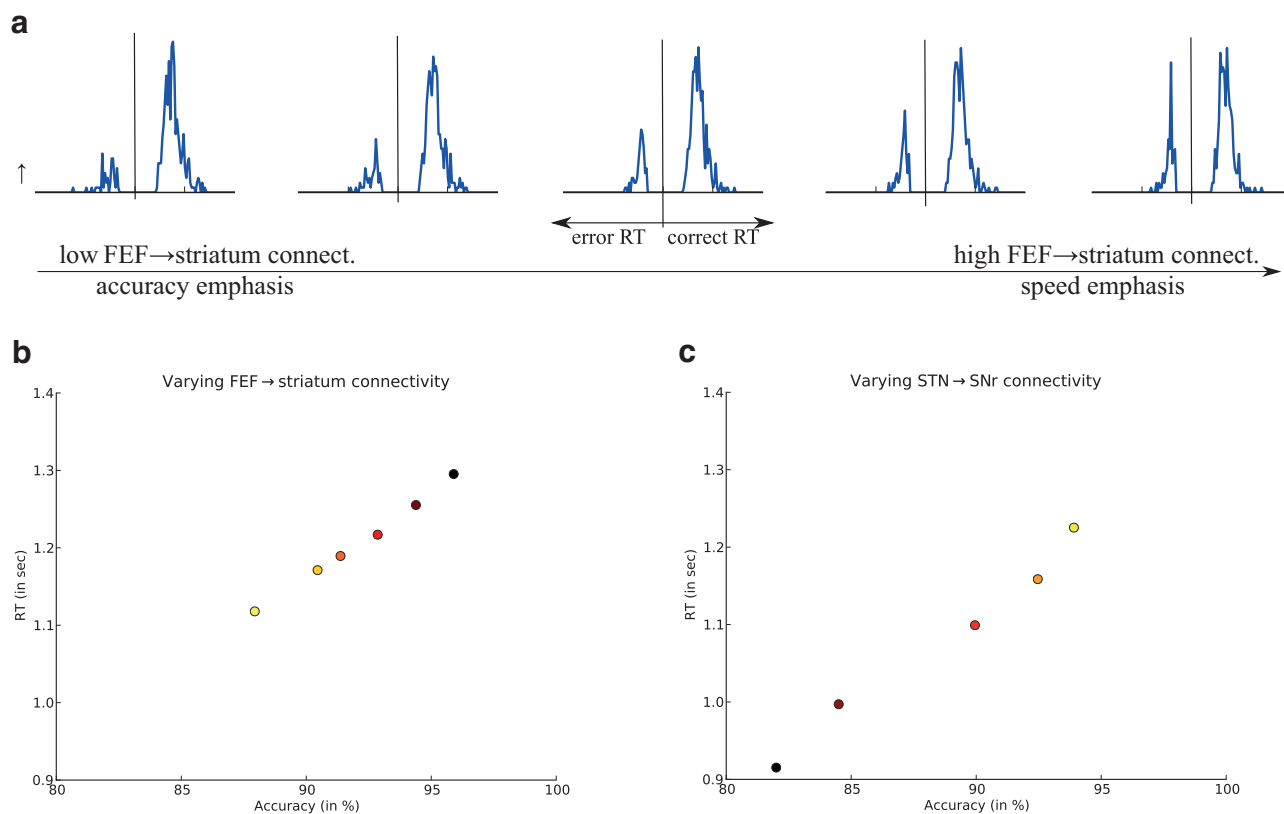
Figure 5. (a) Response time (RT) distributions for incongruent trials by network models. Frontal eye fields (FEF) → striatum projection strengths were varied along the *x*-axis. Correct RT distributions are on the right side of each panel, and incorrect RT distributions are on the left side, mirrored on the *y*-axis. This manipulation is equivalent to a speed–accuracy adjustment, as shown empirically to vary with presupplementory motor area (pre-SMA) → striatal communication (Forstmann et al., 2010, 2008), where here FEF plays the role of pre-SMA for eye movements compared to manual movements studied in Forstmann et al. (2010, 2008). (b + c) Speed–accuracy tradeoff under parametric modulation of (b) FEF → striatum connection strength and (c) subthalamic nucleus (STN) → substantia nigra pars reticulata (SNr) connection strength (color coded). Black represents low connection strength, and yellow represents high connection strength. This pattern is consistent with decision threshold modulation. The absolute values of connection strengths in these different routes are chosen to lie on a sensitive range producing observable effects for demonstration purposes.

related to changes in the decision threshold in sequential sampling models. Our model subsumes both of these mechanisms, and suggests that these different routes are themselves modulated by distinct cognitive variables, such as volitional speed-accuracy modulation and conflict/choice entropy (cortico-striatal and STN). We return to this issue in the Discussion section.

In sum, our model captures key qualitative behavioral patterns described in the literature (see above). Moreover, these patterns hold over varying biologically plausible parameter ranges leading to predictable changes in the behavioral patterns. However, given the complexity of the underlying model, it is also important to establish whether the internal dynamics of the different nodes of the network are consistent with available electrophysiological data in this class of tasks.

Neurophysiology.

DLPFC, SEF, and pre-SMA activity. Our model summarizes the computations of the executive control complex as a single layer

corresponding to DLPFC, SEF, and pre-SMA. One of our central predictions is that DLPFC activation must be delayed relative to the habitual response mechanism in order to produce the desired qualitative patterns. To demonstrate the plausibility of this account, we simulated networks with increased DLPFC speed (time constant of membrane potential updating). Consequently, networks ceased to make fast errors while correct RTs became much faster and more peaked (see Figure 4c). The reason for this pattern is that active executive control now dominates and overrides the prepotent mechanism during early processing. This result implies that some delay in executive control is needed to account for empirical findings in which incongruent RTs are delayed.

SC and FEF activity. Comparing single unit activation patterns of SC (see Figure 6a) to those of FEF (see Figure 6b) reveals that the activation dynamics are very similar between those two regions. Our model thus predicts that FEF can be interpreted as a cortical saccade planning/monitoring area that directly influences saccade generation via its projections to SC (Munoz & Everling, 2004). Moreover, SC
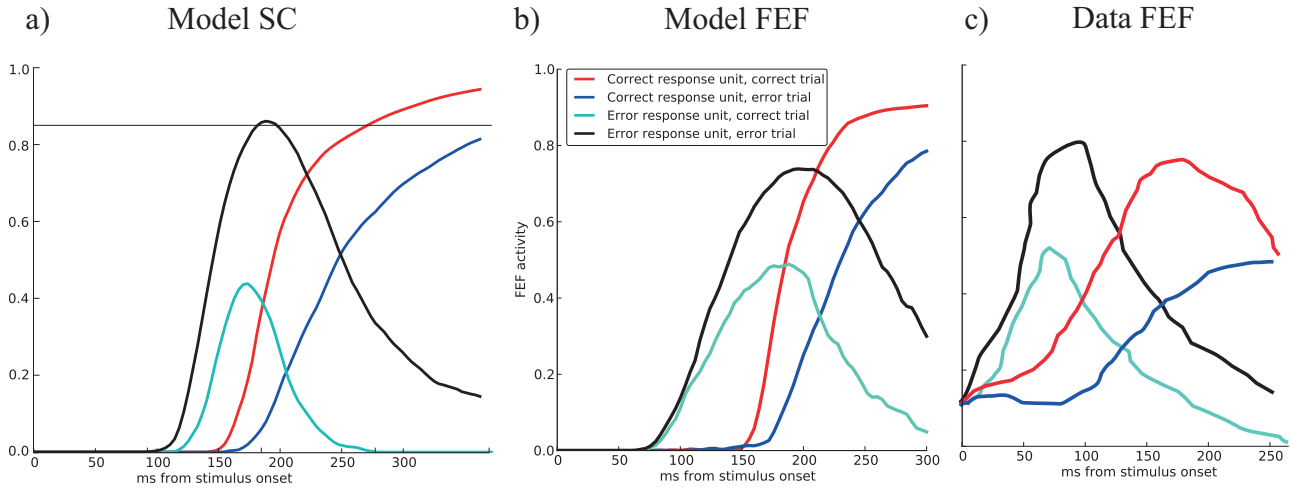
Figure 6. (a) Average activity of individual superior colliculus (SC) units coding for the correct and error response in correct and incorrect trials during incongruent trials aligned to stimulus onset. The prepotent (i.e., erroneous) response comes on before the volitional, correct response. In incorrect trials, the error-unit threshold is crossed before the volitional response unit gets active. In correct trials, the error-unit is inhibited in time. (b) Average activity of individual frontal eye fields (FEF) units coding for prepotent error responses and volitional correct responses during incongruent trials aligned to stimulus onset (Benchmark Pattern 6). (c) Electrophsyiological recordings in FEF of monkeys (data based on Everling & Munoz, 2000).

activity reveals that in both, correct and incorrect incongruent trials, the incorrect prepotent response unit becomes active before the controlled one, thus matching Qualitative Pattern 5.

dACC activity. As described earlier, the dACC computes co-activation of both response units in FEF (i.e., when average activity is > 0.5)—a direct measure of conflict (or value of the alternative action to that initially considered; see above). Conse-

quently, its activity (see Figure 7a) follows a similar pattern as average FEF layer activity: Conflict is present but resolved prior to responding in correct trials while conflict is present *after* responding in error trials. However, dACC does not get active in congruent trials, because it never shifts from one action to the other.

This qualitative pattern of peak conflict activation *before* correct incongruent trials but *after* incorrect incongruent trials matches event-
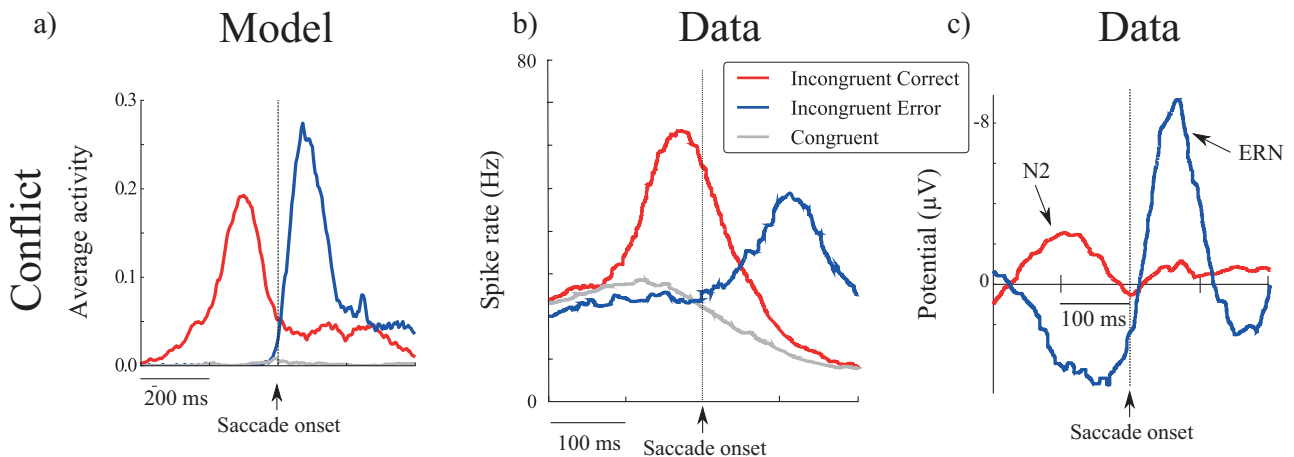
Figure 7. (a) Averaged dorsal anterior cingulate cortex (dACC) activity (corresponding to conflict in frontal eye fields [FEF]) in pro-saccade and correct and incorrect incongruent trials. No conflict is present in congruent trials. During correct incongruent trials, conflict is detected and resolved before the response is gated. During incorrect incongruent trials, an incorrect response is made before conflict is detected. (b) Activity recorded in monkey presupplementary motor area (pre-SMA) during the switch-task (data based on Isoda & Hikosaka, 2007). (c) Electroencephalography (EEG) recordings from the central scalp of humans during the Flanker task (data based on Yeung et al., 2004), thought to originate from dACC. The N2 and error related negativity (ERN) component closely match our modeling results, replicating this aspect of Yeung et al.'s (2004) model.

related potentials (ERPs) commonly observed in human electroencephalography (EEG) studies (see Figure 7c). The so-called *error related negativity* (ERN) which is measured *after* response errors, whereas the so-called *N2* potential is measured *before* correct high conflict responses (Falkenstein, Hohnsbein, Hoormann, & Blanke, 1991; Gehring, Goss, Coles, Meyer, & Donchin, 1993). The idea that these two signals could merely represent "two sides of the same conflict coin" and reflective of underlying dACC activity was first presented in the modeling work by Yeung and colleagues (Yeung et al., 2004; Yeung & Cohen, 2006).

STN activity. As noted in the model description, conflict detection in the dACC results in delayed (and more accurate) responding by recruiting the STN to prevent gating until conflict is resolved. Indeed, this mechanism is in part responsible for the rightward-shifted RT distributions in correct incongruent trials. Accordingly, this same pattern of increased activity before correct responses and increased activity after error responses can be observed in STN (see Figure 8a). Again, this qualitative pattern was also found in STN recordings in monkeys by Isoda and Hikosaka (2008; see Figure 8b), who showed that timing of STN firing relative to pre-SMA was consistent with communication along this hyperdirect pathway.

The neurocomputational model of (Brown, Bullock, & Grossberg, 2004) interprets the role of STN differently. In their model, STN is activated by the output structure (FEF in their case) to lock out the influence of competing responses *after* a response has been selected. This is a critical difference to the account presented herein where STN plays a role in the selection of a response by raising the threshold prior to response selection, thereby delaying execution but increasing accuracy. To show explicitly how our model predictions can be qualitatively differentiated from this alternative model of STN function, we disconnected dACC inputs into the STN and instead allowed only the output structure (SC in our model) to project to it, so that STN function operates as it does in Brown et al. (2004). As can be seen in Figure 8c, the activity pattern changes dramatically. Specifically, there is no more differentiation of activation patterns between the different trial types as is observed in our model and the empirical data (Isoda & Hikosaka, 2008). Because STN only influences processing after response selection, it also does not lead to delayed responding or decision threshold adjustment. This qualitative difference in model predictions is fundamental and not subject to parameter tuning, as it reflects a distinct computational role for the STN. Although we focused on the Brown model for demonstration purposes here, other models of STN function with different connectivity would similarly not account for these data. For example, the biophysical model of Rubchinsky, Kopell, and Sigvardt (2003) assumes that STN neurons provide focused selection of a particular action (by disinhibiting SNr, taking the role of the direct Go pathway) while simultaneously inhibiting competing actions (by exciting SNr in other columns). This model cannot explain this activity pattern because co-activation of multiple cortical inputs does *not* result in increased STN activity (see Figure 6b in Rubchinsky et al., 2003).

Striatal activity. Figure 9a shows striatal activity in congruent and incongruent trials (column I and column II, respectively) for direct-path Go and indirect-path NoGo units (upper and lower rows, respectively). In each case, activity selective to the correct and error responses are color coded. The model closely captures the qualitative pattern across four cell populations (Qualitative Pattern 7) identified in monkey dorsal stria-
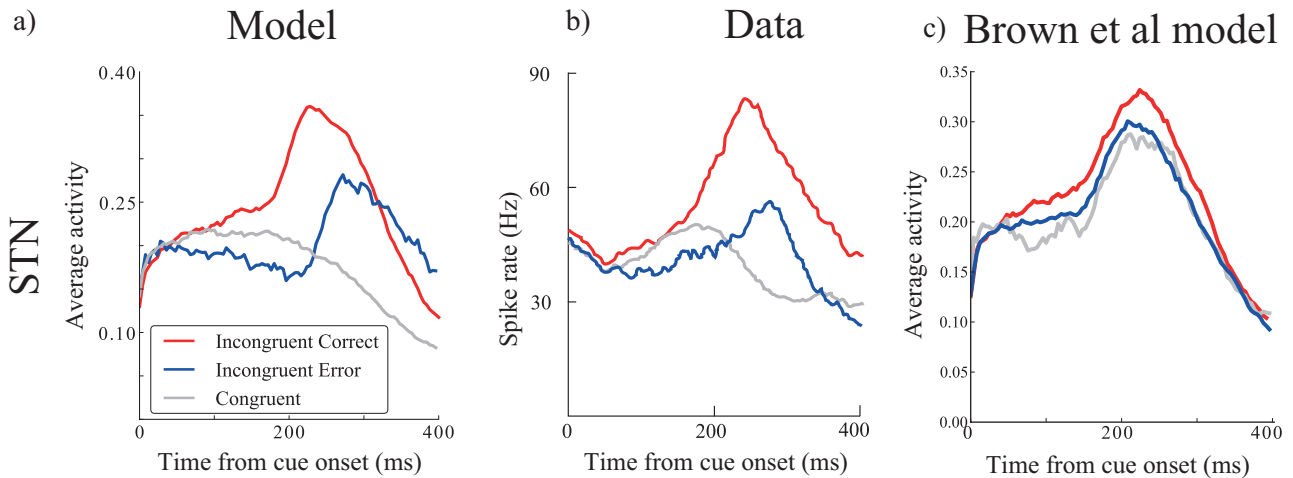
Figure 8. (a) Averaged activity of the model subthalamic nucleus (STN) layer during pro-saccade and correct and incorrect incongruent trials relative to response execution. During congruent trials, STN units exhibit a small early increase in activity that subsides. Correct incongruent trials show increased activity early on in the trial which causes the conflict-induced slowing and prevents prepotent response gating. In error trials, this mechanism is triggered too late, and the incorrect response gets executed. (b) Electrophysiological recordings of the monkey STN (data based on Isoda & Hikosaka, 2008) on correct and incorrect switch trials and non-switch trials. (c) Average activity of the STN layer of an alternative model in which STN is not excited by dorsal anterior cingulate cortex (dACC) but instead by saccadic output (superior colliculus [SC] in our model) as proposed by Brown et al. (2004). This model does not predict differences between trial types.

Figure 9. (a) Averaged striatal activity during correct pro- (first column) and incongruent trials (second column) in Go (first row) and NoGo (second row) neuronal populations. In each case, activity for correct (red) and error (blue) response (resp.) coding units are shown separately. As described in the text, the Go units for the prepotent response become active early in the trial for both trial types, but in anti-saccade trials, these are followed by NoGo units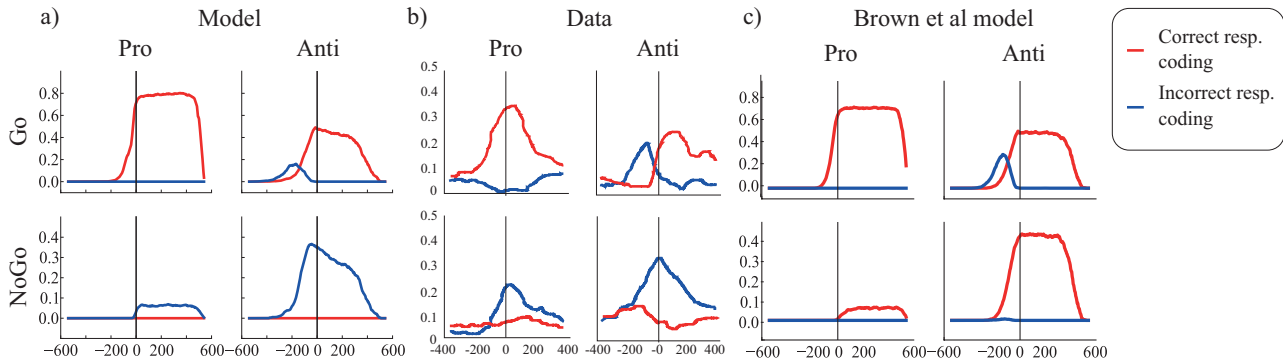 that veto the Go activity and finally Go activity for the controlled response due to top-down dorsolateral prefrontal cortex (DLPFC) activity. (b) Electrophysiological recordings of the monkey striatum (data based on Watanabe & Munoz, 2009). The first row represents neurons coding corresponding to the executed response (i.e., Go neurons), and the second row represents neurons coding that suppress execution of the corresponding action (i.e., NoGo neurons). (c) Alternative model simulating Brown et al.'s (2004) assumption that the indirect pathway acts to defer the execution of the correct response, rather than suppress the alternative response. Note that predictions for Go pathway accord with those of our model and the data, but prediction of NoGo neurons differs.

tum recordings during the anti-saccade task (see Figure 9b and Ford & Everling, 2009; Watanabe & Munoz, 2009). In particular, for congruent trials, correct-coding Go neurons gate the response while error-coding NoGo units suppress the alternative. In incongruent trials, Go neurons for the error-coding prepotent response are initially activated, but are then followed by increased activity of the corresponding NoGo population which then suppresses the initiated Go activity via NoGo → Go inhibitory projections (Taverna, Ilijic, & Surmeier, 2008). Finally, the controlled Go-correct units are activated and an incongruent response is executed. Thus, our model predicts that the pattern of electrophysiology observed in empirical recordings arises due to top-down cognitive control modulation of direct and indirect pathway neurons.

Note again that we can distinguish our model's predictions from those of other models that omit the indirect pathway as a distinct source of computation (there are several) or from models that do include it but assign a different function. The neural network model of Brown et al. (2004) assumes the indirect pathway activation defers execution of the correct action plan until the time is appropriate. This would suggest that the executive control complex would activate NoGo units coding for the *correct* response, not the incorrect response as in our model. To demonstrate how this leads to qualitatively different patterns than is observed in our model and the data (see Pattern 7 and Figure 9c) in which this alternative account is simulated in our model. (However, we note that Brown et al.'s, 2004, model could potentially accord with our model in the sense that they also advocate a mechanism by which negative prediction errors drive learning in the NoGo cells, which after training on the AST may also produce the patterns we observe here given that the prepotent response would be punished.) Similarly, the prominent model of Gurney, Prescott, and Redgrave (2001) suggests that this pathway serves as a control pathway rather

than providing negative evidence against particular actions as in our model, and it is unclear how this control function (while not disputed per se) would reproduce the patterns observed here.

Global Response Inhibition

Method. In SRITs, the selectively inhibited prepotent response must be replaced with another, controlled response. Conversely, the stop-signal task (SST) requires outright response inhibition (e.g., Aron & Poldrack, 2006; Cohen & Poldrack, 2008; Logan & Cowan, 1984) and is used to assess global inhibitory control (Aron, 2011). Specifically, subjects are required to press left and right keys in response to Go-cues appearing on a screen. On a subset of trials *after* the Go-cue has been presented, a stop-signal is presented after variable delay (i.e., stop-signal delay; SSD) instructing the subject to withhold responding.

Here, we show that our model can also simulate the SST after we included the right inferior frontal cortex (rIFG) with direct projections to STN (Aron, Behrens, et al., 2007; see Figure 10). Given the assumptions of the race model (i.e., a race between Go and Stop processes), one can estimate the stop-signal reaction time (SSRT) by measuring the probability of successful inhibition at different SSDs. This inhibition function is then compared to the distribution of Go reaction times in non-stop signal trials. There are several extensive reviews of the SST (Verbruggen & Logan, 2009), so here we focus on how our model captures the available evidence. Note that the SST typically refers to the task involving manual movements (and inhibition thereof), but a well-studied equivalent has been used in the oculomotor domain, where it is referred to as the *countermanding task*. While the neuronal circuitry involved in Go-responding depends on the response modality, the neuronal circuitry involved in the global mechanism may be independent of the response modality (Leung & Cai, 2007).
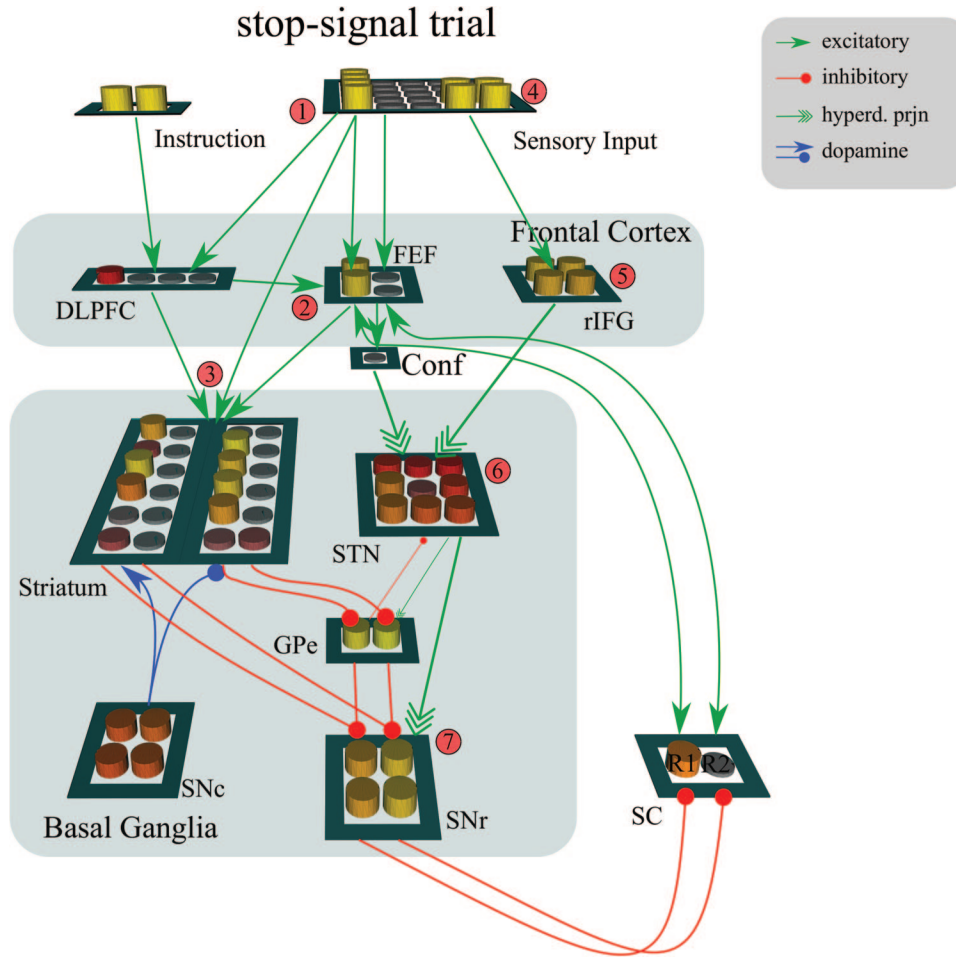
stop-signal trial

Figure 10. Extended neural network model including right inferior frontal gyrus (rIFG) during stop-signal trials. (1) Left input stimulus activates (2) left-coding frontal eye fields (FEF) response units and (3) initiates gating via striatum (similar to pro-saccade trial in a). After a delay, (4) the stop-signal is presented which activates (5) rIFG, which in turn (6) transiently activates the subthalamic nucleus (STN) and finally (7) the whole substantia nigra pars reticulata (SNr) to globally prevent gating. Note that dorsolateral prefrontal cortex (DLPFC) is beginning to get active to initiate selective response inhibition via striatal NoGo units. hyperd. prjn = hyperdirect projection; Conf = conflict; GPe = external segment of the globus pallidus; SNc = substantia nigra pars compacta; SC = superior colliculus; R1 = Response 1; R2 = Response 2.

Networks are presented with one of two input stimuli (left or right), represented by a column of four units each. As in prior simulations, prepotent responses are implemented by weights from the input units to the corresponding FEF response units, such that a left stimulus suggests a left response. On 25% of trials, a stop-signal is presented with variable delay (by activating devoted units in the sensory input layer). The stop signal units send excitatory projections directly to the rIFG layer. rIFG units in the hyperdirect pathway excite the STN (Aron, Behrens, et al., 2007; Neubert, Mars, Buch, Olivier, & Rushworth, 2010) and prevent striatal response gating, and therefore inhibit responding if the SC has not already surpassed threshold. In addition to this global rIFG-STN response suppression mechanism, the DLPFC combines the stop-signal input and the stimulus location to selectively inhibit the associated response via activation of the corresponding population of striatal NoGo

units. Critically, this selective mechanism is slower but remains active after the STN returned to baseline and prevents subsequent responding. Thus, the model uses a fast, global but transient response inhibition mechanism and a slower, selective but lasting mechanism (Aron, 2011). To estimate the SSRT, we use the dynamic one-up/one-down *staircase procedure* for adjusting the SSD (e.g., Logan, Schachar, & Tannock, 1997; Osman, Kornblum, & Meyer, 1986).

We tested the influence of rIFG lesions on the SSRT (Aron, Monsell, Sahakian, & Robbins, 2004) by parametrically reducing the projection strength of rIFG to the STN.

The selective norepinephrine (NE) reuptake inhibitor Atomoxetine increases NE release and improves stop-signal performance in animals, healthy adults, and adult ADHD patients (Chamberlain et al., 2007, 2009). NE is hypothesized to adaptively change the activation gain of neurons in frontal cortex (Aston-Jones & Cohen,

2005). We consequently tested the influence of decreasing the gain parameter in units of the frontal cortex.[2]

Finally, we simulated different motivational influences on stop-signal accuracy. Evidence for the neural underpinnings of motivational biases comes from an fMRI study by Leotti and Wager (2010), who reported that subjects instructed to focus on speed instead of accuracy exhibited a greater increase in activations in brain regions associated with response facilitation, including the FEF and the striatum. Conversely, when instructed to focus on accuracy, subjects exhibited greater activity in IFG regions associated with response inhibition. We thus simulated these activation patterns to account for speed–accuracy tradeoff in a similar manner as in the anti-saccade simulations. In the speed-condition, we manipulated the strength of FEF to striatum connections due to evidence that frontostriatal connectivity is enhanced under speed emphasis (Forstmann et al., 2010, 2008; Mansfield, Karayanidis, Jamadar, Heathcote, & Forstmann, 2011). Conversely, in the accuracy condition we increased baseline excitatory input to rIFG, allowing it to be more excitable and hence facilitating STN recruitment. This simulation approximates the effect of a putative PFC rule based representation to focus on accuracy. Recent data supports the notion that the (right) STN, which receives input from rIFG, shows increased excitability associated with an increased response caution during accuracy focus (Mansfield et al., 2011).

Results. As with the SRITs above, we extracted a list of key qualitative results from the literature that we use to evaluate the fit of our model.

1. The probability of inhibiting a response decreases monotonically as SSD increases (Verbruggen & Logan, 2008).

2. Error responses that escape inhibition are, on average, faster than Go responses on no-stop-signal trials. However, while the distributions begin at the same minimum value, the responses that escape inhibition have a shorter maximum value (Verbruggen & Logan, 2008).

3. STN neurons are excited to stop signals but show little differentiation between stop-signal inhibition and stop-respond error trials (Aron, Behrens, et al., 2007). Contrary, downstream SNr neurons are excited in correct trials but are disinhibited during error trials (Schmidt, Leventhal, Pettibone, Case, & Berke, 2012).

4. SEF neurons are activated in stop-signal and stop-response trials *after* SSRT and can thus not contribute to successful stopping (Stuphorn, Taylor, & Schall, 2000).

Behavior. To illustrate the staircase procedure, Figure 11a shows an example trace of how SSDs are adjusted to assess 50% stop-signal accuracy. As can be seen, the network with rIFG lesion is impaired at stopping and requires shorter SSD on average to inhibit successfully.

As can be seen in Figure 11b, the inhibition function resulting from testing the neural network systematically with different SSDs reveals a monotonically decreasing probability of correctly stopping (Qualitative Pattern 1).

Cumulative RT distributions of Go and non-canceled Stop trials are presented in Figure 12. Both distributions match closely up until SSD + SSRT (Qualitative Pattern 2), suggesting that both are generated by the same process.

Different modulations affect GoRT and SSRT in different ways (see Figures 13a and 13b). While DA manipulations certainly speed GoRT, SSRT remains largely unaffected. On the other hand, when the network is tested with reduced gain (simulating low NE levels), or has lesions to either STN or rIFG, it exhibits SSRT deficits (increases). Finally, simulated accuracy emphasis results in slowed Go RT but faster SSRT (more effective inhibition). The pattern that emerges from these results is that SSRT is changed by modulations of parameters that are part of the global inhibitory pathway: rIFG and STN.

Neurophysiology. To assess the neural correlates of stopping behavior in our model, we analyzed STN and SNr activity aligned to stop-signal onset. As can be seen in Figure 14, there is little differentiation between stop-signal inhibition and error trials while SNr units show a marked dip in error trials that is less pronounced in inhibition trials (Qualitative Pattern 3).

We moreover analyzed the activity pattern of our executive control complex which consists of DLPFC, SEF, and pre-SMA. As can be seen in Figure 14, activation is observed in stop-signal trials (both stop-respond and successful inhibitions) only *after* SSRT and could thus had no influence on the stopping (Qualitative Pattern 3). This result implies that global stopping to salient stop signals is most likely driven by the fast stop process along the rIFG-STN hyperdirect pathway. We ascertain that executive control processes are delayed relative to this global stopping mechanism, and may participate in selective response inhibition (and in the stop-change task, activation of the correct response) after the global response pause has passed.

Discussion

The interaction between executive control and habitual behavior is a central feature of higher-level brain function, and plays a role in various domains from cognitive psychology (under the rubric of "System 1" vs. "System 2"; Evans, 2003) to machine learning (model-free vs. model based control; Daw, Niv, & Dayan, 2005). At the core of this interaction is a mechanism that allows executive control to override the habitual system and guide action selection. A multitude of psychological cognitive tasks have been used to probe the nature of this interaction. The stop-signal task requires outright stopping of a response already in the planning stage. The anti-saccade (Hallett, 1978), Simon (Simon, 1969), and saccade override (Isoda & Hikosaka, 2007, 2008) tasks all involve inhibition of a prepotent action together with initiation of an action incompatible with the prepotent one. Despite the apparent behavioral simplicity of these tasks, various lines of research have revealed a highly complex and tightly interconnected brain network underlying response inhibition consisting of frontal areas including DLPFC, SEF, pre-SMA, FEF, rIFG, and dACC and basal ganglia structures including the striatum and STN.

We presented a dynamic neural network model of selective and global response inhibition which provides a description of the distributed computations carried out by individual brain regions and neurotransmitters. The complexity of this model is grounded by well-established neuroanatomical and physiological considerations, and accounts for a wealth of key data including electrophysiology, psychiatric and pharmacological modulations, behav-

---

[2] Gain modulates how step-like the activation-dynamics of units are in relation to their input activity. Low gain leads to linear activation dynamics, while high gain levels make a unit respond in a binary-like fashion.
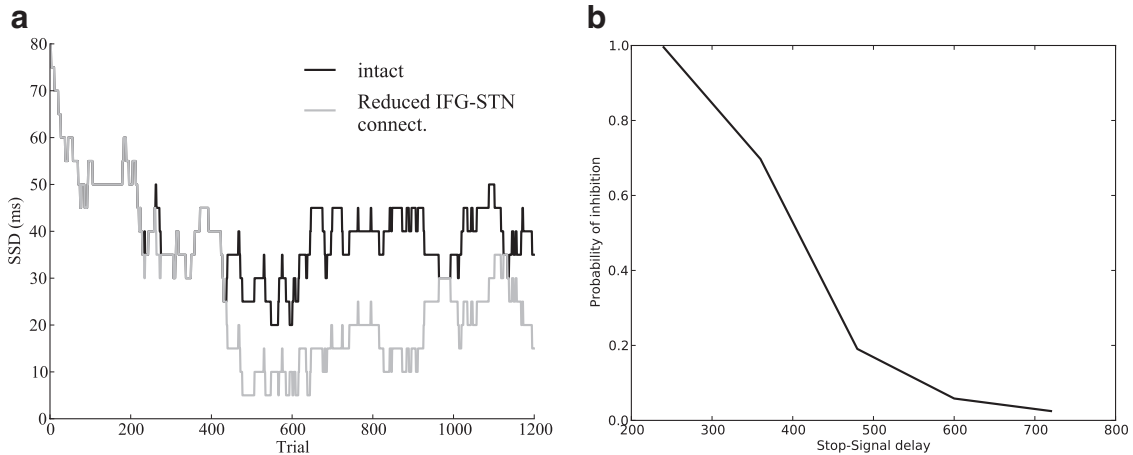
Figure 11. (a) Progression of the staircase procedure for manipulating stop-signal delay (SSD) in networks with reduced right inferior frontal gyrus (rIFG)–subthalamic nucleus (STN) connectivity (connect.). Trial number is plotted on the *x*-axis, and the SSD in milliseconds (converted from simulator time) is plotted on the *y*-axis. If a response is successfully inhibited on stop-signal trial, the SSD is increased by 20 ms to make it harder. If a response is erroneously made on a stop-signal trial, the SSD is decreased by 20 ms. Networks without lesion are highest in general representing the most effective stop-process that is able to withhold responses even when the SSD is quite long. (b) Inhibition function of the neural network model in the stop-signal task. The model is tested on systematically varying levels of SSD in milliseconds, and the proportion of correctly inhibited trials is plotted along the *y*-axis.

ioral, lesion, and imaging studies. Moreover, this model is constrained (i) by using a single parameterization across all simulations of intact function and (ii) by the multitude of qualitative results from different levels of analysis it is required to reproduce. Although we used one parameterization across the intact model simulations, we also generalized the functionality via systematic manipulations across a range of parameters. In other work (Wiecki

& Frank, 2010a), we have shown that the emerging fundamental computational properties of this complex system as a whole are captured by analysis using a modified drift diffusion model, in which distinct mechanisms within the neural model (e.g., STN projection strength, DLPFC speed) are monotonically related to high level decision parameters (e.g., decision threshold, and drift rate of the executive process).
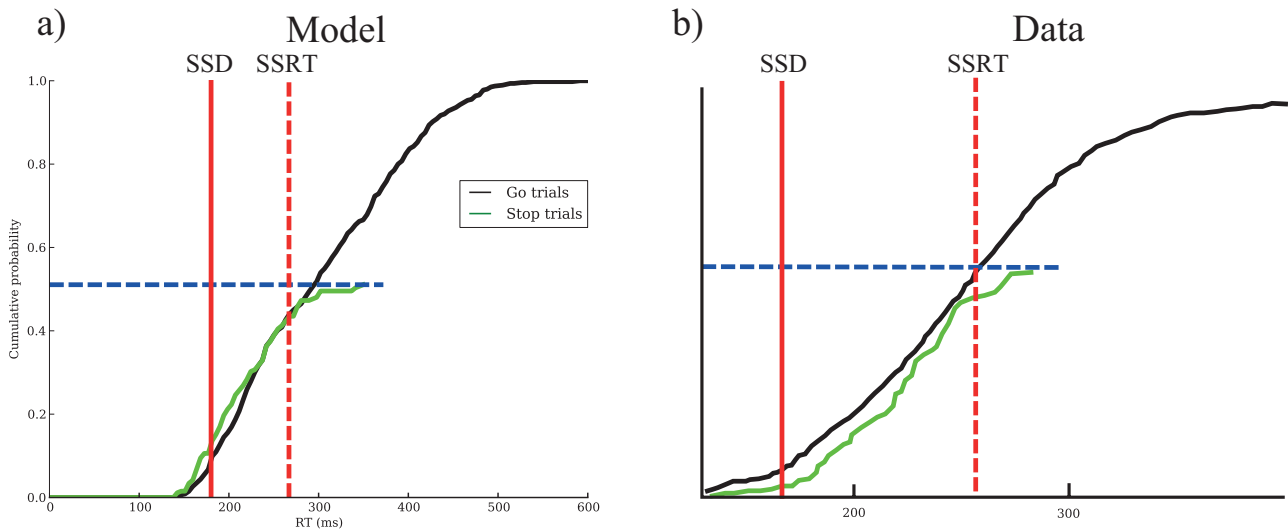
Figure 12. (a) Cumulative reaction time (RT) distributions of the neural network model and from a monkey experiment. (b) Cumulative RT distribution from a monkey experiment for comparison (data based on Lo et al., 2009). The solid red line denotes mean stop-signal delay (SSD); the broken red line denotes stop-signal reaction time (SSRT) offset at SSD. The broken blue horizontal line represents 50% stopping accuracy. Note that the response distribution sums to the response probability—not necessarily to 1.
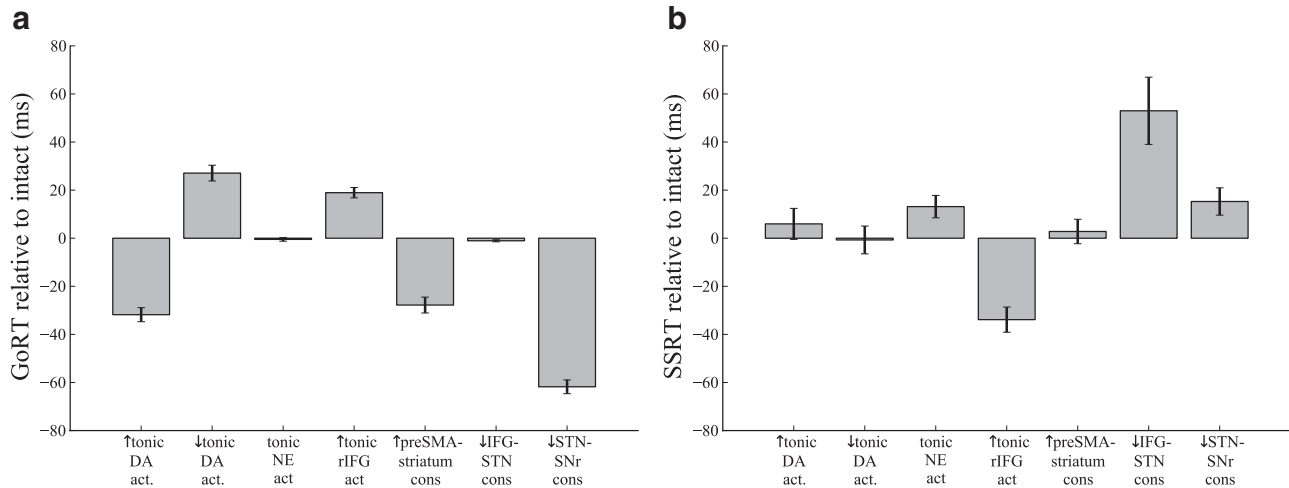
Figure 13. (a) Mean response times (RTs) in milliseconds ± *SEM* (converted from simulator time) for Go trials under different modulations (see text). (b) Mean stop-signal reaction time (SSRTs) in milliseconds ± *SEM* (converted from simulator time) under different modulations (see text). DA = dopamine; act = activity; NE = norepinephrine; rIFG = right inferior frontal gyrus; pre-SMA = presupplementory motor area; IFG = inferior frontal gyrus; STN = subthalamic nucleus; SNr = substantia nigra pars reticulata; act = activity; cons = connections.

Selective Response Inhibition

In our SRIT simulations, the model assumes that prepotent, reflexive actions such as a saccade to an appearing stimulus (e.g., a pro-saccade) are selected via the cortico-cortical route and swiftly gated by the BG. An abundance of data supports the general involvement of the BG in saccade generation and inhibition (e.g., Hikosaka, 1989; Hikosaka et al., 2000; Hikosaka & Wurtz, 1986). Conversely, the cognitive control system not only represents the task rules needed to respond appropriately (e.g., in DLPFC), but also incorporates a downstream mechanism in dACC-STN to detect when these rules indicate an alternative action than was originally initiated. Thus our model synthesizes the popular account of dACC in terms of response conflict (Botvinick, Braver, Barch, Carter, & Cohen, 2001) with recent studies

suggesting that dACC rather reflects the value of the alternative action (Kolling et al., 2012). Moreover, via the hyperdirect pathway to the STN, this mechanism serves to transiently increase the BG gating threshold to prevent prepotent actions from being facilitated and allows more time for the controlled PFC-striatal mechanisms to selectively suppress this response and to facilitate appropriate alternative courses of action. It has also been shown that the SEF, FEF (Munoz & Everling, 2004), dACC (Botvinick et al., 2004), pre-SMA (Isoda & Hikosaka, 2007), and STN (Isoda & Hikosaka, 2008) are involved in detecting conflict between a planned response and the current rule, and for switching from an automatic to a volitional response (e.g., anti-saccades).

To detect conflict between reflexive and controlled responses, the system needs to be able to compute the correct identity of the
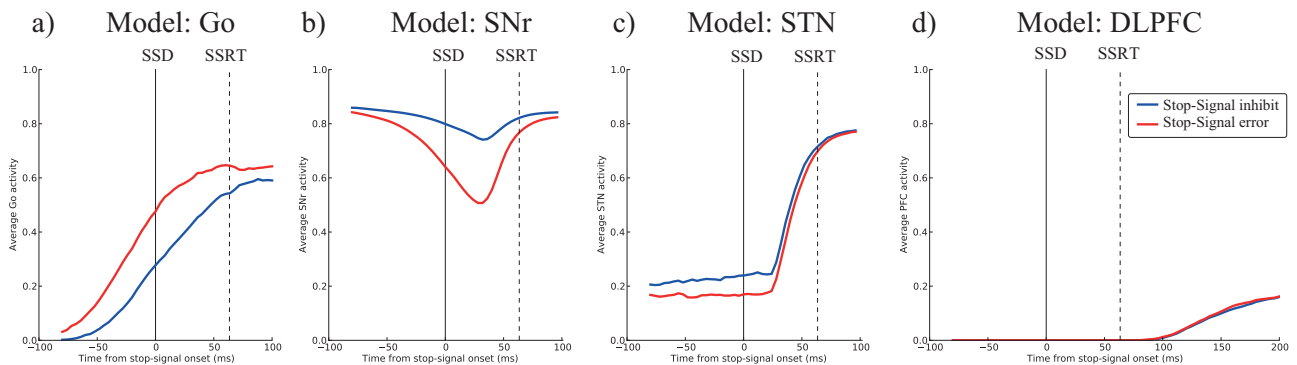
Figure 14. Average activity aligned to stop-signal onset for inhibited and error stop-signal trials. (a) Striatal Go-neuronal activity. (b) Substantia nigra pars reticulata (SNr) activity. (c) Subthalamic nucleus (STN) activity. (d) Activity of the executive control complex consisting of dorsolateral prefrontal cortex (DLPFC), supplementary eye fields (SEF), and presupplementary motor area (pre-SMA). SSD = stop-signal delay; SSRT = stop-signal reaction time; PFC = prefrontal cortex.

controlled response itself. In the model, the DLPFC integrates task instructions and current stimulus location and forms a conjunctive rule representation (Bunge & Wallis, 2008; Wallis & Miller, 2003) that then provides evidence for the associated controlled response via its projection to the FEF, and further biases the gating of this response (and the selective suppression of the reflexive response) via striatum. We demonstrated that this is a necessary condition for our model by showing that a model with faster integration speeds fails to account for key behavioral patterns.

Thus, it should be clear that compared to a congruent response, an incongruent response should (i) be more prone to error because it depends on successful inhibition of prepotent actions which may be close to threshold by the time conflict is detected and (ii) take longer due to (iia) additional computation needed for the DLPFC to perform the requisite vector inversion (activation of correct rule representation among multiple competitors based on an integration of input and instruction), and (iib) the delay in commitment to a response resulting from the increase in decision threshold along the hyperdirect pathway.

Early cognitive models of interference control assumed a dual-route mechanism for action selection, including an automatic response route and a volitional one (Eimer, 1995; Kornblum, Hasbroucq, & Osman, 1990; Ridderinkhof, 2002; D. Wang, Kristjansson, & Nakayama, 2005). This model was extended to include selective suppression of the automatic response by the volitional response mechanism (i.e., the activation-suppression model; Ridderinkhof, 2002; Ridderinkhof et al., 2011). Our model shares these attributes but makes two crucial contributions to this discussion: (i) strong predictions on the neural correlates of these abstract cognitive processes, and (ii) a raise in decision threshold requiring more evidence to gate *any* response. This latter mechanism may not only be adaptive as a fast route to prevent gating of prepotent actions, but could also serve to increase the likelihood that the alternative action selected is the most accurate (particularly when there may be more than one, as is often the case in more realistic executive control scenarios than those typically studied in simple response inhibition tasks).

Response time distributions and errors: Neural underpinnings. At the behavioral level, our intact model reproduces the same patterns found empirically—networks made more errors (see Figure 3a) and were in general slower (see Figure 3b) on incongruent trials compared to congruent trials (e.g., Harris et al., 2006; McDowell et al., 2002; Reilly et al., 2006, 2007). Incongruent errors were more likely to occur when networks responded fast (see Figures 4 and 5a as well as Ridderinkhof et al., 2011). These errors result primarily from variance in the speed of cognitive control (DLPFC), but also in the prepotent response (in some trials gating is faster than others) and in the inhibition process (in some trials the hyperdirect pathway and/or striatal NoGo process is slower). Moreover, reduced DLPFC connectivity also degrades accuracy on incongruent trials, mirroring the empirical performance degradation in anti-saccade tasks during development associated with reduced DLPFC connectivity (Hwang, Velanova, & Luna, 2010). A more explicit investigation into the dynamics of these processes comes from the simulated electrophysiology across brain regions and trial types.

Conflict- and error-related activity: Relation to existing models. The Error Related Negativity (ERN) is an event-related potential associated with errors made in forced-choice reaction time tasks (Falkenstein et al., 1991; Gehring et al., 1993). The ERN reaches its peak within 100 ms *after* the erroneous response. Using a connectionist model, Yeung and colleagues hypothesize the ERN to reflect conflict between the executed, erroneous response and the still-evolving activation of the correct response (Yeung et al., 2004; Yeung & Cohen, 2006). Thus, the error detection mechanism reflects an internal correction of the executed response, leading to a transient period of response conflict. According to this same framework, a similar potential should be observed in high conflict trials *before* correct responses, when conflict is resolved prior to responding. These authors indeed reported that the N2 potential exhibited just this profile and argued that it reflected the same underlying conflict mechanism in the dACC.

Our dACC node exhibits the same qualitative pattern of increased activity (i) before correct incongruent responses, (ii) after incorrect incongruent responses and (iii) baseline activity during congruent responses. However, this pattern is not unique to ERPs thought to originate from dACC, but is also found in electrophysiological recordings in pre-SMA, SEF (Emeric, Leslie, Pouget, & Schall, 2010), and STN (Isoda & Hikosaka, 2008). Our model provides an explicit framework that recapitulates these effects and explores their influences on behavior. Together, these dynamics accord with our earlier assertion that our model synthesizes the conflict model with the notion that the dACC reflects the value of the alternative action: This network only becomes activated when the alternative action is deemed to be more correct than the prepotent one. This process occurs either prior or following response execution (as in the conflict monitoring account), but must always occur after the initial activation of an incorrect (often prepotent) response (not specified by the conflict account but consistent with the alternative action value account).

Global Response Inhibition

By adding a single rIFG layer to our model, we generalized our model to capture data from global response inhibition tasks such as the SST. As we demonstrated above, this model recovers key qualitative behavioral patterns reported in the literature. Moreover, model neurophysiology revealed interesting similarities to recent rat electrophysiological recordings in the SST (Schmidt et al., 2012). Specifically, while STN activity surges in response to the stop signal to the same extent regardless of whether the response is successfully inhibited or not, activity in the SNr strongly differentiates between these trial types. During errors, the striatal Go signals were potent and early enough to inhibit SNr activity in spite of the STN surge. These results suggest that the source of response inhibition errors is variance in the Go process, but that the duration of the stop-process is rather fixed. This conceptualization matches closely with the interactive horse-race model (Verbruggen & Logan, 2009). Here, we hypothesize that the critical point of interaction between the two processes is the SNr.

Why did we add an rIFG layer given that our initial model already contained an executive control complex including DLPFC? As described above, rIFG and STN involvement in the SST is well established, and moreover, simulations showed that the activations in our executive control complex needed to account for SRITs was too slow to account for global response inhibition needed in SST. Nevertheless, the nature of the (potentially sepa-

rable) mechanisms engaged for detecting when inhibitory control is necessary, and how it should be implemented, remains largely elusive. In particular, the role of rIFG is actively debated. Some studies specifically implicate the rIFG in response inhibition (Aron et al., 2003; Chambers et al., 2007; Sakagami et al., 2001; Verbruggen et al., 2010; Xue et al., 2008), whereas others report rIFG activity in tasks lacking pure response inhibition demands, suggesting that it is more involved in monitoring or salience detection (Chatham et al., 2012; Fleming, Thomas, & Dolan, 2010; Hampshire, Chamberlain, Monti, Duncan, & Owen, 2010; Munakata et al., 2011; Sharp et al., 2010; Verbruggen et al., 2010). Our model unifies these two seemingly opposing views by arguing that the rIFG in fact detects salient events and, via downstream processing, engages a stopping mechanism whether or not it is required by the task rules. In both the stop-signal and stop-change task, subjects have to detect an infrequent signal which tells them to update their current action plan. We argue that these signals are salient events and, via noradrenergic modulation, enhance processing in the rIFG which, in turn, causes an *orienting* or *circuit breaker* response by activating the STN (Swann et al., 2011) to pause response selection. This pause enables the volitional DLPFC based response selection mechanism to take control and either inhibit a specific response (as in the stop-signal task) or initiate a new response (as in the stop-change task). This theory of a rIFG triggering a global response-pause is supported by rIFG involvement in the oddball task (Huettel & McCarthy, 2004; Stevens, Skudlarski, Gatenby, & Gore, 2000) which requires no behavior adaptation whatsoever, yet still causes response slowing (Barcelo, Escera, Corral, & Periáñez, 2006; Parmentier, Elford, Escera, Andrés, & San Miguel, 2008). Indeed, in many of the above-reported studies in which rIFG is activated under conditions of monitoring or saliency, when they have been reported, subject RTs were nevertheless delayed despite no overt inhibitory demands (Chatham et al., 2012; Fleming et al., 2010; Sharp et al., 2010).

Different Forms of Response Inhibition

Inhibitory control can be issued globally or selectively (Aron, 2011; Aron & Verbruggen, 2008). The brain seems to revert to a global inhibitory mechanism when unexpected events occur that require quick response adaptation (e.g., stop-signals), and to a selective inhibitory control mechanism when response inhibition can be prepared (Greenhouse, Oldenkamp, Aron, & Diego, 2012; Hu & Li, 2011). We propose that selective inhibition of the prepotent response is initiated by the DLPFC and implemented via the indirect corticostriatal NoGo pathway (Hu & Li, 2011; Jahfari et al., 2011; Watanabe & Munoz, 2009, 2010; Zandbelt & Vink, 2010). Global response inhibition on the other hand is driven by a salience detection mechanism implemented in the rIFG which directly projects to the STN to inhibit responding (Aron, Durston, et al., 2007; Eagle et al., 2008; Isoda & Hikosaka, 2008; Jahfari et al., 2012, 2011; Kühn et al., 2004; Mink, 1996; Nambu et al., 2000; Nambu, Tokuno, & Takada, 2002).

In addition to the selectivity of inhibitory control, differences exist between proactive and reactive initiation of response inhibition (Aron, 2011; Cai, Oldenkamp, & Aron, 2011; Greenhouse et al., 2012; Swann et al., 2012). Our modeling work suggests multiple possible sources for proactive control. Speed–accuracy adjustments are implemented by increasing functional connectivity between frontal motor regions and striatum to decrease the decision threshold under speed emphasis (see Figures 5b and 5c as well as Forstmann et al., 2010; Lo & Wang, 2006). The second proactive mechanism increases response caution by increasing baseline rIFG activity to prime saliency detection and slow responding via the rIFG-STN hyperdirect pathway (see Figure 13b). Interestingly, while FEF → striatum functional connectivity influence speed and accuracy in our SRIT simulations, SSRT in the stop-signal task is unaffected by this modulations and is only improved by an increase in tonic rIFG activity. This suggests that proactive control in form of mere response slowing is ineffective in reducing SSRT—the staircase procedure adapts to slower overall responding—but that enhanced attentional monitoring has preferential influence on global inhibitory control. In other words, although all these mechanisms can lead to adjustments in decision threshold, only those associated with active engagement of the stop process will facilitate inhibitory control per se. If confirmed, this result may have implications for refining therapy of inhibitory control disorders like addiction, obesity and OCD. Nevertheless, it remains important to emphasize that the striatal NoGo pathway is also thought to help to prevent the proactive selection of maladaptive responses.

Multiple Mechanisms of Response Threshold Regulation in Fronto-Basal-Ganglia Circuitry at Different Time Scales

Different mechanisms in our neural network influence the gating threshold for initiating motor responses at distinct time scales, and modulated by distinct cognitive variables. First, the strength of cortico-striatal projections regulate the ease with which cortical motor plans can be gated by the BG, allowing for speed emphasis in the speed-accuracy tradeoff (see Figure 5c). This aspect of our model is quite similar to the model of Lo and Wang (2006) and was subsequently corroborated by Forstmann et al. (2010). Our model converges on the same conclusion but extends this view by showing that gating threshold is also more dynamically regulated on a shorter time-scale by (i) motivational state (changes in DA levels, which are modulated by reinforcement and also facilitate striatal Go signals); and (ii) response conflict and saliency (via the hyperdirect pathway, making it more difficult or Go signals to drive BG gating; Jahfari et al., 2011). Moreover, STN efficacy in the neural model is positively correlated with increases in estimated decision threshold (Ratcliff & Frank, 2012). Evidence for conflict-induced decision threshold adjustment via the hyperdirect pathway has been recently described in a reinforcement-based decision making task (Cavanagh et al., 2011). Increases in frontal EEG activity during high conflict decisions were related to increases in decision threshold estimated by the drift diffusion model. Intracranial recordings directly within the STN also revealed decision conflict-related activity during the same time period and frequency range as observed over frontal electrodes (see also Zaghloul et al., 2012). Moreover, disruption of STN function with deep brain stimulation led to a reversal of the relationship between frontal EEG and decision threshold, without altering frontal activity itself. These data thus support the notion that frontal-STN communication is involved in decision threshold adjustment as a function of conflict. Similarly, proactive preparation to increase decision threshold in the stop signal task when stop

signals are likely is associated with hyperdirect pathway activity (Jahfari et al., 2012).

In our neural models, conflict-related STN activity subsides with time (see Figure 8), due to resolution of conflict in FEF/ACC, feedback inhibition from GPe, and neural accommodation. Thus a more refined description of this transient STN surge is that it initially increases the decision threshold (more so with conflict), followed by a dynamic collapse of the decision threshold over time. Indeed, a recent multilevel computational modeling and behavioral study by Ratcliff and Frank (2012) supported this idea by showing that a collapsing threshold diffusion model provided good fits to both the BG model and to human participant choices in a reinforcement conflict task. Moreover, the temporal profile of the best-fitting collapsing threshold corresponded well to the time course of the collapse in STN activity across time.

Psychiatric Disorders and Differential Effects of Dopamine and Norepinephrine

Abnormal striatal DA signaling is hypothesized to be at the core of many disorders, including PD (Bernheimer, Birkmayer, Horny-kiewicz, Jellinger, & Seitelberger, 1973), SZ (Breier et al., 1998), and ADHD (Casey, Nigg, & Durston, 2007; Frank, Santamaria, O'Reilly, & Willcutt, 2007). Intriguingly, all of these disorders are linked to response inhibition deficits in the stop-signal task. Our earlier BG models have successfully accounted for a wide variety of findings associated with striatal DA manipulations across reinforcement learning and working memory tasks (for review, see Wiecki & Frank, 2010b). Yet, we found here that striatal DA manipulations, while affecting overall RT, had negligible effects on response inhibition deficits as assessed by SSRT (see Figure 13b). This prediction converges with recent evidence (reviewed in Munakata et al., 2011) showing that levodopa, a drug that increases DA levels in striatum (Harden & Grace, 1995), had no influence on SSRT in PD patients (Obeso, Wilkinson, Casabona, et al., 2011; Obeso, Wilkinson, & Jahanshahi, 2011).

This lack of DA effect raises the question of the source of the response inhibition deficits in the aforementioned disorders. One conspicuous candidate is abnormal NE functioning as suggested by evidence in both ADHD (Faraone et al., 2005; Frank, Scheres, & Sherman, 2007; Ramos & Arnsten, 2007) and PD (Farley et al., 1978). In our simulations, NE modulation influences SSRT via its gain-modulatory effects in rIFG (Aston-Jones & Cohen, 2005). Additional support for this account comes from pharmacological experiments using the selective norepinephrine reuptake inhibitor atomoxetine, which improves response inhibition performance in animals, healthy adults, and ADHD patients (Chamberlain et al., 2007, 2009). Moreover, fMRI analysis revealed that atomoxetine exerted its beneficial effects via modulation of rIFG (Chamberlain et al., 2009), providing additional support for the model mechanisms. Finally, this highlights an alternative source for response inhibition deficits observed in PD patients previously linked to DA dysfunction (see Vazey & Aston-Jones, 2012, for a review high-lighting the importance of aberrant NE signaling in cognitive deficits of PD patients).

Limitations

Despite our model's success in reproducing and explaining a wide array of data and offering potential solutions for long stand-ing issues in the field, we certainly acknowledge that there are many errors of omission and—although we did not include any biological features that are unsupported by data—perhaps some errors of commission. We however note that most of our assumptions and simulations are largely orthogonal to each other. Thus, each aspect of the model is falsifiable on its own, without necessarily falsifying other aspects. We discuss a few salient limitations below; it is by no means exhaustive.

Specificity of PFC Regions and Function

While the BG of our neural network model is fairly concrete and solidly grounded on ample anatomical electrophysiological, and functional evidence, the individual contributions of frontal regions including DLPFC, SEF, pre-SMA, FEF, and dACC are not as well-established currently. For example, we identified an executive control network in our model consisting of DLPFC, SEF, and pre-SMA. The task rules and necessary motor commands to follow them are implemented by hard-coded input and output weight patterns of its extended network (i.e., sensory input, instruction, FEF, and striatum). This implementation short-circuits a lot of the computational complexities the biological system has to solve; (i) the executive controller has to selectively retrieve the appropriate rule for the current trial from short or long-term working memory; (ii) integrate the sensory evidence to compute the correct response (e.g., via vector inversion); (iii) compute the necessary motor sequences to perform the correct action; and (iv) identify incorrectly activated prepotent responses and selectively suppress them. While neural network models with a more detailed representation of PFC exist (e.g., O'Reilly & Frank, 2006) in which rule-like representations can develop through experience, how exactly the necessary computations can be implemented dynamically is as-of-today a still unresolved question.

Critically, our focus in this work was on how PFC and BG interact when inhibitory control is required by extending the detailed BG model by Frank (2006). We also account for some electrophysiologcal data in frontal cortex, while acknowledging that there is still some uncertainty in the respective roles of these areas and their interactions, which will be open for revision as more data become available.

Learning

Previous BG models explored the role of DA in feedback driven learning (Wiecki & Frank, 2010b). As humans (but not monkeys) are able to perform this task without learning, we chose to remain agnostic about the type of learning that takes place prior to performing the task. Thus, we hard-coded task rules into the model. An additional driving factor is the lack of published reports on specific learning phenomena in the SRITS and the SST.

Conclusions

We presented a comprehensive, biologically plausible model of global and selective response inhibition which takes known properties of the neuronal underpinnings into account and tries to link them with results from cognitive science, electrophysiology, imaging studies and pharmacological experiments. Here,

we showed that augmenting our previously described BG model with the addition of the FEF, DLPFC, and rIFG allows us to simulate control over prepotent responses and to capture a wealth of data in this domain across multiple levels of analysis. We furthermore provide multiple mechanisms that can lead to disruptions in inhibitory control processes and which have implications for interpretation of data from patients with psychiatric disorders such as SZ and ADHD. Our model shows that the observed deficits in inhibitory control paradigms do not necessarily have to reflect dysfunctional response inhibition per se but could be due to other factors—such as salience, conflict detection, and/or motivation—and could be related to distinct neural mechanisms.

References

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

Alexander, W. H., & Brown, J. W. (2011). Medial prefrontal cortex as an action-outcome predictor. *Nature Neuroscience, 14,* 1338–1344.

Andrés, P. (2003). Frontal cortex as the central executive of working memory: Time to revise our view. *Cortex, 39,* 871–896.

Aron, A. R. (2007). The neural basis of inhibition in cognitive control. *The Neuroscientist, 13,* 214–228.

Aron, A. R. (2011). From reactive to proactive and selective control: Developing a richer model for stopping inappropriate responses. *Biological Psychiatry, 69*(12), e55–e68. doi:10.1016/j.biopsych.2010.07.024

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

Aron, A. R., Durston, S., Eagle, D. M., Logan, G. D., Stinear, C. M., & Stuphorn, V. (2007). Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. *The Journal of Neuroscience, 27,* 11860–11864. doi:10.1523/JNEUROSCI.3644-07.2007

Aron, A. R., Fletcher, P. C., Bullmore, E. T., Sahakian, B. J., & Robbins, T. W. (2003). Stop-signal inhibition disrupted by damage to right inferior frontal gyrus in humans. *Nature Neuroscience, 6,* 115–116. doi:10.1038/nn1003

Aron, A. R., Monsell, S., Sahakian, B. J., & Robbins, T. W. (2004). A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. *Brain: A Journal of Neurology, 127,* 1561–1573.

Aron, A. R., & Poldrack, R. A. (2006). Cortical and subcortical contributions to stop signal response inhibition: Role of the subthalamic nucleus. *The Journal of Neuroscience, 26,* 2424–2433.

Aron, A. R., & Verbruggen, F. (2008). Stop the presses: Dissociating a selective from a global mechanism for stopping. *Psychological Science, 19,* 1146–1153.

Aston-Jones, G., & Cohen, J. D. (2005). An integrative theory of locus coeruleus-norepinephrine function: Adaptive gain and optimal performance. *Annual Review of Neuroscience, 28,* 403–450. doi:10.1146/annurev.neuro.28.061604.135709

Badcock, J. C., Michie, P. T., Johnson, L., & Combrinck, J. (2002). Acts of control in schizophrenia: Dissociating the components of inhibition. *Psychological Medicine, 32,* 287–297. doi:10.1017/S0033291701005128

Barcelo, F., Escera, C., Corral, M. J., & Periáñez, J. A. (2006). Task switching and novelty processing activate a common neural network for cognitive control. *Journal of Cognitive Neuroscience, 18,* 1734–1748.

Bellgrove, M. A., Chambers, C. D., Vance, A., Hall, N., Karamitsios, M., & Bradshaw, J. L. (2006). Lateralized deficit of response inhibition in early-onset schizophrenia. *Psychological Medicine, 36,* 495–505. doi:10.1017/S0033291705006409

Bernheimer, H., Birkmayer, W., Hornykiewicz, O., Jellinger, K., & Seitelberger, F. (1973). Brain dopamine and the syndromes of Parkinson and Huntington clinical, morphological and neurochemical correlations. *Journal of the Neurological Sciences, 20,* 415–455. doi:10.1016/0022-510X(73)90175-5

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. doi:10.1037/0033-295X.108.3.624

Botvinick, M. M., Cohen, J. D., & Carter, C. S. (2004). Conflict monitoring and anterior cingulate cortex: An update. *Trends in Cognitive Sciences, 8,* 539–546. doi:10.1016/j.tics.2004.10.003

Brass, M., Derrfuss, J., Forstmann, B., & von Cramon, D. Y. (2005). The role of the inferior frontal junction area in cognitive control. *Trends in Cognitive Sciences, 9,* 314–316. doi:10.1016/j.tics.2005.05.001

Braver, T. S., Barch, D. M., Gray, J. R., Molfese, D. L., & Snyder, A. (2001). Anterior cingulate cortex and response conflict: Effects of frequency, inhibition and errors. *Cerebral Cortex, 11,* 825–836.

Breier, A., Kestler, L., Adler, C., Elman, I., Wiesenfeld, N., Malhotra, A., & Pickar, D. (1998). Dopamine d2 receptor density and personal detachment in healthy subjects. *The American Journal of Psychiatry, 155,* 1440–1442.

Brown, J. W., Bullock, D., & Grossberg, S. (2004). How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades. *Neural Networks, 17,* 471–510. doi:10.1016/j.neunet.2003.08.006

Bunge, S. A., & Wallis, J. D. (Eds.). (2008). *Neuroscience of rule-guided behavior.* Oxford, England: Oxford University Press.

Burle, B., Possamaï, C.-A., Vidal, F., Bonnet, M., & Hasbroucq, T. (2002). Executive control in the Simon effect: An electromyographic and distributional analysis. *Psychological Research/Psychologische Forschung, 66,* 324–336. doi:10.1007/s00426-002-0105-6

Cai, W., Oldenkamp, C. L., & Aron, A. R. (2011). A proactive mechanism for selective suppression of response tendencies. *The Journal of Neuroscience, 31,* 5965–5969.

Casey, B. J., Nigg, J. T., & Durston, S. (2007). New potential leads in the biology and treatment of attention deficit-hyperactivity disorder. *Current Opinion in Neurology, 20,* 119–124. doi:10.1097/WCO.0b013e3280a02f78

Cavanagh, J. F., Wiecki, T. V., Cohen, M. X., Figueroa, C. M., Samanta, J., Sherman, S. J., & Frank, M. J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. *Nature Neuroscience, 14,* 1462–1467. doi:10.1038/nn.2925

Chamberlain, S. R., Del Campo, N., Dowson, J., Müller, U., Clark, L., Robbins, T. W., & Sahakian, B. J. (2007). Atomoxetine improved response inhibition in adults with attention deficit/hyperactivity disorder. *Biological Psychiatry, 62,* 977–984. doi:10.1016/j.biopsych.2007.03.003

Chamberlain, S. R., Fineberg, N. A., Blackwell, A. D., Robbins, T. W., & Sahakian, B. J. (2006). Motor inhibition and cognitive flexibility in obsessive-compulsive disorder and trichotillomania. *The American Journal of Psychiatry, 163,* 1282–1284. doi:10.1176/appi.ajp.163.7.1282

Chamberlain, S. R., Hampshire, A., Müller, U., Rubia, K., Del Campo, N., Craig, K., . . . Sahakian, B. J. (2009). Atomoxetine modulates right inferior frontal activation during inhibitory control: A pharmacological functional magnetic resonance imaging study. *Biological Psychiatry, 65,* 550–555. doi:10.1016/j.biopsych.2008.10.014

Chambers, C. D., Bellgrove, M. A., Gould, I. C., English, T., Garavan, H., Mcnaught, E., . . . Mattingley, J. B. (2007). Dissociable mechanisms of

cognitive control in prefrontal and premotor cortex. *Journal of Neurophysiology, 98,* 3638–3647. doi:10.1152/jn.00685.2007

Chambers, C. D., Garavan, H., & Bellgrove, M. A. (2009). Insights into the neural basis of response inhibition from cognitive and clinical neuroscience. *Neuroscience and Biobehavioral Reviews, 33,* 631–646. doi:10.1016/j.neubiorev.2008.08.016

Chatham, C. H., Claus, E. D., Kim, A., Curran, T., Banich, M. T., & Munakata, Y. (2012). Cognitive control reflects context monitoring, not motoric stopping, in response inhibition. *PLoS ONE, 7*(2), e31546. doi:10.1371/journal.pone.0031546

Cohen, J. R., & Poldrack, R. A. (2008). Automaticity in motor sequence learning does not impair response inhibition. *Psychonomic Bulletin & Review, 15,* 108–115. doi:10.3758/PBR.15.1.108

Collins, A. G. E., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience, 35,* 1024–1035. doi:10.1111/j.1460-9568.2011.07980.x

Congdon, E., Constable, R. T., Lesch, K. P., & Canli, T. (2009). Influence of SLC6A3 and COMT variation on neural activation during response inhibition. *Biological Psychology, 81,* 144–152. doi:10.1016/j.biopsycho.2009.03.005

Curtis, C. E., & D'Esposito, M. (2003). Persistent activity in the prefrontal cortex during working memory. *Trends in Cognitive Sciences, 7,* 415–423. doi:10.1016/S1364-6613(03)00197-9

Daw, N. D., Niv, Y., & Dayan, P. (2005). Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience, 8,* 1704–1711. doi:10.1038/nn1560

Derrfuss, J., Brass, M., Neumann, J., & von Cramon, D. Y. (2005). Involvement of the inferior frontal junction in cognitive control: Meta-analyses of switching and Stroop studies. *Human Brain Mapping, 25,* 22–34. doi:10.1002/hbm.20127

Derrfuss, J., Brass, M., & von Cramon, D. Y. (2004). Cognitive control in the posterior frontolateral cortex: Evidence from common activations in task coordination, interference control, and working memory. *NeuroImage, 23,* 604–612.

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. doi:10.1016/j.brainres.2009.07.007

Eagle, D. M., Baunez, C., Hutcheson, D. M., Lehmann, O., Shah, A. P., & Robbins, T. W. (2008). Stop-signal reaction-time task performance: Role of prefrontal cortex and subthalamic nucleus. *Cerebral Cortex, 18,* 178–188.

Eimer, M. (1995). S-R compatibility and response selection. *Acta Psychologica, 90,* 301–313. doi:10.1016/0001-6918(95)00022-M

Emeric, E. E., Leslie, M., Pouget, P., & Schall, J. D. (2010). Performance monitoring local field potentials in the medial frontal cortex of primates: Supplementary eye field. *Journal of Neurophysiology, 104,* 1523–1537. doi:10.1152/jn.01001.2009

Evans, J. S. B. T. (2003). In two minds: Dual-process accounts of reasoning. *Trends in Cognitive Sciences, 7,* 454–459. doi:10.1016/j.tics.2003.08.012

Everling, S., Dorris, M. C., Klein, R. M., & Munoz, D. P. (1999). Role of primate superior colliculus in preparation and execution of anti-saccades and pro-saccades. *The Journal of Neuroscience, 19,* 2740–2754.

Everling, S., & Munoz, D. P. (2000). Neuronal correlates for preparatory set associated with pro-saccades and anti-saccades in the primate frontal eye field. *The Journal of Neuroscience, 20,* 387–400.

Falkenstein, M., Hohnsbein, J., Hoormann, J., & Blanke, L. (1991). Effects of cross-modal divided attention on late ERP components: II. Error processing in choice reaction tasks. *Electroencephalography and Clinical Neurophysiology, 78,* 447–455. doi:10.1016/0013-4694(91)90062-9

Faraone, S. V., Perlis, R. H., Doyle, A. E., Smoller, J. W., Goralnick, J. J., Holmgren, M. A., & Sklar, P. (2005). Molecular genetics of attention-deficit/hyperactivity disorder. *Biological Psychiatry, 57,* 1313–1323. doi:10.1016/j.biopsych.2004.11.024

Farley, I. J., Price, K. S., McCullough, E., Deck, J. H., Hordynski, W., & Hornykiewicz, O. (1978, April 28). Norepinephrine in chronic paranoid schizophrenia: Above-normal levels in limbic forebrain. *Science, 200,* 456–458. doi:10.1126/science.644310

Fleming, S. M., Thomas, C. L., & Dolan, R. J. (2010, March 15). Overcoming status quo bias in the human brain. *Proceedings of the National Academy of Sciences, USA.* Advance online publication. doi:10.1073/pnas.0910380107

Ford, K. A., & Everling, S. (2009). Neural activity in primate caudate nucleus associated with pro- and antisaccades. *Journal of Neurophysiology, 102,* 2334–2341. doi:10.1152/jn.00125.2009

Forstmann, B. U., Anwander, A., Schfer, A., Neumann, J., Brown, S., Wagenmakers, E.-J., . . . Turner, R. (2010, August 23). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. *Proceedings of the National Academy of Sciences, USA.* Advance online publication. doi:10.1073/pnas.1004932107

Forstmann, B. U., Dutilh, G., Brown, S., Neumann, J., von Cramon, D. Y., Ridderinkhof, K. R., & Wagenmakers, E.-J. (2008, November 3). Striatum and pre-SMA facilitate decision-making under time pressure. *Proceedings of the National Academy of Sciences, USA.* Advance online publication. doi:10.1073/pnas.0805903105

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. doi:10.1162/0898929052880093

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., & Badre, D. (2012). Mechanisms of hierarchical reinforcement learning in corticostriatal circuits 1: Computational analysis. *Cerebral Cortex, 22,* 509–526. doi:10.1093/cercor/bhr114

Frank, M. J., Samanta, J., Moustafa, A. A., & Sherman, S. J. (2007, October 25). Hold your horses: Impulsivity, deep brain stimulation, and medication in Parkinsonism. *Science, 318,* 1309–1312. doi:10.1126/science.1146157

Frank, M. J., Santamaria, A., O'Reilly, R. C., & Willcutt, E. (2007). Testing computational models of dopamine and noradrenaline dysfunction in attention deficit/hyperactivity disorder. *Neuropsychopharmacology, 32,* 1583–1599.

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 of London Series B: Biological Sciences, 362,* 1641–1654. doi:10.1098/rstb.2007.2058

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004, November 4). By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science, 306,* 1940–1943. doi:10.1126/science.1102941

Funahashi, S., Chafee, M. V., & Goldman-Rakic, P. S. (1993, October 21). Prefrontal neuronal activity in rhesus monkeys performing a delayed anti-saccade task. *Nature, 365,* 753–756. doi:10.1038/365753a0

Garavan, H., Hester, R., Murphy, K., Fassbender, C., & Kelly, C. (2006). Individual differences in the functional neuroanatomy of inhibitory control. *Brain Research, 1105,* 130–142.

Gehring, W. J., Goss, B., Coles, M. G. H., Meyer, D. E., & Donchin, E. (1993). A neural system for error detection and compensation. *Psychological Science, 4,* 385–390.

Goldberg, J. H., Farries, M. A., & Fee, M. S. (2012). Integration of cortical and pallidal inputs in the basal ganglia-recipient thalamus of singing birds. *Journal of Neurophysiology, 108,* 1403–1429. doi:10.1152/jn.00056.2012

Greenhouse, I., Oldenkamp, C. L., Aron, A. R., & Diego, S. (2012). Stopping a response has global or non-global effects on the motor

system depending on preparation. *Journal of Neurophysiology, 107,* 384–392. doi:10.1152/jn.00704.2011

Gurney, K., Prescott, T. J., & Redgrave, P. (2001). A computational model of action selection in the basal ganglia: I. A new functional anatomy. *Biological Cybernetics, 84,* 401–410. doi:10.1007/PL00007984

Haber, S. N. (2003). The primate basal ganglia: Parallel and integrative networks. *Journal of Chemical Neuroanatomy, 26,* 317–330. doi:10.1016/j.jchemneu.2003.10.003

Hallett, P. E. (1978). Primary and secondary saccades to goals defined by instructions. *Vision Research, 18,* 1279–1296.

Hampshire, A., Chamberlain, S. R., Monti, M. M., Duncan, J., & Owen, A. M. (2010). The role of the right inferior frontal gyrus: Inhibition and attentional control. *NeuroImage, 50,* 1313–1319. doi:10.1016/j.neuroimage.2009.12.109

Harden, D. G., & Grace, A. A. (1995). Activation of dopamine cell firing by repeated L-DOPA administration to dopamine-depleted rats: Its potential role in mediating the therapeutic response to L-DOPA treatment. *The Journal of Neuroscience, 15,* 6157–6166.

Harris, M. S., Reilly, J. L., Keshavan, M. S., & Sweeney, J. A. (2006). Longitudinal studies of antisaccades in antipsychotic-naive first-episode schizophrenia. *Psychological Medicine, 36,* 485–494. doi:10.1017/S0033291705006756

Hess, W., Bürgi, S., & Bucher, V. (1946). Motorische Funktion des Tectal- und Tegmentalgebietes [Motor function of the tectal and tegmental area]. *Monatsschrift für Psychiatrie und Neurologie, 112,* 1–26. doi:10.1159/000148295

Hikida, T., Kimura, K., Wada, N., Funabiki, K., & Nakanishi, S. (2010). Distinct roles of synaptic transmission in direct and indirect striatal pathways to reward and aversive behavior. *Neuron, 66,* 896–907. doi:10.1016/j.neuron.2010.05.011

Hikosaka, O. (1989). Role of basal ganglia in initiation of voluntary movements. In M. A. Arbib & S. Amari (Eds.), *Dynamic interactions in neural networks: Models and data* (pp. 153–167). doi:10.1007/978-1-4612-4536-0_9

Hikosaka, O. (2007). GABAergic output of the basal ganglia. *Progress in Brain Research, 160,* 209–226.

Hikosaka, O., & Isoda, M. (2008). Brain mechanisms for switching from automatic to controlled eye movements. *Progress in Brain Research, 171,* 375–382. doi:10.1016/S0079-6123(08)00655-9

Hikosaka, O., Nakamura, K., & Nakahara, H. (2006). Basal ganglia orient eyes to reward. *Journal of Neurophysiology, 95,* 567–584. doi:10.1152/jn.00458.2005

Hikosaka, O., Takikawa, Y., & Kawagoe, R. (2000). Role of the basal ganglia in the control of purposive saccadic eye movements. *Physiological Reviews, 80,* 953–978.

Hikosaka, O., & Wurtz, R. H. (1986). Saccadic eye movements following injection of lidocaine into the superior colliculus. *Experimental Brain Research, 61,* 531–539.

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709. doi:10.1037/0033-295X.109.4.679

Hu, S., & Li, C.-S. (2011). Neural processes of preparatory control for stop signal inhibition. *Human Brain Mapping, 33,* 2785–2796. doi:10.1002/hbm.21399

Huddy, V. C., Aron, A. R., Harrison, M., Barnes, T. R. E., Robbins, T. W., & Joyce, E. M. (2009). Impaired conscious and preserved unconscious inhibitory processing in recent onset schizophrenia. *Psychological Medicine, 39,* 907–916. doi:10.1017/S0033291708004340

Huerta, M. F., Krubitzer, L. A., & Kaas, J. H. (1987). Frontal eye field as defined by intracortical microstimulation in squirrel monkeys, owl monkeys, and macaque monkeys: II. Cortical connections. *The Journal of Comparative Neurology, 265,* 332–361. doi:10.1002/cne.902650304

Huettel, S. A., & McCarthy, G. (2004). What is odd in the oddball task? Prefrontal cortex is activated by dynamic changes in response strategy. *Neuropsychologia, 42,* 379–386. doi:10.1016/j.neuropsychologia.2003.07.009

Hutton, S. B., & Ettinger, U. (2006). The antisaccade task as a research tool in psychopathology: A critical review. *Psychophysiology, 43,* 302–313. doi:10.1111/j.1469-8986.2006.00403.x

Hwang, K., Velanova, K., & Luna, B. (2010). Strengthening of top-down frontal cognitive control networks underlying the development of inhibitory control: A functional magnetic resonance imaging effective connectivity study. *The Journal of Neuroscience, 30,* 15535–15545.

Isoda, M., & Hikosaka, O. (2007). Switching from automatic to controlled action by monkey medial frontal cortex. *Nature Neuroscience, 10,* 240–248. doi:10.1038/nn1830

Isoda, M., & Hikosaka, O. (2008). Role for subthalamic nucleus neurons in switching from automatic to controlled eye movement. *The Journal of Neuroscience, 28,* 7209–7218.

Jahfari, S., Verbruggen, F., Frank, M. J., Waldorp, L. J., Colzato, L., Ridderinkhof, K. R., & Forstmann, B. U. (2012). How preparation changes the need for top-down control of the basal ganglia when inhibiting premature actions. *The Journal of Neuroscience, 32,* 10870–10878. doi:10.1523/JNEUROSCI.0902-12.2012

Jahfari, S., Waldorp, L., van den Wildenberg, W. P. M., 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. *The Journal of Neuroscience, 31,* 6891–6899.

Jocham, G., Klein, T. A., & Ullsperger, M. (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *The Journal of Neuroscience, 31,* 1606–1613. doi:10.1523/JNEUROSCI.3904-10.2011

Johnston, K., & Everling, S. (2006). Monkey dorsolateral prefrontal cortex sends task-selective signals directly to the superior colliculus. *The Journal of Neuroscience, 26,* 12471–12478.

Kolling, N., Behrens, T., Mars, R., & Rushworth, M. (2012, April 6). Neural mechanisms of foraging. *Science, 336,* 95–98. doi:10.1126/science.1216930

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: Cognitive basis for stimulus-response compatibility—A model and taxonomy. *Psychological Review, 97,* 253–270. doi:10.1037/0033-295X.97.2.253

Kravitz, A. V., Freeze, B. S., Parker, P. R. L., Kay, K., Thwin, M. T., Deisseroth, K., & Kreitzer, A. C. (2010, July 29). Regulation of Parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. *Nature, 466,* 622–626. doi:10.1038/nature09159

Kravitz, A. V., Tye, L. D., & Kreitzer, A. C. (2012). Distinct roles for direct and indirect pathway striatal neurons in reinforcement. *Nature Neuroscience, 15,* 816–818. doi:10.1038/nn.3100

Kühn, A. A., Williams, D., Kupsch, A., Limousin, P., Hariz, M., Schneider, G.-H., . . . Brown, P. (2004). Event-related beta desynchronization in human subthalamic nucleus correlates with motor performance. *Brain: A Journal of Neurology, 127,* 735–746. doi:10.1093/brain/awh106

Leotti, L. A., & Wager, T. D. (2010). Motivational influences on response inhibition measures. *Journal of Experimental Psychology: Human Perception and Performance, 36,* 430–447. doi:10.1037/a0016802

Leung, H. C., & Cai, W. (2007). Common and differential ventrolateral prefrontal activity during inhibition of hand and eye movements. *The Journal of Neuroscience, 27,* 9893–9900.

Lo, C.-C., Boucher, L., Paré, M., Schall, J. D., & Wang, X.-J. (2009). Proactive inhibitory control and attractor dynamics in countermanding action: A spiking neural circuit model. *The Journal of Neuroscience, 29,* 9059–9071. doi:10.1523/JNEUROSCI.6164-08.2009

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. doi:10.1038/nn1722

Logan, G. D. (1985). On the ability to inhibit simple thoughts and actions: II. Stop-signal studies of repetition priming. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 675–691.

Logan, G. D., & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91,* 295–327. doi:10.1037/0033-295X.91.3.295

Logan, G. D., Schachar, R. J., & Tannock, R. (1997). Impulsivity and inhibitory control. *Psychological Science, 8,* 60–64. doi:10.1111/j.1467-9280.1997.tb00545.x

Lu, M. T., Preston, J. B., & Strick, P. L. (1994). Interconnections between the prefrontal cortex and the premotor areas in the frontal lobe. *The Journal of Comparative Neurology, 341,* 375–392. doi:10.1002/cne.903410308

Mansfield, E. L., Karayanidis, F., Jamadar, S., Heathcote, A., & Forstmann, B. U. (2011). Adjustments of response threshold during task switching: A model-based functional magnetic resonance imaging study. *The Journal of Neuroscience, 31,* 14688–14692. doi:10.1523/JNEUROSCI.2390-11.2011

McDowell, J. E., Brown, G. G., Paulus, M., Martinez, A., Stewart, S. E., Dubowitz, D. J., & Braff, D. L. (2002). Neural correlates of refixation saccades and antisaccades in normal and schizophrenia subjects. *Biological Psychiatry, 51,* 216–223. doi:10.1016/S0006-3223(01)01204-5

Menzies, L., Achard, S., Chamberlain, S. R., Fineberg, N., Chen, C. H., del Campo, N., . . . Bullmore, E. (2007). Neurocognitive endophenotypes of obsessive-compulsive disorder. *Brain: A Journal of Neurology, 130,* 3223–3236.

Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience, 24,* 167–202. doi:10.1146/annurev.neuro.24.1.167

Mink, J. W. (1996). The basal ganglia: Focused selection and inhibition of competing motor programs. *Progress in Neurobiology, 50,* 381–425. doi:10.1016/S0301-0082(96)00042-1

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognitive Psychology, 41,* 49–100. doi:10.1006/cogp.1999.0734

Montague, P. R., Dayan, P., & Sejnowski, T. J. (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *The Journal of Neuroscience, 16,* 1936–1947.

Monterosso, J. R., Aron, A. R., Cordova, X., Xu, J., & London, E. D. (2005). Deficits in response inhibition associated with chronic methamphetamine abuse. *Drug and Alcohol Dependence, 79,* 273–277. doi:10.1016/j.drugalcdep.2005.02.002

Morein-Zamir, S., Fineberg, N. A., Robbins, T. W., & Sahakian, B. J. (2010). Inhibition of thoughts and actions in obsessive-compulsive disorder: Extending the endophenotype? *Psychological Medicine, 40,* 263–272. doi:10.1017/S003329170999033X

Morein-Zamir, S., & Kingstone, A. (2006). Fixation offset and stop signal intensity effects on saccadic countermanding: A crossmodal investigation. *Experimental Brain Research, 175,* 453–462.

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in Cognitive Sciences, 15,* 453–459. doi:10.1016/j.tics.2011.07.011

Munoz, D. P., & Everling, S. (2004). Look away: The anti-saccade task and the voluntary control of eye movement. *Nature Reviews Neuroscience, 5,* 218–228. doi:10.1038/nrn1345

Nakamura, K., & Hikosaka, O. (2006). Role of dopamine in the primate caudate nucleus in reward modulation of saccades. *The Journal of Neuroscience, 26,* 5360–5369.

Nambu, A., Tokuno, H., Hamada, I., Kita, H., Imanishi, M., Akazawa, T., . . . Hasegawa, N. (2000). Excitatory cortical inputs to pallidal neurons via the subthalamic nucleus in the monkey. *Journal of Neurophysiology, 84,* 289–300.

Nambu, A., Tokuno, H., & Takada, M. (2002). Functional significance of the cortico-subthalamo-pallidal "hyperdirect" pathway. *Neuroscience Research, 43,* 111–117. doi:10.1016/S0168-0102(02)00027-5

Neubert, F.-X., Mars, R. B., Buch, E. R., Olivier, E., & Rushworth, M. F. S. (2010). Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *Proceedings of the National Academy of Sciences, USA, 107,* 13240–13245. doi:10.1073/pnas.1000674107

Nieuwenhuis, S., Broerse, A., Nielen, M. M. A., & de Jong, R. (2004). A goal activation approach to the study of executive function: An application to antisaccade tasks. *Brain and Cognition, 56,* 198–214. doi:10.1016/j.bandc.2003.12.002

Nigg, J. T. (2001). Is ADHD a disinhibitory disorder? *Psychological Bulletin, 127,* 571–598. doi:10.1037/0033-2909.127.5.571

Nigg, J. T., Wong, M. M., Martel, M. M., Jester, J. M., Puttler, L. I., Glass, J. M., . . . Zucker, R. A. (2006). Poor response inhibition as a predictor of problem drinking and illicit drug use in adolescents at risk for alcoholism and other substance use disorders. *Journal of the American Academy of Child & Adolescent Psychiatry, 45,* 468–475. doi:10.1097/01.chi.0000199028.76452.a9

Obeso, I., Wilkinson, L., Casabona, E., Bringas, M. L., Alvarez, M., Alvarez, L., . . . Jahanshahi, M. (2011). Deficits in inhibitory control and conflict resolution on cognitive and motor tasks in Parkinson's disease. *Experimental Brain Research, 212,* 371–384.

Obeso, I., Wilkinson, L., & Jahanshahi, M. (2011). Levodopa medication does not influence motor inhibition or conflict resolution in a conditional stop-signal task in Parkinson's disease. *Experimental Brain Research, 213,* 435–445. doi:10.1007/s00221-011-2793-x

Oosterlaan, J., Logan, G. D., & Sergeant, J. A. (1998). Response inhibition in AD/HD, CD, comorbid AD/HD+CD, anxious, and control children: A meta-analysis of studies with the stop task. *Journal of Child Psychology and Psychiatry, 39,* 411–425. doi:10.1017/S0021963097002072

O'Reilly, R. C., & Frank, M. J. (2006). Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia. *Neural Computation, 18,* 283–328. doi:10.1162/089976606775093909

O'Reilly, R. C., & Munakata, Y. (2000). *Computational explorations in cognitive neuroscience: Understanding the mind by simulating the brain.* Cambridge, MA: MIT Press.

Osman, A., Kornblum, S., & Meyer, D. E. (1986). The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 243–258. doi:10.1037/0096-1523.12.3.243

Palminteri, S., Lebreton, M., Worbe, Y., Grabli, D., Hartmann, A., & Pessiglione, M. (2009). Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proceedings of the National Academy of Sciences, USA, 106,* 19179–19184. doi:10.1073/pnas.0904035106

Paré, M., & Hanes, D. P. (2003). Controlled movement processing: Superior colliculus activity associated with countermanded saccades. *The Journal of Neuroscience, 23,* 6480–6489.

Parent, A., & Hazrati, L. N. (1995). Functional anatomy of the basal ganglia: II. The place of sub-thalamic nucleus and external pallidum in basal ganglia circuitry. *Brain Research Reviews, 20,* 128–154. doi:10.1016/0165-0173(94)00008-D

Parmentier, F. B. R., Elford, G., Escera, C., Andrés, P., & San Miguel, I. (2008). The cognitive locus of distraction by acoustic novelty in the cross-modal oddball task. *Cognition, 106,* 408–432. doi:10.1016/j.cognition.2007.03.008

Penadés, R., Catalán, R., Rubia, K., Andrés, S., Salamero, M., & Gastó, C. (2007). Impaired response inhibition in obsessive compulsive disorder. *European Psychiatry, 22,* 404–410.

Pouget, P., Logan, G. D., Palmeri, T. J., Boucher, L., Paré, M., & Schall, J. D. (2011). Neural basis of adaptive response time adjustment during saccade countermanding. *The Journal of Neuroscience, 31,* 12604–12612. doi:10.1523/JNEUROSCI.1868-11.2011

Ramos, B. P. & Arnsten, A. F. T. (2007). Adrenergic pharmacology and cognition: Focus on the prefrontal cortex. *Pharmacology & Therapeutics, 113,* 523–536. doi:10.1016/j.pharmthera.2006.11.006

Ratcliff, R., & Frank, M. J. (2012). Reinforcement-based decision making in corticostriatal circuits: Mutual constraints by neurocomputational and diffusion models. *Neural Computation, 24,* 1186–1229. doi:10.1162/NECO_a_00270

Ray, N. J., Jenkinson, N., Brittain, J., Holland, P., Joint, C., Nandi, D., . . . Aziz, T. Z. (2009). The role of the subthalamic nucleus in response inhibition: Evidence from deep brain stimulation for Parkinson's disease. *Neuropsychologia, 47,* 2828–2834. doi:10.1016/j.neuropsychologia.2009.06.011

Reilly, J. L., Harris, M. S., Keshavan, M. S., & Sweeney, J. A. (2006). Adverse effects of risperidone on spatial working memory in first-episode schizophrenia. *Archives of General Psychiatry, 63,* 1189–1197. doi:10.1001/archpsyc.63.11.1189

Reilly, J. L., Harris, M. S., Khine, T. T., Keshavan, M. S., & Sweeney, J. A. (2007). Antipsychotic drugs exacerbate impairment on a working memory task in first-episode schizophrenia. *Biological Psychiatry, 62,* 818–821. doi:10.1016/j.biopsych.2006.10.031

Reuter, B., & Kathmann, N. (2004). Using saccade tasks as a tool to analyze executive dysfunctions in schizophrenia. *Acta Psychologica, 115,* 255–269. doi:10.1016/j.actpsy.2003.12.009

Ridderinkhof, K. R. (2002). Micro- and macro-adjustments of task set: Activation and suppression in conflict tasks. *Psychological Research/Psychologische Forschung, 66,* 312–323. doi:10.1007/s00426-002-0104-7

Ridderinkhof, K. R., Forstmann, B. U., Wylie, S. A., Burle, B., & van den Wildenberg, W. P. M. (2011). Neurocognitive mechanisms of action control: Resisting the call of the Sirens. *Wiley Interdisciplinary Reviews: Cognitive Science, 2,* 174–192. doi:10.1002/wcs.99

Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition, 56,* 129–140. doi:10.1016/j.bandc.2004.09.016

Roberts, R. J., Hager, L. D., & Heron, C. (1994). Prefrontal cognitive processes: Working memory and inhibition in the antisaccade task. *Journal of Experimental Psychology: General, 123,* 374–393. doi:10.1037/0096-3445.123.4.374

Rougier, N. P., Noelle, D., Braver, T. S., Cohen, J. D., & O'Reilly, R. C. (2005). Prefrontal cortex and the flexibility of cognitive control: Rules without symbols. *Proceedings of the National Academy of Sciences, USA, 102,* 7338–7343. doi:10.1073/pnas.0502455102

Rowe, J., Friston, K., Frackowiak, R., & Passingham, R. (2002). Attention to action: Specific modulation of corticocortical interactions in humans. *NeuroImage, 17,* 988–998. doi:10.1006/nimg.2002.1156

Rubchinsky, L. L., Kopell, N., & Sigvardt, K. A. (2003). Modeling facilitation and inhibition of competing motor programs in basal ganglia subthalamic nucleus-pallidal circuits. *Proceedings of the National Academy of Sciences, USA, 100,* 14427–14432. doi:10.1073/pnas.2036283100

Sakagami, M., Tsutsui, K., Lauwereyns, J., Koizumi, M., Kobayashi, S., & Hikosaka, O. (2001). A code for behavioral inhibition on the basis of color, but not motion, in ventrolateral prefrontal cortex of macaque monkey. *The Journal of Neuroscience, 21,* 4801–4808.

Schachar, R., & Logan, G. D. (1990). Impulsivity and inhibitory control in normal development and childhood psychopathology. *Developmental Psychology, 26,* 710–720. doi:10.1037/0012-1649.26.5.710

Schlag-Rey, M., Amador, N., Sanchez, H., & Schlag, J. (1997, November 27). Antisaccade performance predicted by neuronal activity in the supplementary eye field. *Nature, 390,* 398–401. doi:10.1038/37114

Schlag-Rey, M., & Schlag, J. (1984). Visuomotor functions of central thalamus in monkey: I. Unit activity related to spontaneous eye movements. *Journal of Neurophysiology, 51,* 1149–1174.

Schmidt, R., Leventhal, D., Pettibone, J., Case, A., & Berke, J. (2012). Suppressing actions in the basal ganglia [Abstract]. In *COSYNE 2012: The 9th Annual Computational and Systems Neuroscience Meeting* (p. 139). Retrieved from http://cosyne.org/cosyne12/Cosyne2012_program_book.pdf

Sharp, D. J., Bonnelle, V., De Boissezon, X., Beckmann, C. F., James, S. G., Patel, M. C., & Mehta, M. A. (2010). Distinct frontal systems for response inhibition, attentional capture, and error processing. *Proceedings of the National Academy of Sciences, USA.* Advance online publication. doi:10.1073/pnas.1000175107

Shen, W., Flajolet, M., Greengard, P., & Surmeier, D. J. (2008, August 8). Dichotomous dopaminergic control of striatal synaptic plasticity. *Science, 321,* 848–851. doi:10.1126/science.1160575

Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of go/no-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia, 46,* 224–232. doi:10.1016/j.neuropsychologia.2007.07.015

Simon, J. R. (1969). Reactions toward the source of stimulation. *Journal of Experimental Psychology, 81,* 174–176. doi:10.1037/h0027448

Sloman, S. A. (1996). The empirical case for two systems of reasoning. *Psychological Bulletin, 119,* 3–22. doi:10.1037/0033-2909.119.1.3

Sommer, M. A., & Wurtz, R. H. (2002, May 24). A pathway in primate brain for internal monitoring of movements. *Science, 296,* 1480–1482. doi:10.1126/science.1069590

Sommer, M. A., & Wurtz, R. H. (2004a). What the brain stem tells the frontal cortex: I. Oculomotor signals sent from superior colliculus to frontal eye field via mediodorsal thalamus. *Journal of Neurophysiology, 91,* 1381–1402. doi:10.1152/jn.00738.2003

Sommer, M. A., & Wurtz, R. H. (2004b). What the brain stem tells the frontal cortex: II. Role of the SC-MD-FEF pathway in corollary discharge. *Journal of Neurophysiology, 91,* 1403–1423. doi:10.1152/jn.00740.2003

Sommer, M. A., & Wurtz, R. H. (2006, November 16). Influence of the thalamus on spatial visual processing in frontal cortex. *Nature, 444,* 374–377. doi:10.1038/nature05279

Sparks, D. L. (2002). The brainstem control of saccadic eye movements. *Nature Reviews Neuroscience, 3,* 952–964. doi:10.1038/nrn986

Stevens, A. A., Skudlarski, P., Gatenby, J. C., & Gore, J. C. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magnetic Resonance Imaging, 18,* 495–502. doi:10.1016/S0730-725X(00)00128-4

Stuphorn, V., & Schall, J. D. (2006). Executive control of countermanding saccades by the supplementary eye field. *Nature Neuroscience, 9,* 925–931. doi:10.1038/nn1714

Stuphorn, V., Taylor, T. L., & Schall, J. D. (2000, December 14). Performance monitoring by the supplementary eye field. *Nature, 408,* 857–860. doi:10.1038/35048576

Swann, N. C., Cai, W., Conner, C. R., Pieters, T. A., Claffey, M. P., George, J. S., . . . Tandon, N. (2012). Roles for the pre-supplementary motor area and the right inferior frontal gyrus in stopping action: Electrophysiological responses and functional and structural connectivity. *NeuroImage, 59,* 2860–2870. doi:10.1016/j.neuroimage.2011.09.049

Swann, N., Poizner, H., Houser, M., Gould, S., Greenhouse, I., Cai, W., . . . Aron, A. R. (2011). Deep brain stimulation of the subthalamic nucleus alters the cortical profile of response inhibition in the beta frequency

band: A scalp EEG study in Parkinson's disease. *The Journal of Neuroscience, 31,* 5721–5729.

Taverna, S., Ilijic, E., & Surmeier, D. J. (2008). Recurrent collateral connections of striatal medium spiny neurons are disrupted in models of Parkinson's disease. *The Journal of Neuroscience, 28,* 5504–5512. doi:10.1523/JNEUROSCI.5493-07.2008

van Koningsbruggen, M. G., Pender, T., Machado, L., & Rafal, R. D. (2009). Impaired control of the oculomotor reflexes in Parkinson's disease. *Neuropsychologia, 47,* 2909–2915. doi:10.1016/j.neuropsychologia.2009.06.018

Vazey, E. M., & Aston-Jones, G. (2012). The emerging role of norepinephrine in cognitive dysfunctions of Parkinson's disease. *Frontiers in Behavioral Neuroscience, 6,* 48. doi: 10.3389/fnbeh.2012.00048

Verbruggen, F., Aron, A. R., Stevens, M. A., & Chambers, C. D. (2010). Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. *Proceedings of the National Academy of Sciences, USA, 107,* 13966–13971. doi:10.1073/pnas.1001957107

Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences, 12,* 418–424. doi:10.1016/j.tics.2008.07.005

Verbruggen, F., & Logan, G. D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience and Biobehavioral Reviews, 33,* 647–661. doi:10.1016/j.neubiorev.2008.08.014

Voon, V., Pessiglione, M., Brezing, C., Gallea, C., Fernandez, H. H., Dolan, R. J., & Hallett, M. (2010). Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron, 65,* 135–142. doi:10.1016/j.neuron.2009.12.027

Wallis, J. D., & Miller, E. K. (2003). Neuronal activity in primate dorsolateral and orbital prefrontal cortex during performance of a reward preference task. *European Journal of Neuroscience, 18,* 2069–2081. doi:10.1046/j.1460-9568.2003.02922.x

Wang, D., Kristjansson, A., & Nakayama, K. (2005). Efficient visual search without top-down or bottom-up guidance. *Perception & Psychophysics, 67,* 239–253. doi:10.3758/BF03206488

Wang, Y., Isoda, M., Matsuzaka, Y., Shima, K., & Tanji, J. (2005). Prefrontal cortical cells projecting to the supplementary eye field and presupplementary motor area in the monkey. *Neuroscience Research, 53,* 1–7. doi:10.1016/j.neures.2005.05.005

Watanabe, M., & Munoz, D. P. (2009). Neural correlates of conflict resolution between automatic and volitional actions by basal ganglia. *European Journal of Neuroscience, 30,* 2165–2176. doi:10.1111/j.1460-9568.2009.06998.x

Watanabe, M., & Munoz, D. P. (2010). Presetting basal ganglia for volitional actions. *The Journal of Neuroscience, 30,* 10144–10157. doi:10.1523/JNEUROSCI.1738-10.2010

Watanabe, M., & Munoz, D. P. (2011). Probing basal ganglia functions by saccade eye movements. *European Journal of Neuroscience, 33,* 2070–2090. doi:10.1111/j.1460-9568.2011.07691.x

Wegener, S. P., Johnston, K., & Everling, S. (2008). Microstimulation of monkey dorsolateral prefrontal cortex impairs antisaccade performance. *Experimental Brain Research, 190,* 463–473.

Wiecki, T., & Frank, M. (2010a, November). *Computational models of response inhibition: Multiple level of analysis.* Poster presented at the 40th annual meeting of the Society for Neuroscience, San Diego, California.

Wiecki, T., & Frank, M. (2010b). Neurocomputational models of motor and cognitive deficits in Parkinson's disease. In A. Bjorklund & M. A. Cenci (Eds.), *Progress in brain research: Recent advances in Parkinson's disease—Part I: Basic research* (Vol. 183, pp. 275–297). doi:10.1016/S0079-6123(10)83014-6

Wylie, S. A., Ridderinkhof, K. R., Elias, W. J., Frysinger, R. C., Bashore, T. R., Downs, K. E., . . . van den Wildenberg, W. P. M. (2010). Subthalamic nucleus stimulation influences expression and suppression of impulsive behaviour in Parkinson's disease. *Brain, 133,* 3611–3624. doi:10.1093/brain/awq239

Xue, G., Aron, A. R., & Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex, 18,* 1923–1932. doi:10.1093/cercor/bhm220

Yeung, N., Botvinick, M. M., & Cohen, J. D. (2004). The neural basis of error detection: Conflict monitoring and the error-related negativity. *Psychological Review, 111,* 931–959. doi:10.1037/0033-295X.111.4.931

Yeung, N., & Cohen, J. D. (2006). The impact of cognitive deficits on conflict monitoring: Predictable dissociations between the error-related negativity and N2. *Psychological Science, 17,* 164–171.

Zaghloul, K. A., Weidemann, C. T., Lega, B. C., Jaggi, J. L., Baltuch, G. H., & Kahana, M. J. (2012). Neuronal activity in the human subthalamic nucleus encodes decision conflict during action selection. *The Journal of Neuroscience, 32,* 2453–2460. doi:10.1523/JNEUROSCI.5815-11.2012

Zandbelt, B. B., & Vink, M. (2010). On the role of the striatum in response inhibition. *PLoS ONE, 5*(11), e13848. doi:10.1371/journal.pone.0013848

(Appendix follows)

Appendix

Implementational and Mathematical Details of the Model

Software

The model and the Python scripts are available in the online supplemental materials.

Implementation Details

Like Frank's (2006) original model, this model is implemented in the Emergent neural modeling software framework (Aisa et al., 2008), which can be downloaded here: http://grey.colorado.edu/emergent/index.php/Main Page.

Emergent measures simulator time in cycles. Here, we convert this time to milliseconds by multiplying cycles by 4 to roughly match behavioral and electrophysiological data.

Emergent uses point neurons with excitatory, inhibitory, and leak conductances contributing to an integrated membrane potential, which is then thresholded and transformed via an $\frac{x}{x+1}$ sigmoidal function to produce a rate code output communicated to other neurons (discrete spiking can also be used, but produces noisier results).

The membrane potential V_m is a function of ionic conductances g with reversal (driving) potentials E as follows:

$$\Delta V_m(t) = \tau \sum_c g_c(t) \overline{g}_c (E_c - V_m(t)), \qquad (A1)$$

with 3 channels (c) corresponding to the following: e excitatory input, l leak current, and i inhibitory input. Following electrophysiological convention, the overall conductance is decomposed into a time-varying component $g_c(t)$ computed as a function of the dynamic state of the model, and a constant $\overline{g}_c$ 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:

$$V_m^{\infty} = \frac{g_e \overline{g}_e}{g_e \overline{g}_e + g_l \overline{g}_e + g_i \overline{g}_e}, \qquad (A2)$$

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 (O'Reilly & Munakata, 2000).

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

$$\eta_j = g_e(t) = \langle x_i w_{ij} \rangle = \frac{1}{n} \sum_i x_i w_{ij}. \qquad (A3)$$

The inhibitory conductance can either be computed by the kWTA (k-Winners-Take-All) function described in the next section or by modeling inhibitory interneurons. Leak is a constant.

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

$$y_j(t) = \frac{1}{\left(1 + \frac{1}{\gamma [V_m(t) - \Theta]_+}\right)}, \qquad (A4)$$

where $[x]_+$ is a threshold function that returns 0 if x0 and x if x0. To avoid dividing by 0, we assume $y_j(t) = 0$ if it returns 0. This activation is subject to scaling factors (wt_scale.abs and wt_scale.rel) that modify how much impact the projections have on the post-synaptic neurons.

Inhibition Within and Between Layers

Inhibition between layers (i.e., for GABAergic projections between BG layers and striatal inhibitory interneurons) 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. For *within* layer lateral inhibition (used here in premotor cortex), Leabra uses a kWTA (k-Winners-Take-All) function to achieve inhibitory competition among neurons within each layer (area). The kWTA function computes a uniform level of inhibitory current for all neurons in the layer, such that the k + 1th 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 (O'Reilly & Munakata, 2000). Thus, although the kWTA function is somewhat biologically implausible in its implementation (e.g., requiring global information about activation states and using sorting mechanisms), it provides a computationally effective approximation to biologically plausible inhibitory dynamics. kWTA is computed via a uniform level of inhibitory current for all neurons in the layer as follows:

$$g_i = g_{k+1}^{\Theta} + q(g_k^{\Theta} - g_{k+1}^{\Theta}), \qquad (A5)$$

where 0q1 (0.25 default) is a parameter $\square$ for setting the inhibition between the upper bound of g_k and $\square$. These boundary inhibition values are the lower bound of g_{k+1} computed as a function of the level of inhibition necessary to keep a unit right at threshold:

$$g_i = g_{k+1}^{\Theta} + q(g_k^{\Theta} - g_{k+1}^{\Theta}). \qquad (A6)$$

(Appendix continues)

In the basic version of the kWTA function, which is relatively rigid about the kWTA constraint and is therefore used for output layers, g_k^Θ and g_{k+1}^Θ are set to the threshold inhibition value for the kth and k + 1th most excited neurons, respectively. Thus, the inhibition is placed exactly to allow k neurons to be above threshold, and the remainder below threshold. For this version, the q parameter is almost always .25, allowing the kth unit to be sufficiently above the inhibitory threshold.

The premotor cortex uses the average-based kWTA version, g_k^Θ is the average g_i^Θ value for the top k most excited neurons, and g_{k+1}^Θ is the average of g_i^Θ for the remaining $n - k$ neurons. This version allows for more flexibility in the actual number of neurons active depending on the nature of the activation distribution in the layer and the value of the q parameter (which is typically .6), and is therefore used for hidden layers.

Hysterisis and Accommodation

$$I_a(t) = g_a(t)\overline{g}_a\big(V_m(t) - E_a\big). \tag{A7}$$

$$I_h(t) = g_h(t)\overline{g}_h\big(V_m(t) - E_h\big). \tag{A8}$$

E_h is excitatory; E_a is inhibitory.

g_a and g_h are time-varying functions that depend on previous activity, integrated over different time periods.

$$g_a(t) = \begin{cases} g_a(t-1) + dt_{g_a}\big(1 - g_a(t-1)\big); & \text{if } \big(b_a(t) = \Theta_a\big) \\ g_a(t-1) + dt_{g_a}\big(0 - g_a(t-1)\big); & \text{if } \big(b_a(t) = \Theta_d\big) \end{cases} \tag{A9}$$

Computation of Conflict

Dorsal anterior cingulate cortex (dACC) activity is the Hopfield energy of the presupplementary motor area:

$$\text{dACC}_{act} = \text{FEF}_{\text{left}_{act}} * \text{FEF}_{\text{right}_{act}} \tag{A10}$$