Inhibitory Control in the Cortico-Basal Ganglia-Thalamocortical Loop: Complex Regulation and Interplay with Memory and Decision Processes

Highlights

- Modeling inhibitory control as a distributed process in the cortico-subcortical system
- Long-distance recurrent dynamics of persistent activity enhance inhibitory control
- Stop signal reaction time is insensitive to task difficulty in perceptual decision
- Executive control differentially depends on various pathways in the basal ganglia

Authors

Wei Wei, Xiao-Jing Wang

Correspondence

xjwang@nyu.edu

In Brief

Wei et al. propose a cortico-basal ganglia-thalamocortical distributed attractor network model for inhibitory control. Besides reproducing recent experimental findings and making testable predictions, it provides a unified framework to investigate inhibitory control, perceptual decision making, working memory, and their interactions.
Inhibitory Control in the Cortico-Basal Ganglia-Thalamocortical Loop: Complex Regulation and Interplay with Memory and Decision Processes

Wei Wei¹ and Xiao-Jing Wang¹,²,³,*
¹Center for Neural Science, New York University, New York, NY 10003, USA
²NYU-ECNU Institute of Brain and Cognitive Science, NYU Shanghai, 200122 Shanghai, China
³Lead Contact
*Correspondence: xjwang@nyu.edu
http://dx.doi.org/10.1016/j.neuron.2016.10.031

SUMMARY

We developed a circuit model of spiking neurons that includes multiple pathways in the basal ganglia (BG) and is endowed with feedback mechanisms at three levels: cortical microcircuit, corticothalamic loop, and cortico-BG-thalamocortical system. We focused on executive control in a stop signal task, which is known to depend on BG across species. The model reproduces a range of experimental observations and shows that the newly discovered feedback projection from external globus pallidus to striatum is crucial for inhibitory control. Moreover, stopping process is enhanced by the cortico-subcortical reverberatory dynamics underlying persistent activity, establishing interdependence between working memory and inhibitory control. Surprisingly, the stop signal reaction time (SSRT) can be adjusted by weights of certain connections but is insensitive to other connections in this complex circuit, suggesting novel circuit-based intervention for inhibitory control deficits associated with mental illness. Our model provides a unified framework for inhibitory control, decision making, and working memory.

INTRODUCTION

Across mammalian species, from rodent to monkey to human, a conserved brain system is the prefrontal cortex-basal ganglia (BG)-thalamic circuit. This cortico-subcortical system plays a critical role in diverse cognitive functions, including perceptual decision making (Ding and Gold, 2012; Forstmann et al., 2010; Grinband et al., 2006), inhibitory control (Aron et al., 2003, 2007b; Aron and Poldrack, 2006; Jahanshahi et al., 2015), and working memory (Floresco et al., 1999; Isseroff et al., 1982; Mills et al., 2012; Parnaudeau et al., 2013; Wang, 2001; Watanabe and Funahashi, 2012). A recent fMRI study revealed an overlapping cortical representation of executive control and working memory (Harding et al., 2016). It has also been found that degraded executive control function is usually associated with deficits in working memory in schizophrenia (Grégoire et al., 2012; Zandbelt et al., 2011). Intriguingly, in a task that engaged both perceptual decision making and inhibitory control, the two processes were found to be functionally independent of each other (Middlebrooks and Schall, 2014), raising the question of whether they could still share a common underlying circuit. Understanding the circuit mechanism for the interplay between these cognitive functions poses a challenge for both experimental and modeling studies.

Inhibitory control, the ability to cancel a response when a planned action becomes inappropriate, constitutes an essential part of executive function (Schall and Godlove, 2012; Stuphorn, 2015; Verbruggen and Logan, 2009). Impaired inhibition function has been found to be implicated in several neurological disorders such as Parkinson’s disease (PD) (Gauggel et al., 2004; Mirabella et al., 2012), attention-deficit hyperactivity disorder (Castelanos et al., 2006; McAlonan et al., 2009), schizophrenia (Hughes et al., 2012; Thakkar et al., 2011, 2015; Zandbelt et al., 2011), and also normal aging (Andrés et al., 2008; Coxon et al., 2012; Hu et al., 2014). The stop signal task, in which a subject is required to suppress a planned action upon the occurrence of an unexpected stop signal, provides a standard paradigm for investigating inhibitory control (Verbruggen and Logan, 2008, 2009). It allows for the estimation of the latency of a covert stop process, i.e., the stop signal reaction time (SSRT), according to the race model (Boucher et al., 2007; Logan and Cowan, 1984; Logan et al., 2014). SSRT measures the effectiveness of inhibitory control (the shorter the SSRT, the better one’s ability to suppress prepotent but inappropriate action), and it has been widely used to assess impaired inhibitory control in psychiatric patients. A recent experiment provides direct neurophysiological evidence for the involvement of the BG in the stop process: a transient surge of neural activity in the subthalamic nucleus (STN) upon a stop signal presentation and the corresponding increase of activity in the substantia nigra pars reticulata (SNr) in successful stop signal trials were well within the window of the SSRT (Schmidt et al., 2013). In pro/anti-saccade tasks that involve response inhibition, the higher-order thalamus (Th) also showed strong modulation during response initiation (Tanaka and Kunimatsu, 2011). Meanwhile, higher-order Th receives topographic inputs from the SNr (Gulcebi et al., 2012) and forms reciprocal connections with the frontal cortex (Alexander et al., 1986; Haber and Calzavara, 2009; Parent and Hazrati, 1995). How inhibitory control depends on the multiple pathways through the BG and...
corticothalamic loop is poorly understood. For instance, which connection pathways in this complex circuit determine SSRT remains unknown. Progress in this direction would provide insight into the operation of a core large-scale brain system as well as specific circuit deficits associated with mental illness.

In this work, we developed a physiologically based network model of spiking neurons with feedback mechanisms at three levels: local cortical microcircuit, corticothalamic loop, and cortico-BG-thalamocortical system. We found that the SSRT can be effectively modulated by some specific intrinsic BG connections; in particular, the newly discovered pathway from the external segment of the globus pallidus (GPe) to the striatum (Str) (Abdi et al., 2015; Dodson et al., 2015; Hegeman et al., 2016; Hernández et al., 2015; Mallet et al., 2012) plays an essential role in cancelling a planned response in midcourse, in support of the disinhibition to the Th; as a result, the stop signal can disrupt the ramping activity in the Str and therefore interfere with the disinhibition to the Th; as a result, the thalamic drive to the Cx may become insufficient for the initiation of the planned response in a successful stop signal trial. In the model, neural spiking activity is highly irregular; a response may or may not be cancelled by a stop signal probabilistically depending on the stochastic network dynamics in each single trial.

RESULTS

Distributed Attractor Network Model for Inhibitory Control
The distributed network model we have constructed includes the cortical circuit (Cx), the BG, and the Th, which form a closed loop (Figure 1). This closed loop underlies the inhibitory control function in our model, with a putative circuit as depicted in Figure 1A. The Cx shows ramping neuronal activity during movement initiation due to both local reverberation in the Cx and the excitatory thalamic drive to the Cx. In our model, a stop signal is presented as an input to the STN in the BG. The STN receives stop signal-related inputs from two sources: directly from the midbrain, e.g., pedunculopontine tegmental nucleus (PPN), which causes a rapid response in STN (Schmidt et al., 2013), and indirectly from medial frontal areas, which are related to inhibitory control (see Discussion). We did not explicitly model the frontal circuit involved in inhibitory control in the present model; therefore, the stop signal received in the STN represents a combination of both sources. After reaching the STN, the stop signal flows through two routes (Figure 1C). One is from the STN directly to the SNr, which provides a rapid suppression of response initiation by increasing the SNr activity to disrupt the disinhibition to the Th. The second route is from the STN to the Str through the GPe (Mallet et al., 2016; see Discussion). Through this route, the stop signal can disrupt the ramping activity in the Str and therefore interfere with the disinhibition to the Th; as a result, the thalamic drive to the Cx may become insufficient for the initiation of the planned response in a successful stop signal trial. In the model, neural spiking activity is highly irregular; a response may or may not be cancelled by a stop signal probabilistically depending on the stochastic network dynamics in each single trial.

To examine the mechanism of inhibitory control in our model, we simulated a stop signal task with a constant go signal and a constant stop signal, as usually used in monkey and human experiments. The neuronal activities of each area are shown in...
Figure 2, Single-Trial Simulations of the Stop Signal Task

Spike trains (raster plots) are shown for a go trial (black) and a successful stop signal trial (blue) in each area (A–F; upper panels). Population firing rates in each area (A–F; lower panels) are shown for a go trial (black), and successful (blue) and failed (green) stop signal trials. The schematic in (A) illustrates a constant go stimulus that is turned on at time 0 (black) and two SSDs for the stop signal at 120 ms (blue) and 170 ms (green), respectively.

Figure 2 for the go trial (black), successful (blue), and failed (non-cancelled; green) stop signal trials, respectively. The raster plots in each area represent spike trains of the population selective to the go stimulus in the go trial (black) and successful stop signal trial (blue). The lower panel of Figure 2A shows a schematic of the go stimulus and two stop signal delays (SSDs) (blue, 120 ms; green, 170 ms). In the go trial, the ramping activity in the Cx (Figure 2A) leads to an increase of Str activity (Figure 2D) and a decrease of the SNr activity (Figure 2C), which releases the inhibition to the Th (Figure 2F) and further facilitates the ramping activity in the Cx. The stop signal elevates the STN activity (Figure 2E). In a successful stop signal trial (blue), at the time of stop signal onset, the SNr activity is not suppressed too much by the striatal input yet, and the elevated STN activity can increase the SNr activity. This leads to decreased Th activity and terminates the ramping in the Cx. The stop signal flow through STN-to-GPe-to-Str increases the GPe activity (Figure 2B) and then reduces the Str activity, leading to a weakened go process. In a non-cancelled stop signal trial (green), the SNr activity is strongly suppressed by the striatal input at the time of the stop signal arrival. In this case, the elevated STN activity is insufficient to counteract the strong striatal input to the SNr and disrupt the disinhibition to the Th, and the response is not cancelled.

Successful stopping happens only in a fraction of the trials for a given SSD. In our model, the competition between two selective populations in the Cx results in variably ramping slopes of the winner population, which provides the main drive for the fluctuating neuronal activity in the Cx and the downstream circuits from trial to trial. In trials with faster cortical ramping, the SNr also decreases faster and is less likely to increase its activity at the occurrence of the stop signal due to the stronger inhibitory drive from the Str. The fraction of non-cancelled stop signal trials (normalized by the performance in go trials) gives the non-cancelled probability, which can be used to estimate the SSRT through the integration method (Boucher et al., 2007; Logan and Cowan, 1984). The idea is that SSRT quantifies the time needed for a stop process to “win the race” over the go process; therefore, non-cancelled trials are those with RTs shorter than the sum of SSD and SSRT; see Figure 3A and Supplemental Experimental Procedures, available online, for details. This suggests that the large-scale spiking network model of the cortico-BG-thalamocortical system we have introduced is behaviorally consistent with the race model (Boucher et al., 2007; Logan and Cowan, 1984). The non-cancelled probability as a function of the SSD defines the inhibition function, as shown in Figure 3B. The non-cancelled probability increases with the SSD, from completely cancelling at small SSD to totally failing to cancel at large SSD, as reported in numerous experiments (e.g., Hanes and Schall, 1995; Logan and Cowan, 1984). The mean neuronal activities for non-cancelled and cancelled stop signal trials, and the corresponding latency-matched fast and slow go trials, were shown in Figure S1. Here, latency-matched fast and slow go trials refer to those go trials with RTs shorter and longer than SSD plus SSRT, respectively. Consistent with recent experimental findings (Schmidt et al., 2013), the variable timing of the striatal activity for go trials determines whether the cancellation of response by the stop signal is successful or not (Figures S1D and S1G). Note that despite the similar activities between Cx and Str in our model (Figure S1), the cortical activity itself does not provide such a determinant since it only receives stop-related activity through the BG-to-Th-to-Cx route, while as suggested by both experiments and our modeling study (see next
SSRT Depends on Specific BG Circuit Connections

The stop signal in our model is conveyed through the hyperdirect pathway of the BG to the STN. To cancel a response, the activation of the stop signal on the SNr has to counterbalance the direct pathway input to the SNr. We found that the non-cancelled probability increases with direct pathway strength $g_{\text{Str-SNr}}$ (Figure 4A, left), since it is harder for the stop signal to counterbalance the direct pathway input to the SNr and disrupt the disinhibition to the Th when the strength of the direct pathway is stronger. The SSRT increases with $g_{\text{Str-SNr}}$ (Figure 4A, right), indicating a reduction of stopping function with the increase of the direct pathway strength. At the neuronal level, the SNr has a lower spontaneous firing rate and decreases faster during response initiation for a larger $g_{\text{Str-SNr}}$ (Figure S2A), increasing the chance that SNr activity drops to a critically low level by the onset of the stop signal.

We further checked how the backpropagation of the stop signal from the STN to the Str through the GPe might impact inhibitory control. When the strength of the GPe-to-Str connection, $g_{\text{GPe-Str}}$, is increased, we found that the non-cancelled probability is reduced (Figure 4B, left) and the SSRT is decreased (Figure 4B, right). On the other hand, in the absence of the GPe-to-Str connection, the model failed completely in performing inhibitory control, as indicated by the fact that the non-cancelled probabilities are near 1 (black curve in the left panel of Figure 4B; $g_{\text{GPe-Str}} = 0$). Note that for fair comparison, we adjusted the background input to the SNr and kept the spontaneous SNr firing rate unchanged for different $g_{\text{GPe-Str}}$ to distinguish from the impact of direct pathway strength $g_{\text{Str-SNr}}$, which influences the stopping function by modulating the SNr activity.

With a stronger $g_{\text{GPe-Str}}$, the stop signal backpropagating to the Str can more easily curtail the ramping activity there, which makes the SNr activity decrease slower, thereby facilitating the stopping function (Figure S2B). This result resonates with the recent experimental finding of an important role in the stopping of akrypallidal cells in the GPe that mediate this feedback projection to the Str (Mallet et al., 2016). The striatal activity therefore provides a determinant for stopping behavior, where the two routes of stop-related activity transmission, from GPe to Str and from SNr to Cx through Th, converge (Schmidt et al., 2013; see also Figures S1D and S1G). The inhibition functions for different $g_{\text{Str-SNr}}$ (Figures S3A and S3B) and $g_{\text{GPe-Str}}$ (Figures S3C and S3D) are aligned with each other when plotting against the z score of the relative finishing time (ZRFT) (Logan and Cowan, 1984).

We have also performed simulations for an extended circuit, in which the Str neurons are segregated into direct (D1-expressing) and indirect (D2-expressing) pathway neurons, and the GPe neurons are segregated into Ark and Pro types of neurons, respectively (Figure 4C). With this extended circuit, we performed simulations on the impact of Ark-to-Str back-projection on inhibitory control, in which we assumed symmetric projection from Ark neurons to D1- and D2-expressing striatal neurons since no such data are available yet (Hegeman et al., 2016). We found a similar impact as that shown in Figure 4B (Figure 4D). Therefore, segregation of the Str and GPe into sub-populations does not change the results we have obtained from the simplified circuit. We also showed that choosing a different level of decision threshold at the Cx (Figures S4A–S4E) or scaling the sparse GPe–STN connections to all-to-all connections (Figures S4F and S4G) does not change the results.

Self-Sustained Persistent Activity Facilitates Stopping Function

When excitatory feedback connections are sufficiently strong, a recurrent neural circuit can generate self-sustained persistent
In Figure 5B, we showed the neuronal activity of the winner population in the Cx for two representative samples in the state space (blue and green squares in Figure 5A, lower panel; \( w^+ = 1.7 \)) with \( g_{\text{Th-Cx}} = 0.26 \) nS (blue) and 0.32 nS (green), respectively. When a transient stimulus is shown for 500 ms (Figure 5B, lower panel), it triggers persistent activity for \( g_{\text{Th-Cx}} = 0.32 \) nS (green curve), but not for \( g_{\text{Th-Cx}} = 0.26 \) nS (blue curve). Therefore, by increasing the feedback strength of the big loop, the circuit can be switched from a state without persistent activity to a state that supports self-sustained persistent activity and therefore working memory, under the assumption that local reverberation in the Cx is insufficient to maintain persistent activity. We further note that when \( g_{\text{Th-Cx}} \) is above some critical value (\( \sim 0.2 \) nS here for \( w^+ = 1.7 \)), the Cx shows ramping neuronal activity during the presentation of the stimulus even though only the SS is stable after the stimulus offset. A local circuit with ramping activity but without persistent activity has been investigated in Wong and Wang (2006).

We compared inhibitory control function for these two cases. When there is self-sustained persistent activity, the inhibition function is left shifted (Figure 5C), indicating a faster go process due to a stronger Cx-BG-Th-Cx feedback loop, as compared to the case when there is no persistent activity. Meanwhile, the SSRT is significantly reduced when there is persistent activity (Figure 5D, colored points), indicating also a faster stop process. This suggests an enhancement of stopping function due to self-sustained persistent activity. To further elucidate the relationship between persistent activity dynamics and SSRT, we increased \( g_{\text{Th-Cx}} \) gradually from the regime without persistent activity to the regime supporting persistent activity with a fixed \( w^+ \) (\( w^+ = 1.7 \)). We observed a two-level profile for the SSRT: when \( g_{\text{Th-Cx}} \) is varied but without crossing the phase boundary (vertical dashed line in Figure 5D), the SSRT is insensitive to this change. There are two different levels of the SSRT, corresponding to the
states with and without persistent activity, when the thalamocortical projection strength is modulated. Note that adjusting $g_{th-Cx}$ modulates not only the positive feedback strength to the Cx, but also the strength of stop-related activity transmitting to the Cx. In Figures S5A–S5C, we showed that the same conclusion holds when there is a corticothalamic sub-loop. When the corticothalamic projection is not very strong (here $g_{Cx-Th}=0.2$ nS), the boundary between the regions “Stable SS” and “Bistable” is left-shifted, but the structure of the state space remains qualitatively the same as for the case without corticothalamic projection (Figure S5A, lower panel). For the two points indicated in the state space, the Cx shows persistent activity for one, but not for the other, when receiving a transient stimulus (Figure S5B). We found that the two-level modulation of SSRT by the thalamocortical connection strength still holds when there exists a corticothalamic sub-loop (Figure S5C).

It is interesting to check further how the SSRT depends on $w^+$. From Figure S5D, we found that the SSRT decreases monotonically with the increase of $w^+$, i.e., the stopping function is enhanced with stronger locally reverberatory dynamics. This differential modulation of the SSRT by $g_{th-Cx}$ and $w^+$ results from the fact that in addition to providing positive feedback to the Cx, the thalamocortical connection, but not the recurrent connection within the Cx, also transmits the stop-related activity that terminates ramping activity and motor response.

**Functional Independence of Perceptual Decision and Inhibitory Control**

Our model provides a framework to explore the interaction between inhibitory control and perceptual decision making when both processes are engaged in a task. To study this, we introduced a stop signal in the random-dot motion discrimination task, in which both stimulus-selective populations receive inputs, with the relative strength of the inputs encoding the motion coherence $c'$ (Wang, 2002; Wong and Wang, 2006). When there is no stop signal, the model performances and mean RTs as a function of $c'$ are shown in Figures 6A and 6B. For non-cancelled stop signal trials, the performances and mean RTs show similar trends as those for go trials (Figure S6). In stop signal trials, the non-cancelled probability increases with $c'$ and therefore decreases with the task difficulty (Figure 6C). However, the SSRT is insensitive to $c'$ (Figure 6D; one-way ANOVA, $F(4, 31) = 0.72, p = 0.59$). This suggests that the task difficulty, which strongly influences the perceptual decision process (Figures 6A and 6B), has no impact on the stop process. This can also be seen from Figure 6E, which shows that the mean neuronal activities of cancelled (successful) stop signal trials are almost indistinguishable for low and high coherences. It therefore indicates the functional independence between the choosing and stopping processes, consistent with a recent behavioral experiment (Middlebrooks and Schall, 2014). Figure S7 shows that the inhibition functions are aligned with each other for different $c'$ when plotted against the ZRFT.

Note that we have made no assumptions about how the stop signal is presented or how the response is executed. Our model therefore suggests that the functional independence between choosing and stopping processes generalizes across modalities of the stop signal and response. This can be tested, for instance, by performing motion discrimination task with auditory stop signal and/or key press response.
It has been noted that the corticothalamic connection strength, \( g_{\text{Cx}}/C_0 \), enhances SNr activity and diminishes ramping Cx activity. This is reflected by the very steep rising of Cx activity for trials with RTs located near the second peak. The RT distributions for different SSDs are shown in Figures S8A and S8B with \( g_{\text{Cx}}/C_0 = 0.4 \) nS. Interestingly, this computational finding is supported by empirical observations from experiments using rats (Schmidt et al., 2013).

Our model thus suggests that the bimodal RT distribution for rats performing a stop signal task might result from a strong corticothalamic sub-loop. Note that our model assumes that even when the sensory stimulus for signaling stop is transient, the internal representation of the stop signal is sustained. This assumption is crucial and remains to be tested in future experiments.

**Impact of Corticothalamic Sub-loop on Stopping Function**

Up to this point, we have not included the corticothalamic connection in the model (except in Figures S5A–S5C), so we can examine the cortico-BG-thalamocortical loop. Now we investigate the impact of such a connection on stopping function by including non-zero corticothalamic connection strength, \( g_{\text{Cx}}/C_0 \). This sub-loop makes it possible for the reverberatory dynamics to rebound in spite of transient suppression by a stop signal, leading to slow responses and a bimodal distribution of RTs in non-cancelled trials (Figure 7A, lower panel; \( g_{\text{Cx}}/C_0 = 0.4 \) nS). Note that the second peak has a totally different origin from the first one. While the first peak is driven by stimulus and reflects the most likely time for ramping Cx activity crossing the threshold, the second peak is driven mainly by noise when the constant stop signal enhances SNr activity and diminishes ramping Cx activity. This is reflected by the very steep rising of Cx activity for trials with RTs located near the second peak. The RT distributions for different SSDs are shown in Figures S8A and S8B with \( g_{\text{Cx}}/C_0 = 0.4 \) nS. Interestingly, this computational finding is supported by empirical observations from experiments using rats (Schmidt et al., 2013).

Our model thus suggests that the bimodal RT distribution for rats performing a stop signal task might result from a strong corticothalamic sub-loop. Note that our model assumes that even when the sensory stimulus for signaling stop is transient, the internal representation of the stop signal is sustained. This assumption is crucial and remains to be tested in future experiments.
We found that the SSRT is independent of perdirect pathway, and then interacts with the direct and indirect for executive control. A stop signal enters the BG through the hy-

dence between inhibitory control ability and working memory in patients. Another conceptual advance is to establish interdepen-
ture of how different connection pathways modulate SSRT. This

strained model with explanatory power revealed a complex pic-

Figure 7. Impact of Cx-Th Sub-loop on Inhibitory Control

(A) Lower panel: the RTs of go trials (black) and non-
cancelled stop signal trials (green). Note the double peaks in the distribution of the non-cancelled stop signal trials. Here the corticothalamic connection strength $g_{\text{cx-th}} = 0.4$ nS. Upper panel: modified integration method for estimating SSRT, illustrated with SSD = 220 ms. Black curve represents the cdf for RTs of the go trials, and green curve represents the cdf for the non-cancelled trials, scaled by the non-cancelled probability (blue dashed line). The dashed black curve represents the 99.9% confi-
dence interval (C) of the cdf for RTs of go trials. The vertical red dashed line represents the intersection time of the green curve with the black dashed curve. The red solid line indicates the SSD. The interval between the vertical solid and dashed red lines gives an estimate of the SSRT.

(B) Upper panel: schematic of the circuit. Middle panel: neuronal activities of the winner population in the Cx in response to a brief stimulus for $g_{\text{cx-th}} = 0$ (blue) and 0.4 nS (green), respectively. The gray-shaded area indicates neuronal activities during stimulus onset, which lasts for 0.5 s (lower panel). The stimulus is a random-dot input with $c^2 = 25.6\%$.

(C) SSRT is insensitive to $g_{\text{cx-th}}$. Error bars indicate SD of SSRT values estimated for different SSDs. In this figure, $g_{\text{th-cx}} = 0.26$ nS.

DISCUSSION

Inhibitory control of action is central for flexible behavior, charac-
terized by SSRT, which has been used to quantify inhibitory control ability as well its impairment in psychiatric illness. The BG are

known to play an important role in inhibitory control, but only recent work has begun to dissect its underlying circuit mecha-
nism in the BG. The spiking neural circuit model presented in this paper accounts for salient experimental observations including the bimodal distribution of RTs in a stop signal task (Schmidt et al., 2013), as well as the surprising finding that SSRT is independent of task difficulty in perceptual decision making (Middlebrooks and Schall, 2014). This biologically con-
strained model with explanatory power revealed a complex pic-
ture of how different connection pathways modulate SSRT. This is an important new insight into BG function, opening up new ex-
periments to test the model and ideas about potential circuit-
specific ways to remedy inhibitory control deficits in psychiatric patients. Another conceptual advance is to establish interdepen-
dence between inhibitory control ability and working memory in the cortico-BG-thalamocortical system.

In this work, we propose a distributed attractor network model for executive control. A stop signal enters the BG through the hyperdirect pathway, and then interacts with the direct and indirect pathways to subserve inhibitory control in stop signal tasks. This cortico-BG-thalamocortical system involves three feedback loops at multiple levels, and a main finding of this work is the complex dependence of the effectiveness of inhibitory control on various pathways, as summarized in Figure 8. The SSRT is increased when the “go” process becomes more potent by increasing the Str-to-SNr strength. On the other hand, the SSRT is decreased when the GPMe-to-Str connection is stronger, which permits more potent backpropagation of the stop signal to the Str. The SSRT is also decreased when the network is switched from a state without persistent activity to a state supporting persistent activity, establishing a link between working memory and inhibitory control. Our model offers an explanation for several recent experimen-
tal observations in stop signal tasks. First, the GPMe-to-Str feedback connection, mediated by arkyppallidal cells, is essential for inhibitory control (Mallet et al., 2016). Second, when inhibitory control is combined with perceptual decision making, SSRT is in-
dependent of the task difficulty (Middlebrooks and Schall, 2014). Third, our model suggests that a non-negligible corticothalamic sub-loop leads to a bimodal distribution of RTs for non-cancelled stop signal trials (Schmidt et al., 2013) and predicts that the strength of corticothalamic projection does not change SSRT. This model prediction can be tested in future experiments.

Previously, a closed cortico-BG-thalamocortical loop has been proposed to implement different cognitive functions. This loop has been suggested to implement working memory retrieval and response selection (Humphries and Gurney, 2002; Schroll et al., 2012; Vitay and Hamker, 2010). In another modeling study, unbalanced direct and hyperdirect pathways led to the loss of action selection and the generation of oscillatory activity (Leblois et al., 2006). Recently, the competition between hyperdirect and
indirect pathways has been investigated using a nested diffusion model to differentiate the stopping and go-no processes (Dunovan et al., 2015). Wiecki and Frank (2013) proposed a Cx-BG-Th-Cx loop model, including several frontal cortical areas that implemented different response inhibition tasks. None of these studies, however, have investigated how the effectiveness of inhibitory control, quantified by SSRT, depends on various connection pathways. In this work, we build a distributed attractor network model instantiated by the Cx-BG-Th-Cx closed loop, such that the local recurrent connection within the Cx is by itself not enough to maintain persistent activity. It is conceivable that persistent activity can be generated intrinsically within certain local areas (like the prefrontal cortex in monkey) but depends on the thalamocortical loop in other areas (such as the premotor/motor system in rodents; K. Svoboda, personal communication). Our model provides a general framework for inhibitory control, perceptual decision making, and working memory, and allows for investigating how the SSRT depends on connections and dynamic properties of the network.

The role of the BG in action control was commonly hypothesized to be exerted through disinhibition to downstream structures (Hikosaka et al., 2000; but see Turner and Desmurget, 2010 for an alternative view), although disinhibition may be context dependent (Goldberg et al., 2013). Interruption of the disinhibition mechanism of the BG lay in the core of our model for inhibitory control and has been supported by recent experiments (Schmidt et al., 2013). The indirect pathway of the BG has important roles in proactive inhibitory control (Majid et al., 2013) and perceptual decision making (Wei et al., 2015). Here we showed that modulation of the relative weights of the BG pathways could effectively modulate the SSRT, which provides a possible source for individual difference of the SSRT. Recently, goal-directed or habitual behavior in rodents was associated with latency changes for the activation of direct and indirect pathway striatal neurons (O’Hare et al., 2016). It has been found that in goal-directed response, the indirect pathway striatal neurons showed a shorter latency than the direct pathway neurons. The difference between the latencies is about 20 ms, which is much shorter than the time for evidence accumulation and the SSD and therefore unlikely to impact the process we investigated. The GPe has been found to also be involved in anti-saccade tasks (Yoshida and Tanaka, 2016). Recent experiments have identified two distinct classes of GPe neurons, with one class forming reciprocal connections with the Str (Ark type) and the other forming reciprocal connections with the STN (Pro type) (Dodson et al., 2015; Mallet et al., 2012). The input sources for stop-related activity of Ark GPe neurons are not clarified yet, with transmission of stop-related activity from slowly activated STN neurons to the GPe as a possible source (Mallet et al., 2016). In this work, we assumed a constant input to the STN representing the stop signal, which is a combination of rapid input directly from the midbrain (e.g., PPN) and slower input indirectly from the medial frontal areas, and that the stop signal received by the GPe is transmitted from the STN. Since how the two GPe classes interact with each other during behavioral tasks is not experimentally characterized yet, we consider one single class of GPe neurons that project both to the STN and the Str. Under this simplification, the GPe activity increases and the striatal activity decreases with the occurrence of the stop signal, consistent with recent experiments (Mallet et al., 2016). Similarly, we also simplified the two classes of striatal neurons expressing D1 and D2 receptors, respectively, to a homogeneous class due to their non-distinction in existing experiments (Ding and Gold, 2010; Schmidt et al., 2013). We have shown that taking into account the subclasses of striatal and GPe neurons does not qualitatively change results obtained from the simplified circuit, assuming homogeneous populations in these nuclei. Further experimental data about interaction between the two types of striatal and GPe neurons will help future models to reveal the functional roles of these segregations under normal and pathological states. Moreover, since our model is biophysically realistic with spiking neurons and synaptic dynamics, it can be used, in a separate study, to explore different dynamical operating regimes (such as spike-to-spike synchrony or population rhythms) and their impact on inhibitory control function, as well as differential contributions of various (such as AMPA versus NMDA) receptor types to executive control, bridging levels from molecules to circuit dynamics to behavior.

Another direction for future extensions of our model is to elucidate differential contributions of several frontal areas to inhibitory control, including the right inferior frontal cortex (rIFC) and pre-supplementary motor area (preSMA) (Aron, 2011; Jha et al., 2015; Swann et al., 2012), whose roles in stopping are still controversial. The rIFC might directly participate in inhibitory control and has been identified by recent studies (Swann et al., 2009), or be important for detection of the stop signal (Duann et al., 2009; Hampshire et al., 2010; Sharp et al., 2010). The preSMA is more related to rule-based, motivational, or contextual modulation (Isoda and Hikosaka, 2007; Jha et al., 2015; Ridderinkhof et al., 2004; Scangos and Stuphorn, 2010). These two frontal areas could interact with the BG by projecting to the STN (Aron et al., 2007a; Cavanagh et al., 2011; Isoda and Hikosaka, 2008), representing a frontal source of stop-related signal flowing to the BG whose strength correlates with efficacy in stopping for normal subjects (Forstmann et al., 2010).
Both the stimulus and the stop signal are constant. We simulated a two-alternative choice task and considered two types of stimuli. The first kind is a go stimulus, indicating the position of the target (left or right) for the response as used in the classical stop signal task (Boucher et al., 2007; Logan and Cowan, 1984). The second one is a random-dot stimulus, as used in the motion discrimination task (Newsome et al., 1989; Roitman and Shadlen, 2002). A correct go trial or failed (non-cancelled) stop signal trial is defined as a trial in which the firing rate of the selective population in the Cx crosses a threshold at 15 Hz. A successful stop signal trial is defined as a trial in which the firing rate of the selective population in the Cx never crosses the threshold. Note that the results in the paper do not depend on the choice of threshold value in the Cx (Figures S4A–S4E). When there is no stop signal, model performance is defined as the fraction of correct go trials. In stop signal trials, the non-cancelled probability is defined as the ratio between the fraction of non-cancelled stop signal trials and the performance when there is no stop signal. This definition ensures that the non-cancelled probability approaches 1 for large SSD.

Network Modeling
The full circuit includes three brain areas: the Cx, the BG, and the Th. The Cx is the same as described in previous work (Wang, 2002), but with a lower background input to support distributed attractor state. The present model extends an earlier one (Wei et al., 2015) by including the Th and replacing the effect of SC in that model by a constant threshold in the Cx. We used the leaky integrate-and-fire model for all the neurons.

SUPPLEMENTAL INFORMATION
Supplemental Information includes Supplemental Experimental Procedures and eight figures and can be found with this article online at http://dx.doi.org/10.1016/j.neuron.2016.10.031.

ACKNOWLEDGMENTS
This work was supported by a Swartz Foundation Fellowship (W.W.), the NIH grant R01 MH62349, and the Naval Research grant N00014-13-1-0297 (K.-J.W.). We thank J. Rubin for insightful discussion, and F. Song and J. Murray for comments and carefully reading an early version of the manuscript.

REFERENCES


Jahfari, S., W aldorp, L., van den Wildenberg, W.P.M., Scholte, H.S., Ridderinkhof, K.R., and Forstmann, B.U. (2011). Effective connectivity reveals important roles for both the hyperdirect (fronto-subthalamic) and the indirect


