Article

Neuron

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.

Wei & Wang, 2016, Neuron *92*, 1–13
December 7, 2016
http://dx.doi.org/10.1016/j.neuron.2016.10.031

CellPress

Please cite this article in press as: Wei and Wang, Inhibitory Control in the Cortico-Basal Ganglia-Thalamocortical Loop: Complex Regulation and Interplay with Memory and Decision Processes, Neuron (2016), http://dx.doi.org/10.1016/j.neuron.2016.10.031

Neuron
Article

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

Wei Wei[1] and Xiao-Jing Wang[1,2,3,*]
[1]Center for Neural Science, New York University, New York, NY 10003, USA
[2]NYU-ECNU Institute of Brain and Cognitive Science, NYU Shanghai, 200122 Shanghai, China
[3]Lead 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 (Castellanos 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 reticulate (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

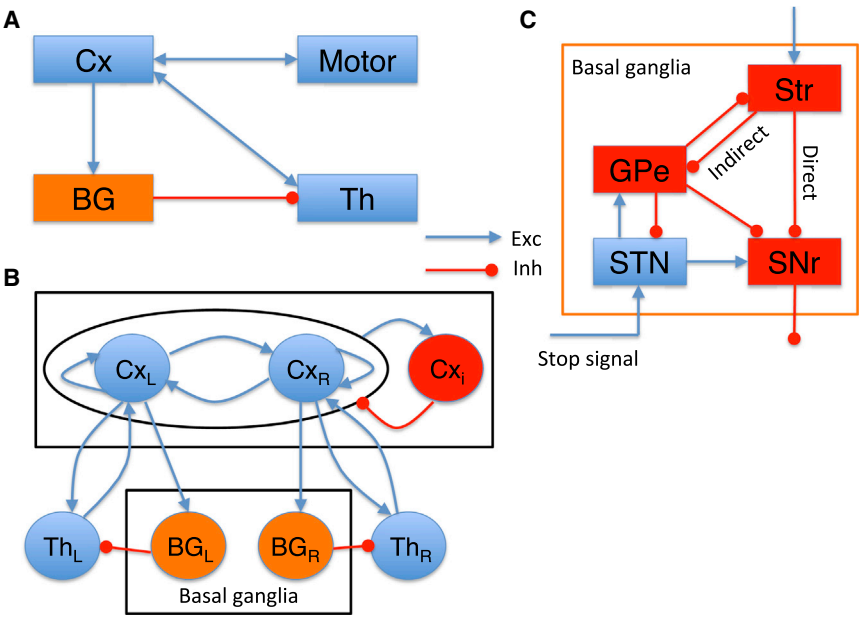

Figure 1. Schematic of the Cortico-Basal Ganglia-Thalamocortical Loop Model

(A) Circuit for inhibitory control implemented by the Cx-BG-Th loop.

(B) The zoomed-in circuit diagram from (A). In each local area, there are two neural populations selective for targets (L, left; R, right). This setting is suitable for simulating a two-alternative choice task. The model is endowed with feedback loops at three levels: local cortical recurrent network, thalamocortical loop, and global cortico-BG-thalamocortical loop.

(C) The zoomed-in BG circuit diagram. The Str projects to the output nucleus (SNr) through two pathways: the direct pathway (denoted as "Direct") from the Str directly to the SNr, and the indirect pathway (denoted as "Indirect") from the Str through the GPe, STN to the SNr. For simplicity, we use SNr to represent both SNr and GPi, the two output nuclei of the BG. The stop signal originates from a separate cortical and/or subcortical area and is simplified as a direct input to the STN, entering the BG via the hyperdirect pathway. Note that the projection from the GPe to the Str allows the stop signal to be fed back to the Str. Cx_i represents the inhibitory population in the Cx.

Cx, cortical circuit; BG, basal ganglia; Str, striatum; GPe, external segment of the globus pallidus (GP); GPi, internal segment of the GP; STN, subthalamic nucleus; SNr, substantia nigra pars reticulata; Motor, cortical/subcortical motor area; Exc, excitatory; Inh, inhibitory.

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 recent physiological report (Mallet et al., 2016). The thalamocortical connection exerts a two-level modulation of SSRT by switching the distributed attractor network between states with or without persistent activity. On the other hand, the corticothalamic feedback connection does not affect SSRT, even though it gives rise to a bimodal distribution of reaction times (RTs), and SSRT is independent of task difficulty when both inhibitory control and perceptual decision making are engaged in a task. These results revealed a complex picture of how SSRT is regulated in the cortico-BG-thalamocortical system.

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 drive from the Th. The Cx-BG-Th-Cx network forms a positive feedback loop (Figure 1B), which is essential for the ramping in the Cx in the model. 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

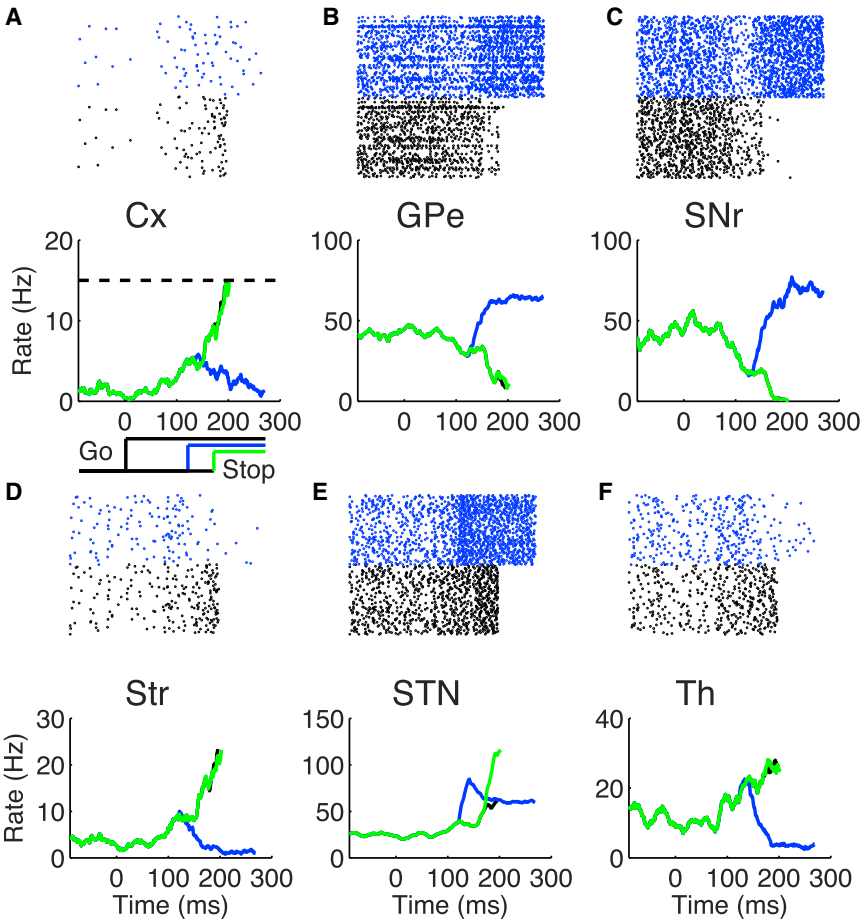

Figure 2. Single-Trial Simulations of the Stop Signal Task

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 (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 flowing 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 (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

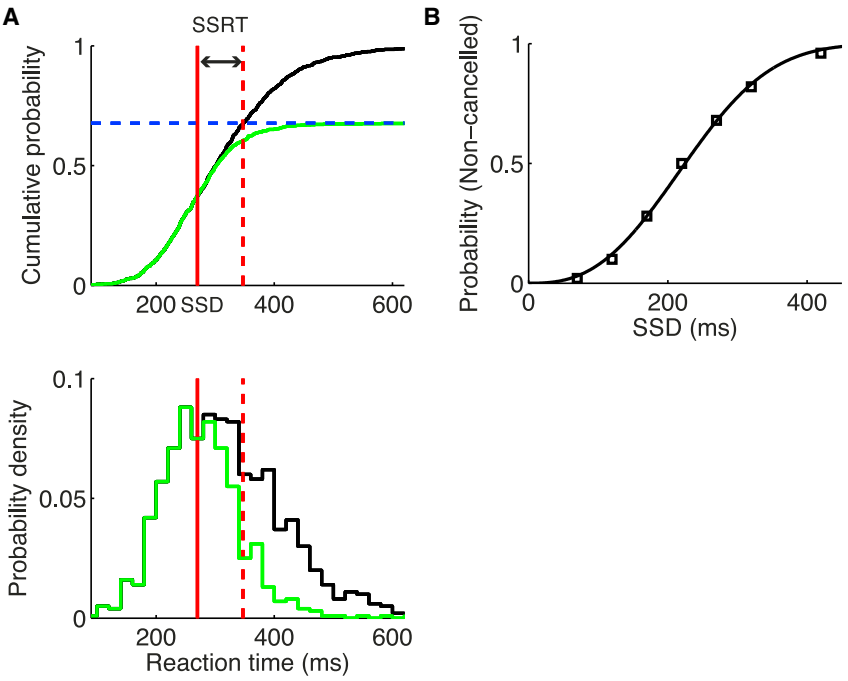

Figure 3. SSRT and Inhibition Function from the Model

(A) Upper panel: integration method for estimating the SSRT. The black curve represents the cumulative density function (cdf) for the reaction times (RTs) of go trials, and the green curve represents the cdf for the RTs of non-cancelled stop signal trials with SSD = 270 ms, scaled by the overall non-cancelled probability (fraction of non-cancelled stop signal trials, blue dashed line). The vertical red dashed line represents the intersection time of the blue dashed line and the black curve. The solid red line indicates the SSD. The interval between the vertical solid and dashed red lines gives an estimate of the SSRT. Lower panel: RT probability densities for the go trials (black curve) and non-cancelled stop signal trials (green curve; scaled by non-cancelled probability). (B) The non-cancelled probability as a function of the SSD (the inhibition function), which is fitted by the Weibull function.

section), the GPe-to-Str route for stop signal transmission is essential for successful stopping.

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 $g_{Str-SNr}$ (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 $g_{Str-SNr}$ (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 $g_{Str-SNr}$ (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, $g_{GPe-Str}$, 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; $g_{GPe-Str}$ = 0). Note that for fair comparison, we adjusted the background input to the SNr and kept the spontaneous SNr firing rate unchanged for different $g_{GPe-Str}$, to distinguish from the impact of direct pathway strength $g_{Str-SNr}$, which influences the stopping function by modulating the SNr activity.

With a stronger $g_{GPe-Str}$, 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 arkypallidal 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 $g_{Str-SNr}$ (Figures S3A and S3B) and $g_{GPe-Str}$ (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

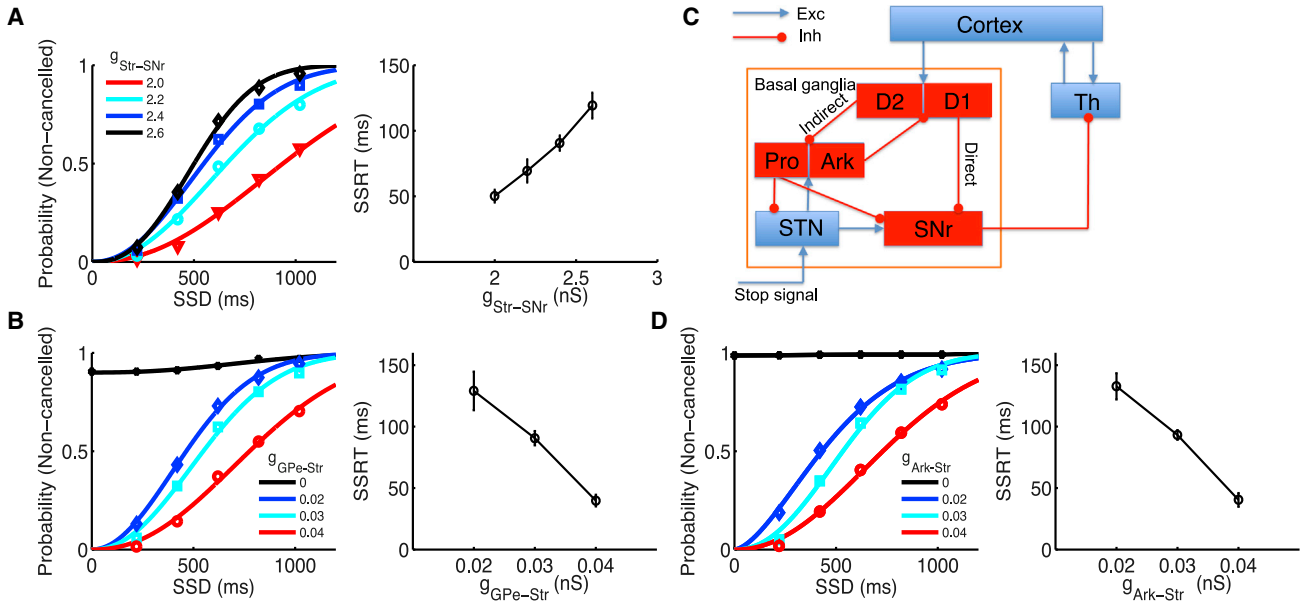

Figure 4. SSRT Depends on Weights of Specific Connections in the BG

(A) Left: dependence of inhibition functions on $g_{Str-SNr}$. Right: the SSRT increases with a larger $g_{Str-SNr}$.

(B) Left: dependence of inhibition functions on $g_{GPe-Str}$. Right: the SSRT decreases with a larger $g_{GPe-Str}$.

(C) Schematic of the extended BG circuit. 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. The Ark neurons project back to both types of Str neurons, and the Pro neurons project to the STN and SNr. The indirect pathway striatal neurons and STN neurons project to both Ark and Pro neurons.

(D) Left: dependence of inhibition functions on the back-projection strength from the Ark neurons to the Str, $g_{Ark-Str}$. Right: the SSRT decreases with a larger $g_{Ark-Str}$.

Error bars indicate SD of SSRT values estimated for each SSD. Note that in (B) and (D), for fair comparison when $g_{GPe-Str}$ or $g_{Ark-Str}$ is varied, the spontaneous firing rate of the SNr is fixed by adjusting the external input rate υ_{ext} to it. The inhibition functions in (A), (B), and (D) are fitted by the Weibull function.

activity (attractor states) (Wang, 2001). The existence of an attractor state in our model is determined by both the local recurrent strength w^+ in the Cx and the feedback strength of the cortico-subcortical loop, for instance, g_{Th-Cx} (see a schematic of the closed Cx-BG-Th loop in Figure 5A, upper panel). We assessed self-sustained activity states of the model system (in the absence of external input) as a function of w^+ and g_{Th-Cx} (Figure 5A, lower panel). In the left corner, the spontaneous state (SS) is the only stable state, and the neuronal activity of the selective population decays back to the SS after the stimulus is offset. The upper right corner represents a bistable region, where both the SS and the elevated state are stable. When g_{Th-Cx} increases further for a given w^+ (region not shown), the SS becomes unstable and both stimulus-selective (for left and right, L and R) populations can be found in that state. Note that the bistable region exists only when w^+ is larger than some critical value, such that the competition between the two (L and R) populations in the Cx is strong enough to implement the winner-take-all mechanism (Wang, 2002). 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 $g_{Th-Cx} = 0.26$ 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 $g_{Th-Cx} = 0.32$ nS (green curve), but not for $g_{Th-Cx} = 0.26$ 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_{Th-Cx} is above some critical value (~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_{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_{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

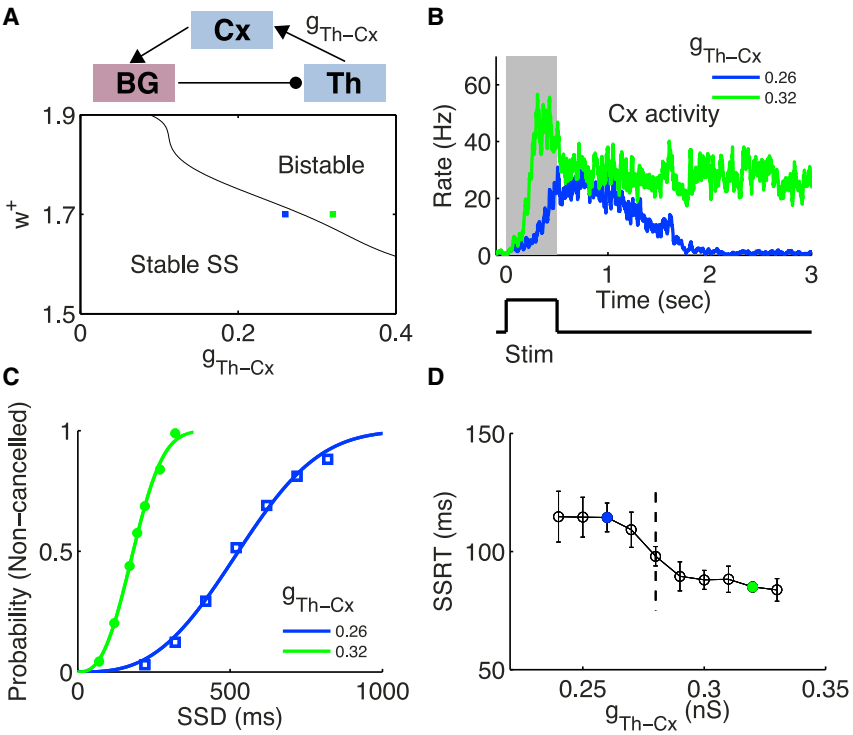

Figure 5. Inhibitory Control Measured by SSRT Is Enhanced by Stronger Reverberatory Dynamics

(A) Upper panel: schematic of the Cx-BG-Th loop. Lower panel: the state diagram as a function of the local recurrent feedback strength w^+ in the Cx and the subcortico-cortical loop feedback strength g_{Th-Cx}. Stable SS, region where only the spontaneous state (SS) is stable; Bistable, region where both the SS and target-selective elevated state are stable. Two sample networks are indicated in the state diagram at g_{Th-Cx} = 0.26 nS (blue) and 0.32 nS (green) with w^+ = 1.7.

(B) Neuronal activity of the winner population in the Cx in response to a brief stimulus. The green and blue curves correspond to the green and blue squares in (A). The gray-shaded area shows the neuronal activities during stimulus onset, which lasts for 0.5 s (lower panel).

(C) Inhibition functions for g_{Th-Cx} = 0.26 nS and 0.32 nS, respectively.

(D) Two-level modulation of the SSRT when increasing g_{Th-Cx} gradually from a regime not supporting persistent activity to a regime supporting persistent activity with a fixed w^+ (w^+ = 1.7). The colored points correspond to those in (A). The vertical dashed line indicates the value of g_{Th-Cx} when crossing the boundary for the two regimes.

Error bars indicate SD of SSRT values estimated for different SSDs. The inhibition functions in (C) are fitted by the Weibull function.

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.

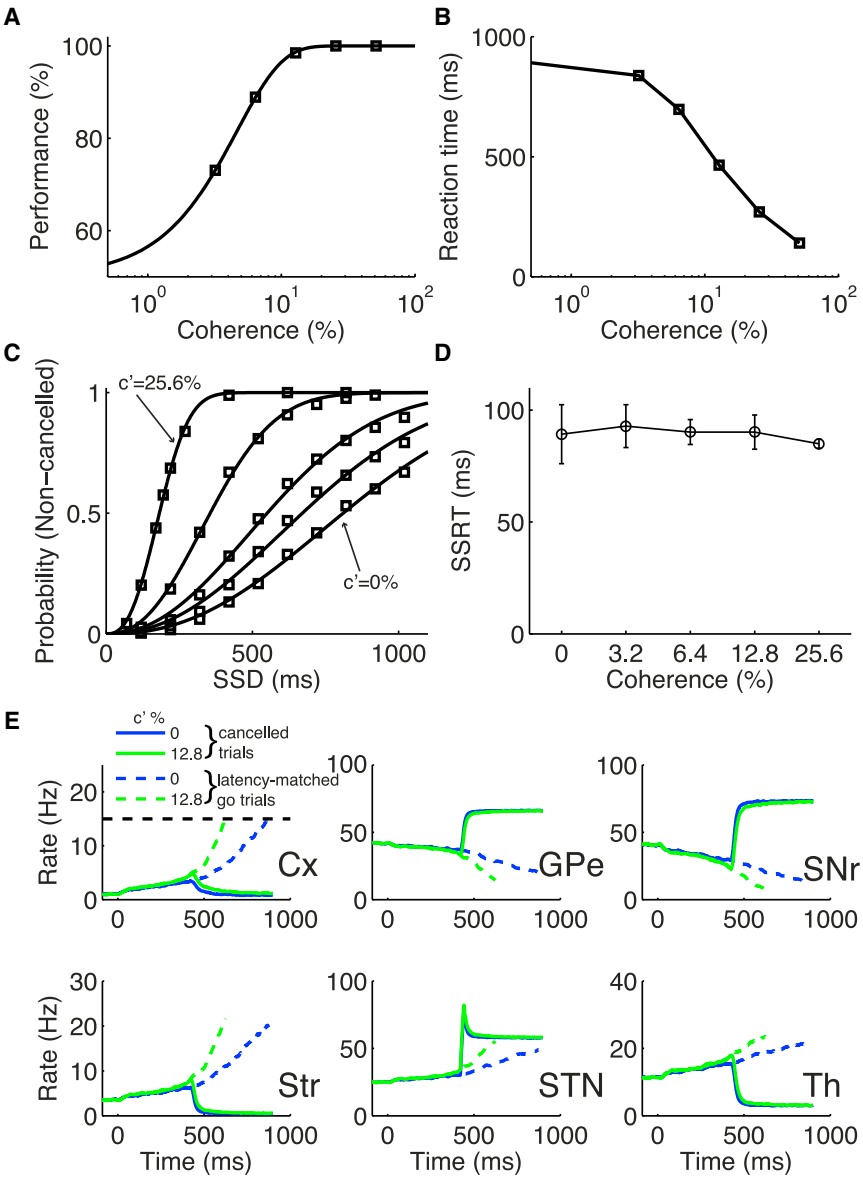

Figure 6. SSRT Is Independent of Task Difficulty in Perceptual Decision Making

(A and B) Performances (A) and mean RTs (B) from model simulation of a random-dot motion direction discrimination task.

(C) Inhibition functions for different task difficulty represented by coherence of the random moving dot input, c'. The values of c' from bottom to top are 0%, 3.2%, 6.4%, 12.8%, and 25.6%. The inhibition functions are fitted by the Weibull function.

(D) The SSRT as a function of c', showing no statistically significant dependence. Error bars indicate SD of SSRT values estimated for those SSDs with corresponding non-cancelled probabilities within the interval (0.1, 0.9).

(E) Mean neuronal activities for successful stop signal trials (solid curves) and latency-matched go trials (dashed curves) with SSD = 420 ms when c' = 0% (blue) and 12.8% (green). The mean activities are obtained from averaging over 200 trials of the same types.

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_{Cx-Th}. 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_{Cx-Th} = 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_{Cx-Th} = 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.

Figure 7A (upper panel) shows the cumulative density function (cdf) of RTs. The classical integration method for estimating SSRT is no longer applicable, since there is a substantial fraction of non-cancelled stop signal trials (with long RTs) for which there are no latency-matched go trials. We therefore estimated the SSRT using a modified integration method (Mayse et al., 2014), where the divergent point of the cdf for non-cancelled stop signal trials and go trials was used to estimate the SSRT (see Supplemental Experimental Procedures for details).

We then investigated how the strength of the Cx-Th sub-loop influences the SSRT. Figure 7B (upper panel) shows a schematic of the circuit. We increased g_{Cx-Th} from 0 to 0.6 nS and estimated the SSRT for each value of g_{Cx-Th} with g_{Th-Cx} = 0.26 nS, for which there is no persistent activity in the network (Figure 7B, middle panel) when receiving a transient stimulus (Figure 7B, lower panel). As shown in Figure 7C, the SSRT is independent of

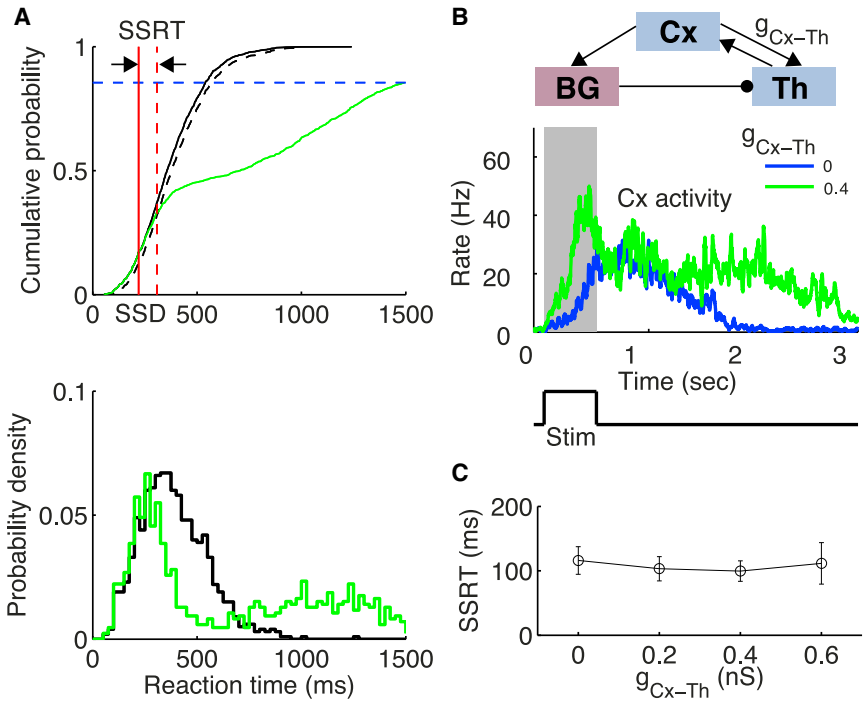

Figure 7. Impact of Cx-Th Sub-loop on Inhibitory Control

(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 $g_{Cx-Th} = 0.4$ 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% confidence interval (CI) 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 $g_{Cx-Th} = 0$ (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 $c' = 25.6\%$.

(C) SSRT is insensitive to g_{Cx-Th}.

Error bars indicate SD of SSRT values estimated for different SSDs. In this figure, $g_{Th-Cx} = 0.26$ nS.

g_{Cx-Th} (one-way ANOVA, F(3, 12) = 0.38, p = 0.77). Note that the same conclusion also holds when there is persistent activity in the circuit. In Figure S8C, the Cx circuit shows persistent activity with $g_{Th-Cx} = 0.32$ nS (upper panel) when receiving a transient stimulus (lower panel). We found that the SSRT is independent of g_{Cx-Th} (Figure S8D; one-way ANOVA, F(2, 9) = 0.54, p = 0.60).

DISCUSSION

Inhibitory control of action is central for flexible behavior, characterized 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 mechanism 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 constrained model with explanatory power revealed a complex picture of how different connection pathways modulate SSRT. This is an important new insight into BG function, opening up new experiments 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 interdependence 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 GPe-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 experimental observations in stop signal tasks. First, the GPe-to-Str feedback connection, mediated by arkypallidal cells, is essential for inhibitory control (Mallet et al., 2016). Second, when inhibitory control is combined with perceptual decision making, SSRT is independent 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, imbalanced 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

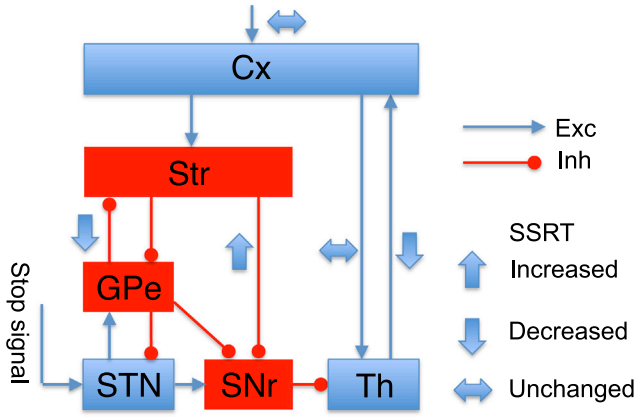

Figure 8. Summary of the Dependence of SSRT on Various Interarea Connections or the External Stimulus

The SSRT may be increased (wide up arrows), decreased (wide down arrows), or unchanged (wide horizontal arrows) by an increase of the connection strength from one area to another area, or of the task difficulty reflected in the input to the Cx. Other conventions are the same as in Figure 1.

indirect pathways has been investigated using a nested diffusion model to differentiate the stopping and no-go 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 process by projecting to the BG (Aron et al., 2007a, 2014; 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.,

2012) and deficits in inhibitory control during aging (Coxon et al., 2012). They can also selectively activate the indirect pathway striatal neurons (Ghahremani et al., 2012; Jahfari et al., 2011). Fixation neurons in the frontal eye field (FEF) have been found to show activity that resembles the stop process (Boucher et al., 2007; Hanes et al., 1998). A local network mechanism for inhibitory control through fixation neurons in the FEF accompanied by a top-down control signal has been proposed (Lo et al., 2009). Primary motor cortex also displayed neuronal activity related to action inhibition (Stinear et al., 2009), where fixation-like neurons were observed recently (Zagha et al., 2015). Other subcortical areas, such as the superior colliculus (SC) (Paré and Hanes, 2003) and basal forebrain (Mayse et al., 2015), have also been found to be involved in inhibitory control. Whether these areas play a causal role or are downstream of a core inhibitory control circuit remains to be better understood in the future. Ultimately, a large-scale brain circuit model is needed in order to systematically investigate multiple input sources for any given brain region and how interconnections between many cortical and subcortical structures actually work.

Both the thalamocortical and corticostriatal dysfunctions have been related to cognitive deficits in mental disorders (Anticevic et al., 2014; Grégoire et al., 2012; Parnaudeau et al., 2013; Zandbelt et al., 2011). Our model predicts that adjusting thalamocortical strength leads to a two-level modulation of the SSRT, which could be tested experimentally by selective optogenetic inactivation of neurons in corresponding thalamic nuclei (Paz et al., 2013). Adjusting the corticostriatal strength, on the contrary, will lead to a gradual modulation of the SSRT, since that does not change the maximal drive from the Th to the Cx and influence the existence of persistent activity. Impact of corticostriatal projection on inhibitory control could be exerted through its efficacy in modulating decision threshold level (Lo and Wang, 2006): a degraded corticostriatal projection will lead to a higher decision threshold, which then increases the SSRT (Figures S4A–S4E). The distinct impact on the SSRT induced by adjusting thalamocortical and corticostriatal connections, i.e., two-level versus gradual modulation, could be a differential signature for the involvement of these two loci. Inasmuch as distinct pathways may be impaired in different psychiatric patient groups (e.g., schizophrenia versus obsessive-compulsive disorder), the identification of specific connections that most effectively modulate SSRT could suggest better loci of "circuit-based" treatment for improving inhibitory control.

In conclusion, our distributed attractor network model offers a unified framework for investigating inhibitory control, perceptual decision making, and working memory. This model provides experimentally testable predictions about interaction among these functions. It can also be applied to the study of neurological diseases such as PD and schizophrenia, and shed further insights about deficits in the executive functions from a distributed attractor network point of view.

EXPERIMENTAL PROCEDURES

Behavioral Task Simulation

We considered a classical paradigm for inhibitory control, the stop signal task. 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.

For full details on the experimental procedures, please refer to the Supplemental Experimental Procedures.

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.

AUTHOR CONTRIBUTIONS

W.W. and X.-J.W. designed the research. W.W. performed the research and analyzed the data. W.W. and X.-J.W. wrote the paper.

ACKNOWLEDGMENTS

This work was supported by a Swartz Foundation Fellowship (W.W.), the NIH grant R01 MH062349, and the Naval Research grant N00014-13-1-0297 (X.-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.

Received: February 9, 2016
Revised: September 12, 2016
Accepted: October 12, 2016
Published: November 17, 2016

REFERENCES

Abdi, A., Mallet, N., Mohamed, F.Y., Sharott, A., Dodson, P.D., Nakamura, K.C., Suri, S., Avery, S.V., Larvin, J.T., Garas, F.N., et al. (2015). Prototypic and arkypallidal neurons in the dopamine-intact external globus pallidus. J. Neurosci. *35*, 6667–6688.

Alexander, G.E., DeLong, M.R., and Strick, P.L. (1986). Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu. Rev. Neurosci. *9*, 357–381.

Andrés, P., Guerrini, C., Phillips, L.H., and Perfect, T.J. (2008). Differential effects of aging on executive and automatic inhibition. Dev. Neuropsychol. *33*, 101–123.

Anticevic, A., Cole, M.W., Repovs, G., Murray, J.D., Brumbaugh, M.S., Winkler, A.M., Savic, A., Krystal, J.H., Pearlson, G.D., and Glahn, D.C. (2014). Characterizing thalamo-cortical disturbances in schizophrenia and bipolar illness. Cereb. Cortex *24*, 3116–3130.

Aron, A.R. (2011). From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. Biol. Psychiatry *69*, e55–e68.

Aron, A.R., and Poldrack, R.A. (2006). Cortical and subcortical contributions to Stop signal response inhibition: role of the subthalamic nucleus. J. Neurosci. *26*, 2424–2433.

Aron, A.R., Schlaghecken, F., Fletcher, P.C., Bullmore, E.T., Eimer, M., Barker, R., Sahakian, B.J., and Robbins, T.W. (2003). Inhibition of subliminally primed responses is mediated by the caudate and thalamus: evidence from functional MRI and Huntington's disease. Brain *126*, 713–723.

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

Aron, A.R., Durston, S., Eagle, D.M., Logan, G.D., Stinear, C.M., and Stuphorn, V. (2007b). Converging evidence for a fronto-basal-ganglia network for inhibitory control of action and cognition. J. Neurosci. *27*, 11860–11864.

Aron, A.R., Robbins, T.W., and Poldrack, R.A. (2014). Inhibition and the right inferior frontal cortex: one decade on. Trends Cogn. Sci. *18*, 177–185.

Boucher, L., Palmeri, T.J., Logan, G.D., and Schall, J.D. (2007). Inhibitory control in mind and brain: an interactive race model of countermanding saccades. Psychol. Rev. *114*, 376–397.

Castellanos, F.X., Sonuga-Barke, E.J., Milham, M.P., and Tannock, R. (2006). Characterizing cognition in ADHD: beyond executive dysfunction. Trends Cogn. Sci. *10*, 117–123.

Cavanagh, J.F., Wiecki, T.V., Cohen, M.X., Figueroa, C.M., Samanta, J., Sherman, S.J., and Frank, M.J. (2011). Subthalamic nucleus stimulation reverses mediofrontal influence over decision threshold. Nat. Neurosci. *14*, 1462–1467.

Coxon, J.P., Van Impe, A., Wenderoth, N., and Swinnen, S.P. (2012). Aging and inhibitory control of action: cortico-subthalamic connection strength predicts stopping performance. J. Neurosci. *32*, 8401–8412.

Ding, L., and Gold, J.I. (2010). Caudate encodes multiple computations for perceptual decisions. J. Neurosci. *30*, 15747–15759.

Ding, L., and Gold, J.I. (2012). Separate, causal roles of the caudate in saccadic choice and execution in a perceptual decision task. Neuron *75*, 865–874.

Dodson, P.D., Larvin, J.T., Duffell, J.M., Garas, F.N., Doig, N.M., Kessaris, N., Duguid, I.C., Bogacz, R., Butt, S.J., and Magill, P.J. (2015). Distinct developmental origins manifest in the specialized encoding of movement by adult neurons of the external globus pallidus. Neuron *86*, 501–513.

Duann, J.R., Ide, J.S., Luo, X., and Li, C.S. (2009). Functional connectivity delineates distinct roles of the inferior frontal cortex and presupplementary motor area in stop signal inhibition. J. Neurosci. *29*, 10171–10179.

Dunovan, K., Lynch, B., Molesworth, T., and Verstynen, T. (2015). Competing basal ganglia pathways determine the difference between stopping and deciding not to go. eLife *4*, e08723.

Floresco, S.B., Braaksma, D.N., and Phillips, A.G. (1999). Thalamic-cortical-striatal circuitry subserves working memory during delayed responding on a radial arm maze. J. Neurosci. *19*, 11061–11071.

Forstmann, B.U., Anwander, A., Schäfer, A., Neumann, J., Brown, S., Wagenmakers, E.J., Bogacz, R., and Turner, R. (2010). Cortico-striatal connections predict control over speed and accuracy in perceptual decision making. Proc. Natl. Acad. Sci. USA *107*, 15916–15920.

Forstmann, B.U., Keuken, M.C., Jahfari, S., Bazin, P.L., Neumann, J., Schäfer, A., Anwander, A., and Turner, R. (2012). Cortico-subthalamic white matter tract strength predicts interindividual efficacy in stopping a motor response. Neuroimage *60*, 370–375.

Gauggel, S., Rieger, M., and Feghoff, T.A. (2004). Inhibition of ongoing responses in patients with Parkinson's disease. J. Neurol. Neurosurg. Psychiatry *75*, 539–544.

Ghahremani, D.G., Lee, B., Robertson, C.L., Tabibnia, G., Morgan, A.T., De Shetler, N., Brown, A.K., Monterosso, J.R., Aron, A.R., Mandelkern, M.A., et al. (2012). Striatal dopamine D_2/D_3 receptors mediate response inhibition and related activity in frontostriatal neural circuitry in humans. J. Neurosci. *32*, 7316–7324.

Goldberg, J.H., Farries, M.A., and Fee, M.S. (2013). Basal ganglia output to the thalamus: still a paradox. Trends Neurosci. *36*, 695–705.

Grégoire, S., Rivalan, M., Le Moine, C., and Dellu-Hagedorn, F. (2012). The synergy of working memory and inhibitory control: behavioral, pharmacological and neural functional evidences. Neurobiol. Learn. Mem. *97*, 202–212.

Grinband, J., Hirsch, J., and Ferrera, V.P. (2006). A neural representation of categorization uncertainty in the human brain. Neuron *49*, 757–763.

Gulcebi, M.I., Ketenci, S., Linke, R., Hacıoğlu, H., Yanalı, H., Veliskova, J., Moshé, S.L., Onat, F., and Çavdar, S. (2012). Topographical connections of the substantia nigra pars reticulata to higher-order thalamic nuclei in the rat. Brain Res. Bull. *87*, 312–318.

Haber, S.N., and Calzavara, R. (2009). The cortico-basal ganglia integrative network: the role of the thalamus. Brain Res. Bull. *78*, 69–74.

Hampshire, A., Chamberlain, S.R., Monti, M.M., Duncan, J., and Owen, A.M. (2010). The role of the right inferior frontal gyrus: inhibition and attentional control. Neuroimage *50*, 1313–1319.

Hanes, D.P., and Schall, J.D. (1995). Countermanding saccades in macaque. Vis. Neurosci. *12*, 929–937.

Hanes, D.P., Patterson, W.F., 2nd, and Schall, J.D. (1998). Role of frontal eye fields in countermanding saccades: visual, movement, and fixation activity. J. Neurophysiol. *79*, 817–834.

Harding, I.H., Harrison, B.J., Breakspear, M., Pantelis, C., and Yücel, M. (2016). Cortical representations of cognitive control and working memory are dependent yet non-interacting. Cereb. Cortex *26*, 557–565.

Hegeman, D.J., Hong, E.S., Hernández, V.M., and Chan, C.S. (2016). The external globus pallidus: progress and perspectives. Eur. J. Neurosci. *43*, 1239–1265.

Hernández, V.M., Hegeman, D.J., Cui, Q., Kelver, D.A., Fiske, M.P., Glajch, K.E., Pitt, J.E., Huang, T.Y., Justice, N.J., and Chan, C.S. (2015). Parvalbumin+ neurons and Npas1+ neurons are distinct neuron classes in the mouse external globus pallidus. J. Neurosci. *35*, 11830–11847.

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

Hu, S., Chao, H.H., Zhang, S., Ide, J.S., and Li, C.S. (2014). Changes in cerebral morphometry and amplitude of low-frequency fluctuations of BOLD signals during healthy aging: correlation with inhibitory control. Brain Struct. Funct. *219*, 983–994.

Hughes, M.E., Fulham, W.R., Johnston, P.J., and Michie, P.T. (2012). Stop-signal response inhibition in schizophrenia: behavioural, event-related potential and functional neuroimaging data. Biol. Psychol. *89*, 220–231.

Humphries, M.D., and Gurney, K.N. (2002). The role of intra-thalamic and thalamocortical circuits in action selection. Network *13*, 131–156.

Isoda, M., and Hikosaka, O. (2007). Switching from automatic to controlled action by monkey medial frontal cortex. Nat. Neurosci. *10*, 240–248.

Isoda, M., and Hikosaka, O. (2008). Role for subthalamic nucleus neurons in switching from automatic to controlled eye movement. J. Neurosci. *28*, 7209–7218.

Isseroff, A., Rosvold, H.E., Galkin, T.W., and Goldman-Rakic, P.S. (1982). Spatial memory impairments following damage to the mediodorsal nucleus of the thalamus in rhesus monkeys. Brain Res. *232*, 97–113.

Jahanshahi, M., Obeso, I., Rothwell, J.C., and Obeso, J.A. (2015). A fronto-striato-subthalamic-pallidal network for goal-directed and habitual inhibition. Nat. Rev. Neurosci. *16*, 719–732.

Jahfari, S., Waldorp, 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

(fronto-striatal-pallidal) fronto-basal ganglia pathways during response inhibition. J. Neurosci. *31*, 6891–6899.

Jha, A., Nachev, P., Barnes, G., Husain, M., Brown, P., and Litvak, V. (2015). The frontal control of stopping. Cereb. Cortex *25*, 4392–4406.

Leblois, A., Boraud, T., Meissner, W., Bergman, H., and Hansel, D. (2006). Competition between feedback loops underlies normal and pathological dynamics in the basal ganglia. J. Neurosci. *26*, 3567–3583.

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

Lo, C.C., Boucher, L., Paré, M., Schall, J.D., and Wang, X.J. (2009). Proactive inhibitory control and attractor dynamics in countermanding action: a spiking neural circuit model. J. Neurosci. *29*, 9059–9071.

Logan, G.D., and Cowan, W.B. (1984). On the ability to inhibit thought and action: a theory of an act of control. Psychol. Rev. *91*, 295–327.

Logan, G.D., Van Zandt, T., Verbruggen, F., and Wagenmakers, E.J. (2014). On the ability to inhibit thought and action: general and special theories of an act of control. Psychol. Rev. *121*, 66–95.

Majid, D.S., Cai, W., Corey-Bloom, J., and Aron, A.R. (2013). Proactive selective response suppression is implemented via the basal ganglia. J. Neurosci. *33*, 13259–13269.

Mallet, N., Micklem, B.R., Henny, P., Brown, M.T., Williams, C., Bolam, J.P., Nakamura, K.C., and Magill, P.J. (2012). Dichotomous organization of the external globus pallidus. Neuron *74*, 1075–1086.

Mallet, N., Schmidt, R., Leventhal, D., Chen, F., Amer, N., Boraud, T., and Berke, J.D. (2016). Arkypallidal cells send a stop signal to striatum. Neuron *89*, 308–316.

Mayse, J.D., Nelson, G.M., Park, P., Gallagher, M., and Lin, S.C. (2014). Proactive and reactive inhibitory control in rats. Front. Neurosci. *8*, 104.

Mayse, J.D., Nelson, G.M., Avila, I., Gallagher, M., and Lin, S.C. (2015). Basal forebrain neuronal inhibition enables rapid behavioral stopping. Nat. Neurosci. *18*, 1501–1508.

McAlonan, G.M., Cheung, V., Chua, S.E., Oosterlaan, J., Hung, S.F., Tang, C.P., Lee, C.C., Kwong, S.L., Ho, T.P., Cheung, C., et al. (2009). Age-related grey matter volume correlates of response inhibition and shifting in attention-deficit hyperactivity disorder. Br. J. Psychiatry *194*, 123–129.

Middlebrooks, P.G., and Schall, J.D. (2014). Response inhibition during perceptual decision making in humans and macaques. Atten. Percept. Psychophys. *76*, 353–366.

Mills, K.L., Bathula, D., Dias, T.G., Iyer, S.P., Fenesy, M.C., Musser, E.D., Stevens, C.A., Thurlow, B.L., Carpenter, S.D., Nagel, B.J., et al. (2012). Altered cortico-striatal-thalamic connectivity in relation to spatial working memory capacity in children with ADHD. Front. Psychiatry *3*, 2.

Mirabella, G., Iaconelli, S., Romanelli, P., Modugno, N., Lena, F., Manfredi, M., and Cantore, G. (2012). Deep brain stimulation of subthalamic nuclei affects arm response inhibition in Parkinson's patients. Cereb. Cortex *22*, 1124–1132.

Newsome, W.T., Britten, K.H., and Movshon, J.A. (1989). Neuronal correlates of a perceptual decision. Nature *341*, 52–54.

O'Hare, J.K., Ade, K.K., Sukharnikova, T., Van Hooser, S.D., Palmeri, M.L., Yin, H.H., and Calakos, N. (2016). Pathway-specific striatal substrates for habitual behavior. Neuron *89*, 472–479.

Paré, M., and Hanes, D.P. (2003). Controlled movement processing: superior colliculus activity associated with countermanded saccades. J. Neurosci. *23*, 6480–6489.

Parent, A., and Hazrati, L.N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Res. Brain Res. Rev. *20*, 91–127.

Parnaudeau, S., O'Neill, P.K., Bolkan, S.S., Ward, R.D., Abbas, A.I., Roth, B.L., Balsam, P.D., Gordon, J.A., and Kellendonk, C. (2013). Inhibition of mediodorsal thalamus disrupts thalamofrontal connectivity and cognition. Neuron *77*, 1151–1162.

Paz, J.T., Davidson, T.J., Frechette, E.S., Delord, B., Parada, I., Peng, K., Deisseroth, K., and Huguenard, J.R. (2013). Closed-loop optogenetic control of thalamus as a tool for interrupting seizures after cortical injury. Nat. Neurosci. *16*, 64–70.

Ridderinkhof, K.R., Ullsperger, M., Crone, E.A., and Nieuwenhuis, S. (2004). The role of the medial frontal cortex in cognitive control. Science *306*, 443–447.

Roitman, J.D., and Shadlen, M.N. (2002). Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. J. Neurosci. *22*, 9475–9489.

Scangos, K.W., and Stuphorn, V. (2010). Medial frontal cortex motivates but does not control movement initiation in the countermanding task. J. Neurosci. *30*, 1968–1982.

Schall, J.D., and Godlove, D.C. (2012). Current advances and pressing problems in studies of stopping. Curr. Opin. Neurobiol. *22*, 1012–1021.

Schmidt, R., Leventhal, D.K., Mallet, N., Chen, F., and Berke, J.D. (2013). Canceling actions involves a race between basal ganglia pathways. Nat. Neurosci. *16*, 1118–1124.

Schroll, H., Vitay, J., and Hamker, F.H. (2012). Working memory and response selection: a computational account of interactions among cortico-basalganglio-thalamic loops. Neural Netw. *26*, 59–74.

Sharp, D.J., Bonnelle, V., De Boissezon, X., Beckmann, C.F., James, S.G., Patel, M.C., and Mehta, M.A. (2010). Distinct frontal systems for response inhibition, attentional capture, and error processing. Proc. Natl. Acad. Sci. USA *107*, 6106–6111.

Stinear, C.M., Coxon, J.P., and Byblow, W.D. (2009). Primary motor cortex and movement prevention: where Stop meets Go. Neurosci. Biobehav. Rev. *33*, 662–673.

Stuphorn, V. (2015). Neural mechanisms of response inhibition. Curr. Opin. Behav. Sci. *1*, 64–71.

Swann, N., Tandon, N., Canolty, R., Ellmore, T.M., McEvoy, L.K., Dreyer, S., DiSano, M., and Aron, A.R. (2009). Intracranial EEG reveals a time- and frequency-specific role for the right inferior frontal gyrus and primary motor cortex in stopping initiated responses. J. Neurosci. *29*, 12675–12685.

Swann, N.C., Cai, W., Conner, C.R., Pieters, T.A., Claffey, M.P., George, J.S., Aron, A.R., and 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.

Tanaka, M., and Kunimatsu, J. (2011). Contribution of the central thalamus to the generation of volitional saccades. Eur. J. Neurosci. *33*, 2046–2057.

Thakkar, K.N., Schall, J.D., Boucher, L., Logan, G.D., and Park, S. (2011). Response inhibition and response monitoring in a saccadic countermanding task in schizophrenia. Biol. Psychiatry *69*, 55–62.

Thakkar, K.N., Schall, J.D., Logan, G.D., and Park, S. (2015). Cognitive control of gaze in bipolar disorder and schizophrenia. Psychiatry Res. *225*, 254–262.

Turner, R.S., and Desmurget, M. (2010). Basal ganglia contributions to motor control: a vigorous tutor. Curr. Opin. Neurobiol. *20*, 704–716.

Verbruggen, F., and Logan, G.D. (2008). Response inhibition in the stop-signal paradigm. Trends Cogn. Sci. *12*, 418–424.

Verbruggen, F., and Logan, G.D. (2009). Models of response inhibition in the stop-signal and stop-change paradigms. Neurosci. Biobehav. Rev. *33*, 647–661.

Vitay, J., and Hamker, F.H. (2010). A computational model of Basal Ganglia and its role in memory retrieval in rewarded visual memory tasks. Front. Comput. Neurosci. *4*, 13.

Wang, X.J. (2001). Synaptic reverberation underlying mnemonic persistent activity. Trends Neurosci. *24*, 455–463.

Wang, X.J. (2002). Probabilistic decision making by slow reverberation in cortical circuits. Neuron *36*, 955–968.

Watanabe, Y., and Funahashi, S. (2012). Thalamic mediodorsal nucleus and working memory. Neurosci. Biobehav. Rev. *36*, 134–142.

Wei, W., Rubin, J.E., and Wang, X.J. (2015). Role of the indirect pathway of the basal ganglia in perceptual decision making. J. Neurosci. *35*, 4052–4064.

Wiecki, T.V., and Frank, M.J. (2013). A computational model of inhibitory control in frontal cortex and basal ganglia. Psychol. Rev. *120*, 329–355.

Wong, K.F., and Wang, X.J. (2006). A recurrent network mechanism of time integration in perceptual decisions. J. Neurosci. *26*, 1314–1328.

Yoshida, A., and Tanaka, M. (2016). Two types of neurons in the primate globus pallidus external segment play distinct roles in antisaccade generation. Cereb. Cortex *26*, 1187–1199.

Zagha, E., Ge, X., and McCormick, D.A. (2015). Competing neural ensembles in motor cortex gate goal-directed motor output. Neuron *88*, 565–577.

Zandbelt, B.B., van Buuren, M., Kahn, R.S., and Vink, M. (2011). Reduced proactive inhibition in schizophrenia is related to corticostriatal dysfunction and poor working memory. Biol. Psychiatry *70*, 1151–1158.

Neuron, Volume *92*

Supplemental Information

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

Wei Wei and Xiao-Jing Wang

Supplemental Figures

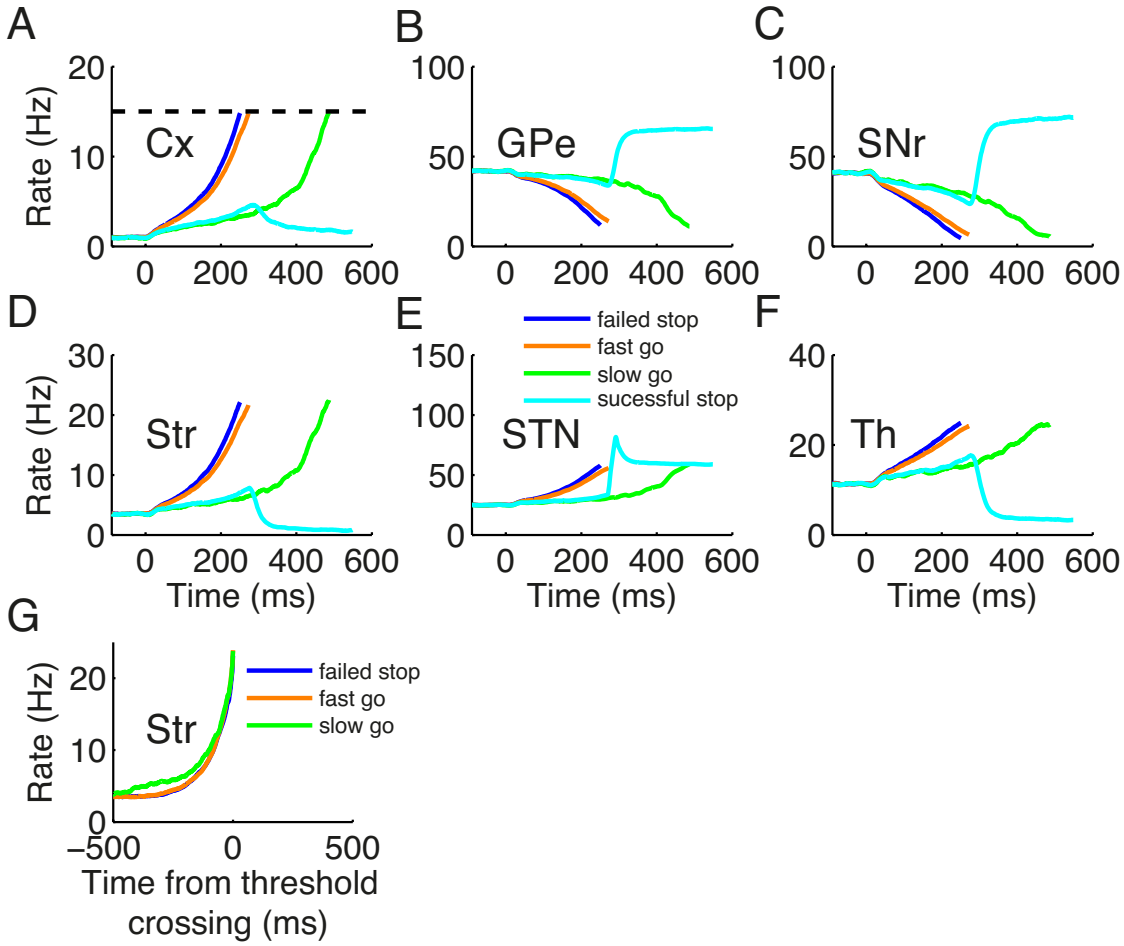

Figure S1 (related to Figure 3): Mean neuronal activities for different types of trials.
(A-F) Mean neuronal activities in each area within the circuit, aligned with the onset of go stimulus. (G) Mean neuronal activities in the Str, aligned with the time for threshold crossing in the Cx. The blue, orange, green and cyan curves represent failed stop signal trials, latency-matched fast go trials, latency-matched slow go trials, and successful stop signal trials. The latency-matched fast/slow go trials are go trials with RTs shorter/longer than SSD+SSRT, where SSD = 270 ms and SSRT = 155 ms, as estimated in Fig. 3.

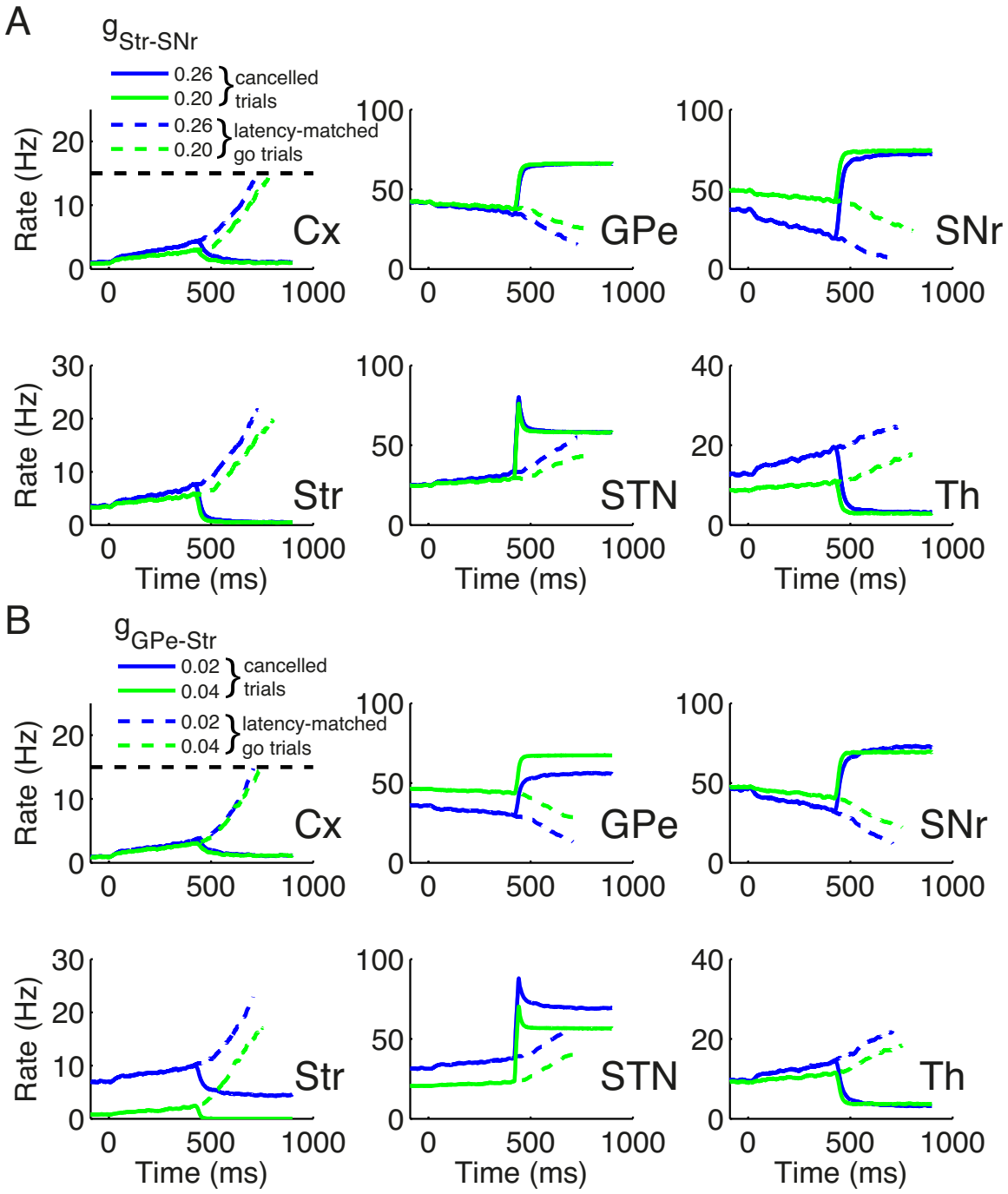

Figure S2 (related to Figure 4): **Dependence of mean neuronal activities for successful stop signal trials and latency-matched go trials on $g_{Str-SNr}$ and $g_{GPe-Str}$.** (A) Mean neuronal activities in each area within the circuit with $g_{Str-SNr}$ = 0.20 nS (green) and 0.26 nS (blue). (B) Mean neuronal activities in each area within the circuit with $g_{GPe-Str}$ = 0.02 nS (blue) and 0.04 nS (green). The solid and dashed curves represent successful stop signal trials and latency-matched go trials, respectively. The mean activities are obtained from averaging over 200 trials of the same types. Note that in (B), for fair comparison the spontaneous firing rate of the SNr is fixed by adjusting the external input rate υ_{ext} to it.

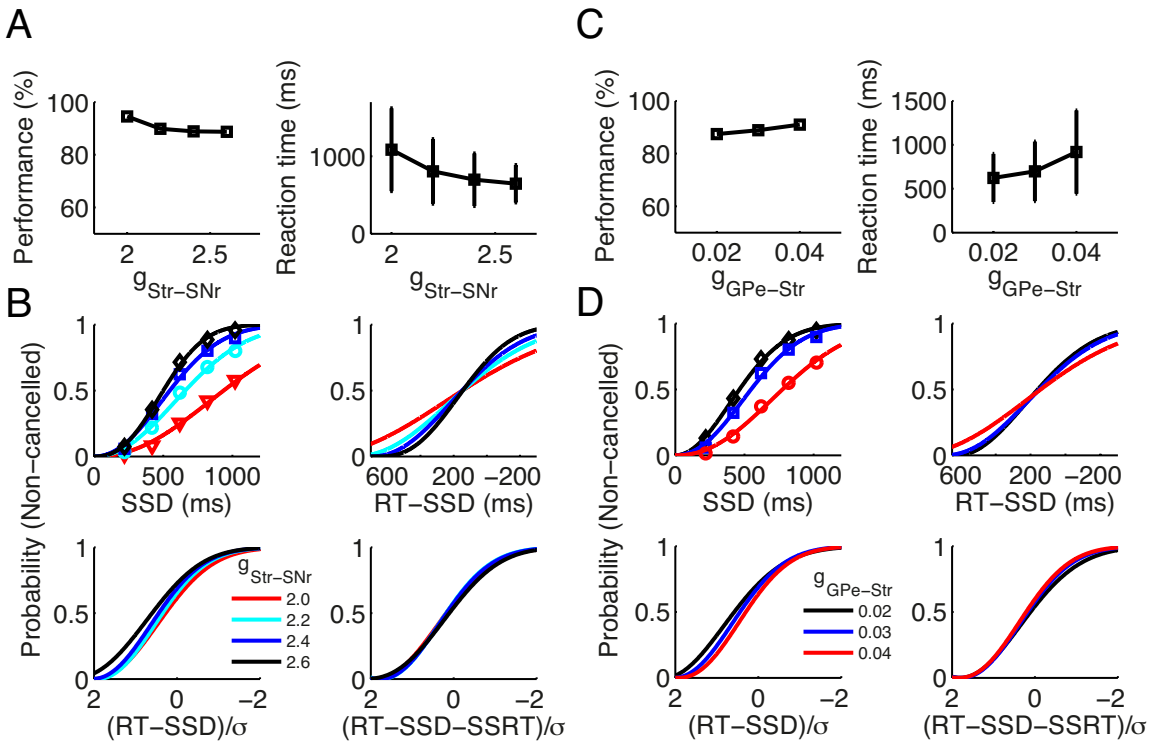

Figure S3 (related to Figure 4): Performances and RTs of the go trials and alignment of inhibition functions when $g_{Str-SNr}$ and $g_{GPe-Str}$ are modulated. (A) Performance (Left), and mean and SD of RTs of go trials for different $g_{Str-SNr}$. (B) Non-cancelled probabilities for different $g_{Str-SNr}$ as a function of SSD, RT-SSD, (RT-SSD)/σ, and (RT-SSD-SSRT)/σ, respectively. (C-D) Same as (A-B), but for the modulation of $g_{GPe-Str}$. Here RT and σ represent mean and SD of the RTs of go trials. The inhibition functions for different $g_{Str-SNr}$ ($g_{GPe-Str}$) are aligned when the dependence on SSD is transformed to the dependence on the z score of the relative finishing time (ZRFT), (RT-SSD-SSRT)/σ.

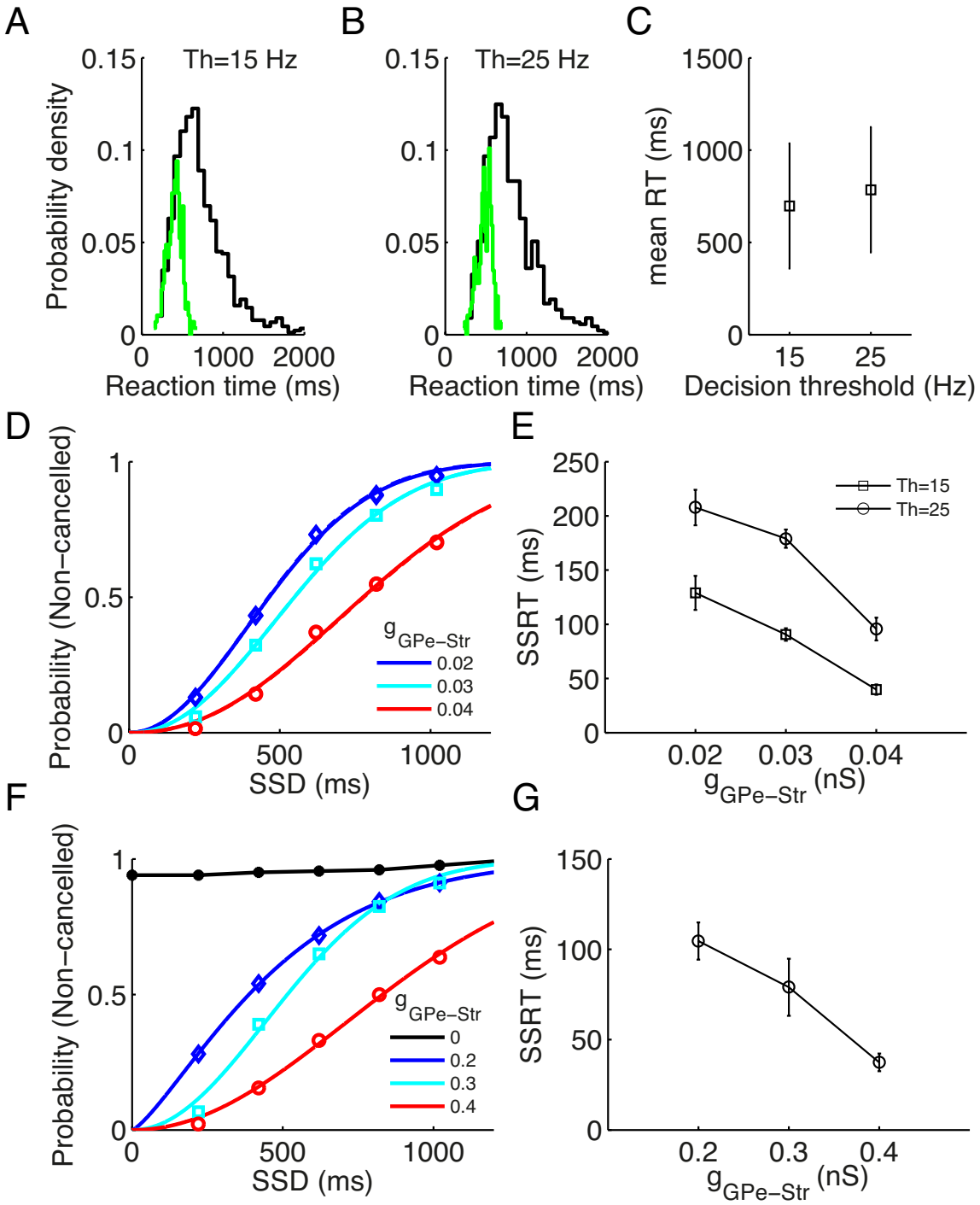

Figure S4 (related to Figure 4): Impact of back-projection from the GPe to the Str for different levels of decision threshold and in a scaled circuit with all-to-all connections in all the areas. (A-B) RTs for go trials and stop signal trials with SSD = 420 ms when the decision threshold (Th) in the Cx is setting at 15 Hz (A) and 25 Hz (B), respectively. (C) Mean and SD for the go trials RTs. (D) Inhibition functions for different GPe-to-Str projection strengths ($g_{GPe-Str}$) with Th = 15 Hz (solid lines) and Th = 25 Hz (dashed lines). Note that the solid and dashed lines are almost indistinguishable. (E) Modulation of the SSRT by $g_{GPe-Str}$ shows similar trends for different levels of decision threshold. (F-G) Dependence of inhibition functions and SSRT on $g_{GPe-Str}$ in a scaled circuit with all-to-all connections in all the areas. In (F-G), the

numbers of neurons in the GPe and STN are scaled down from 5000 in original model to 500. The connection probabilities within the STN-GPe sub-circuit are all scaled to 1 now, such that the connection strengths are scaled by a factor 10*p, where p is the connection probability for the corresponding synaptic connection in the original model. Error bars indicate SD of SSRT values estimated for each SSD. The inhibition functions in (D) and (F) are fitted by the Weibull function.

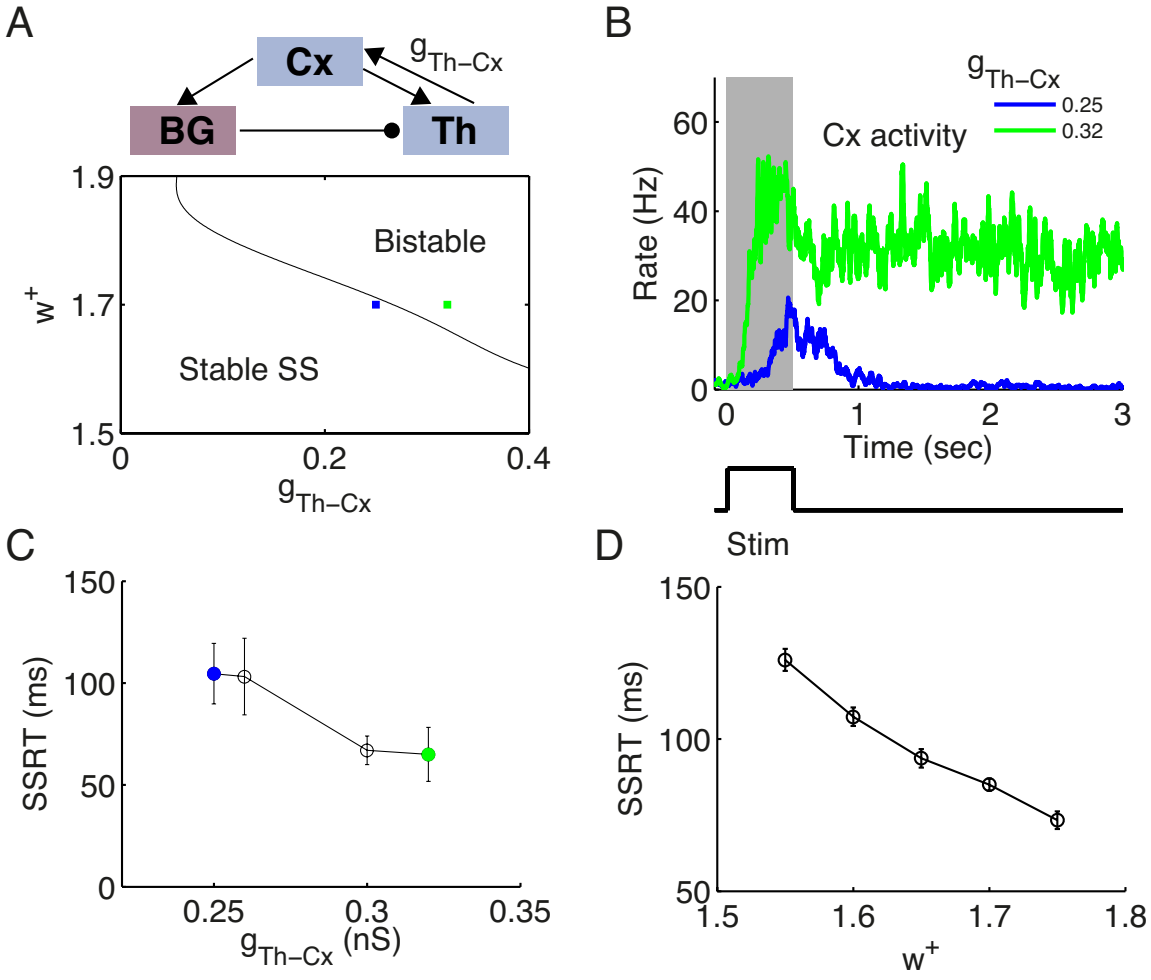

Figure S5 (related to Figure 5): Dependence of SSRT on thalamocortical connection strength when there is a corticothalamic sub-loop and on local cortical recurrent connection strength w^+. (A) Upper panel: schematic of the Cx-BG-Th loop with a Cx-Th sub-loop. Lower panel: the state diagram as a function of the local recurrent feedback strength w^+ in the Cx and the subcortico-cortical loop feedback strength g_{Th-Cx}. Stable SS: region where only the spontaneous state (SS) is stable. Bistable: region where the SS and target-selective elevated state are stable. Two sample networks are indicated in the parameter space at g_{Th-Cx} = 0.25 nS (blue) and 0.32 nS (green) with w^+ = 1.7. (B) Neuronal activities of the winner population in the Cx in response to a brief stimulus. The green and blue curves correspond to the green and blue squares in (A). The gray-shaded area shows the neuronal activities during stimulus onset, which lasts for 0.5 sec (lower panel). (C) SSRT shows a two-level modulation by g_{Th-Cx}, corresponding to whether there exists persistent activity or not. (D) SSRT decrease monotonically when w^+ is increased. In (A-C), w^+ = 1.7, g_{Cx-Th} = 0.2 nS. In (D), g_{Th-Cx} = 0.32 nS, g_{Cx-Th} = 0. Error bars indicate SD of SSRT values estimated for different SSDs.

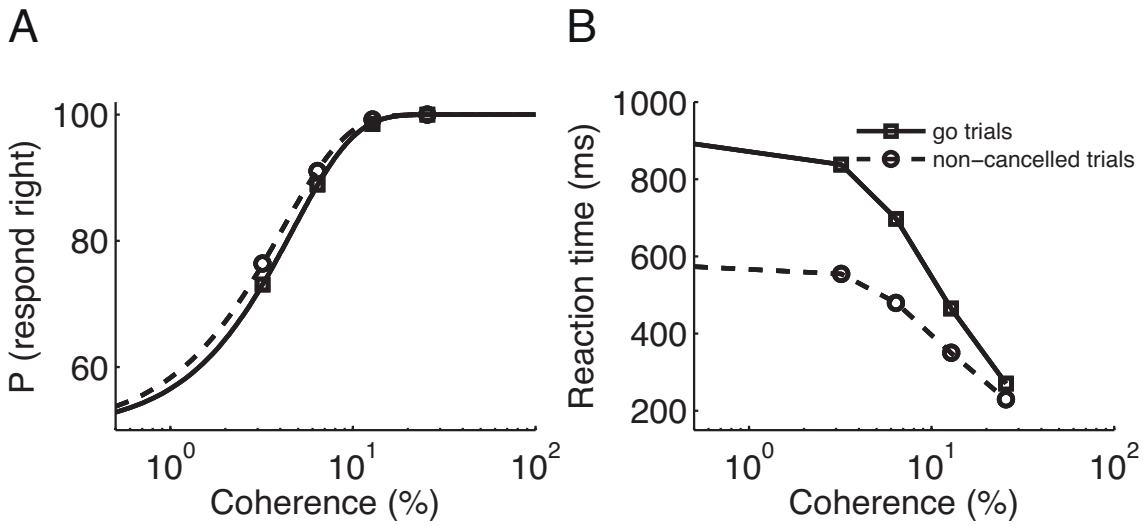

Figure S6 (related to Figure 6): Probabilities of choosing right (A) and mean RTs (B) in go trials and non-cancelled stop signal trials when the net direction of random-dot stimulus is right with different coherences. For non-cancelled stop signal trials, the probability of choosing right increases and mean reaction time decreases with the increase of motion coherence, which are similar to that for go trials. Note that in experiments the SSD is usually adjusted following a staircase procedure to ensure total non-cancelled probability around 0.5. In modeling study we performed simulations for each SSD, and here the performances and mean RTs for non-cancelled trials are average values over two SSDs that have non-cancelled probabilities close to 0.5 for each coherence level of the random-dot input.

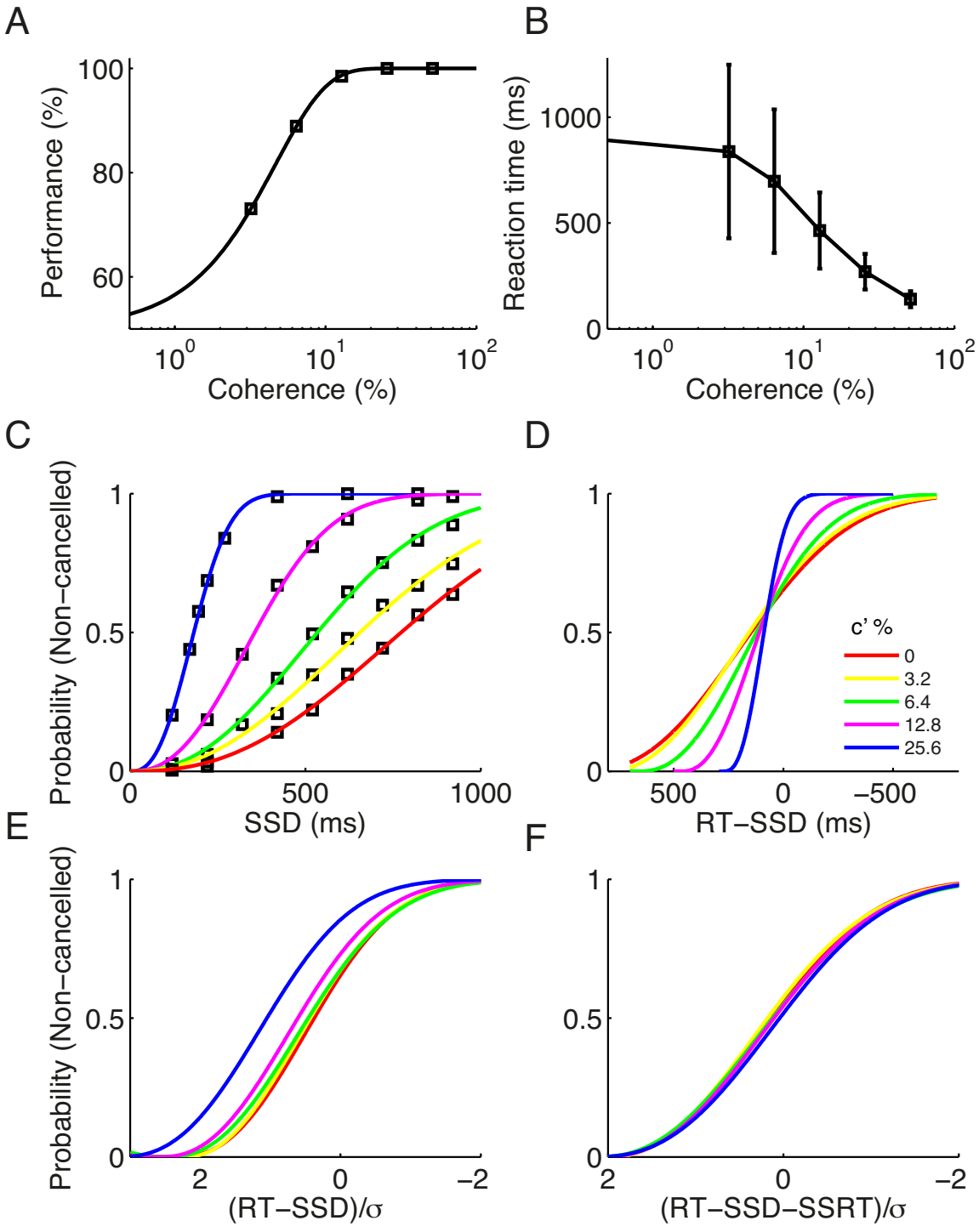

Figure S7 (related to Figure 6): Performances and RTs of the go trials and alignment of inhibition functions when the motion coherence c' is varied. Similar to Figure S3, but for the variation of c'. The inhibition functions for different c' are aligned when the dependence on SSD is transformed to the dependence on ZRFT, (RT-SSD-SSRT)/σ.

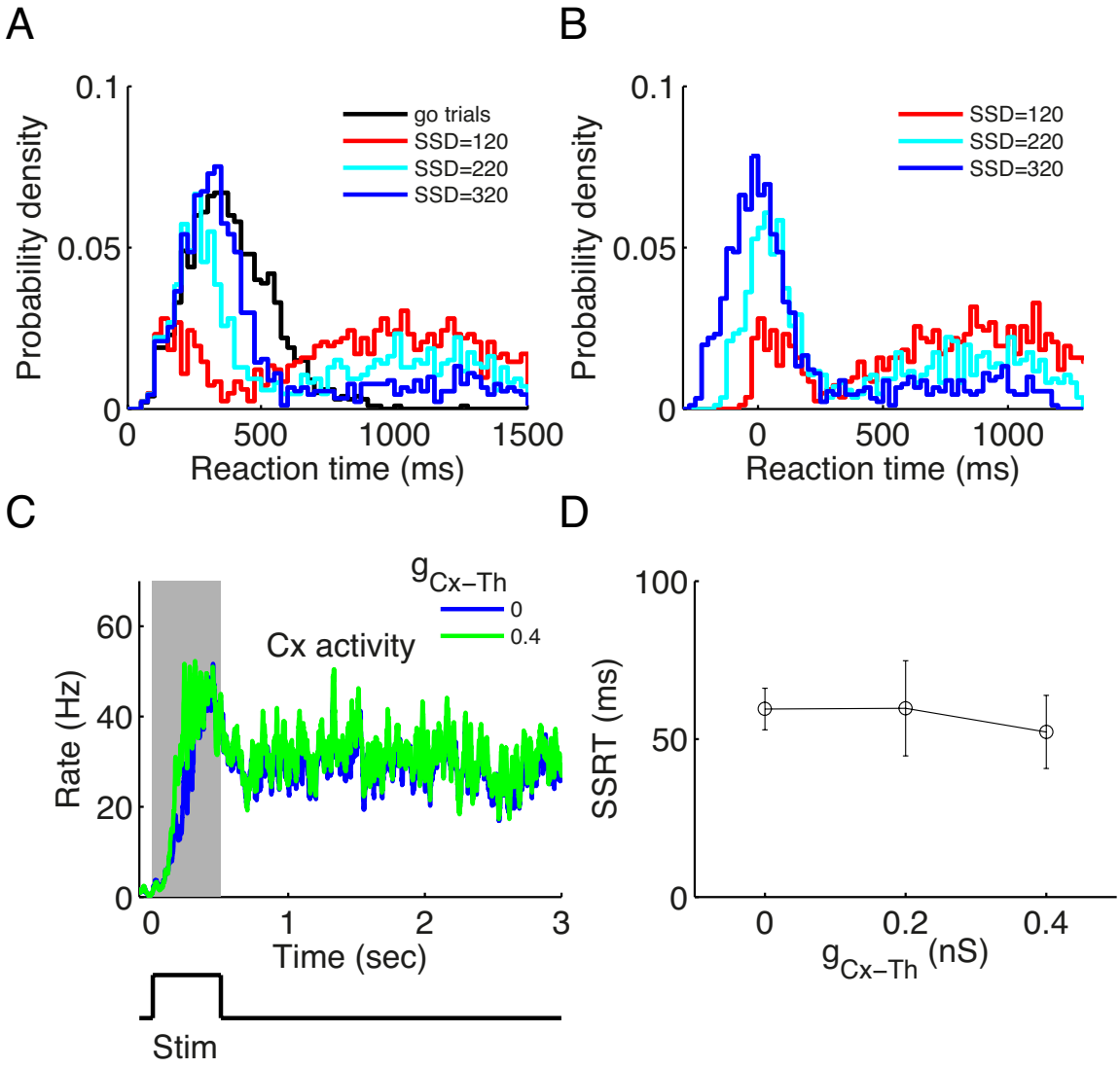

Figure S8 (related to Figure 7): Impact of corticothalamic sub-loop on RT distributions and SSRT. (A-B) RT distributions for the go trials and non-cancelled stop signal trials aligned with the go stimulus onset (A) and stop signal onset (B). (C) Neuronal activities of the winner population in the Cx for $g_{Cx-Th} = 0$ (blue) and $g_{Cx-Th} = 0.4$ nS (green), respectively. The Cx shows persistent activity in both cases. The gray-shaded area shows the neuronal activities during stimulus onset, which lasts for 0.5 sec (lower panel). The stimulus is a random-dot input with $c' = 25.6\%$. (D) SSRT is insensitive to g_{Cx-Th}. Error bars indicate SD of SSRT values estimated for different SSDs. In (A-B), $g_{Cx-Th} = 0.4$ nS, $g_{Th-Cx} = 0.26$ nS. In (C-D), $g_{Th-Cx} = 0.32$ nS.

Supplemental Experimental Procedures

Behavioral task simulation

We simulated a two-alternative choice task and considered two types of stimulus. The go stimulus is implemented by an increase of the rate of external background inputs, modeled as Poissonian spike trains, to the selective population (left or right) in the Cx from 2.4 kHz to 2.464 kHz. For the random-dot stimulus, the direction of coherent motion is assumed to be left- or rightward. Two signals, representing sensory inputs induced by the leftward and rightward moving dots, are fed into left- and right-preferring cortical modules Cx_L and Cx_R, respectively. The firing times of each input neuron are sampled from a Poisson process with a mean rate μ. The mean μ depends linearly on the dot coherence level and is given by the equations $\mu_0 + \mu_A \times c'$ for the preferred direction and $\mu_0 - \mu_B \times c'$ for the non-preferred direction, where μ_0 (20 Hz) is the baseline input, c' (between 0% and 100%) is the coherence level, and μ_A (60 Hz) and μ_B (20 Hz) are proportionality factors (Wang, 2002). The stop signal is simplified as a constant input to the STN, entering the BG via the hyperdirect pathway. The stop signal is implemented by an increase of the external Poisson input rate to all the STN neurons from 4.0 kHz to 5.2 kHz.

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. From Figure S4 (A-C), we saw that the mean RT distribution is shifted toward right for a higher threshold, but the standard deviation is not changed. The inhibition function is not influenced by the choice of the decision threshold (Figure S4D). This is expected, since whether a trial is cancelled or not by a stop signal is determined much earlier than the time for threshold crossing where the stop signal is unlikely to cancel a trial due to the attractor dynamics. We saw that the SSRT is increased for a higher threshold, since the mean RT for go trials is increased and the inhibition function is not changed (the SSRT is roughly given by the mean RT for go trials subtracted by the SSD corresponding to 50% non-cancelled probability). The trend of SSRT modulation by model parameters, illustrated here by the strength of GPe-to-Str back-projection, is conserved (Figure 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 to 1 for large SSD.

Integration method for estimating the SSRT

The integration method has been widely used in estimating the SSRT, which is illustrated in Figure 3A. When the cumulative density function (cdf) of RTs is plotted for go trials (black) and non-cancelled stop signal trials (green) in Figure 3A, the crossing point of the horizontal dashed line (representing the non-cancelled probability) with the black curve gives an estimate of the SSD plus SSRT (Figure 3A, upper panel). Below this time (red vertical dashed line), the probabilities for the RTs of go trials and non-cancelled trials roughly coincide with each other (Figure 3A, lower panel).

In Figure 7A, a modified integration method as introduced in Mayse et al. (2014) was illustrated. The intersection of the cdf for the RTs of the non-cancelled stop signal trials (scaled with the non-cancelled probability; green solid curve) with the 99.9% confidence interval of the cdf for the RTs of go trials (black dashed curve) gives an estimate of the SSD plus the SSRT.

Neuron model

We used the leaky integrate-and-fire model for all the neurons. The membrane dynamics is given by

$$C\frac{dV}{dt} = -g_L(V - V_L) - I_{syn}\,, \quad (1)$$

where C = 0.5 nF, g_L= 25 nS, V_L = -70 mV. When the membrane potential reaches a boundary V_b, it is reset to V_r, where V_b = -50 mV and V_r = -55 mV. The synaptic current I_{syn} is given by

$$I_{syn} = g_1 s(V - V_E) + \frac{g_2 s(V - V_E)}{1 + e^{-0.062V/3.57}} + g_3 s(V - V_I)\,, \quad (2)$$

where the reversal potentials V_E = 0 and V_I = -70 mV. g_k is the synaptic efficacy, where the indices k = 1, 2, 3 indicate AMPA, NMDA and $GABA_A$ synapses, respectively. The gating variable s satisfies

$$\frac{ds}{dt} = \sum_j \delta(t - t^j) - \frac{s}{\tau} \quad (3)$$

for AMPA and $GABA_A$ receptor mediated currents and

$$\frac{ds}{dt} = \alpha(1 - s)\sum_j \delta(t - t^j) - \frac{s}{\tau} \quad (4)$$

for NMDA receptor mediated current, where t^j is the time for the jth spike and $\alpha = 0.63$. For the decay time constant τ, we use 2 ms for AMPA, 5 ms for $GABA_A$, and 100 ms for NMDA mediated currents. I_{syn} describes both the synaptic inputs from other neurons within the circuit and background inputs modeled as Poissonian spike trains representing sources beyond the circuit. We use a short time delay, 0.2 ms, for synaptic transmission.

To soften the positive feedback of the cortico-subcortico-cortical loop, we introduced short-term depression (STD) at thalamocortical synapses, but the results do not depend on this assumption. The STD is implemented by including a factor D that multiplies all terms on the right side of Eq. (2) (Hempel et al., 2000), which satisfies

$$\frac{dD}{dt} = -pD\sum_j \delta(t - t^j) + \frac{1-D}{\tau_D}, \quad (5)$$

where p = 0.45 and τ_D = 600 ms.

Network structure

Each BG nucleus includes two populations of neurons selective to left and right targets or left- and rightward moving dots, respectively. Following Wei et al. (2015), each population in the Str and SNr includes 250 neurons with all-to-all connections, while each population in the STN and GPe includes 2500 neurons with sparse connection probabilities: 0.05 for STN→GPe, 0.05 for GPe→GPe, and 0.02 for GPe→STN. There are no known recurrent connections within the STN (Sato et al., 2000), hence such connections are not included. All other projections are all-to-all for the left and right selective populations, respectively. The sparse connections in the STN-GPe sub-circuit are inherited from our previous model (Wei et al., 2015), where we proposed a mechanism for excessive beta oscillation in Parkinson's disease that depends on the sparse connections of the STN-GPe circuit. Here we adapted this structure with the hope to provide a framework allowing for investigating such a parkinsonian state in the future. To confirm that the sparse connectivity in the STN-GPe circuit is not essential, we performed additional simulations (Figure S4F-S4G), where we scaled the STN and GPe circuits to have all-to-all connections and 500 neurons in both areas. We observed qualitatively the same results as in the original model, illustrated by the impact of GPe-to-Str back-projection strength. Similarly, the Th includes also two populations due to the topographic connections from the SNr to higher order thalamus (Gulcebi et al., 2012), with each population including 250 neurons with all-to-all connections.

The connection efficacies used in the model are the following (in unit of nS): g_{Cx-Str} = 1.0, $g_{Str-Str}$ = 1.0, $g_{Str-SNr}$ = 2.4, $g_{Str-GPe}$ = 3.0, $g_{GPe-GPe}$ = 1.5, $g_{GPe-Str}$ = 0.03, $g_{GPe-STN}$ = 0.8, $g_{GPe-SNr}$ = 0.04, $g^{NMDA}_{STN-GPe}$ = 2.0, $g_{STN-SNr}$ = 0.06, g_{SNr-Th} = 0.09. We will use g_{Th-Cx} = 0.32 nS and $g_{Cx-Th} = 0$, unless otherwise indicated. For the Cx, we used the same synaptic connection properties as in past studies (Lo and Wang, 2006; Wang, 2002), except that the efficacy of background inputs to the two selective populations are reduced from 2.1 nS to 2.0 nS. This modification weakens the cortical background input to support a distributed attractor state.

Background inputs, which are incorporated into equation (4), are described as Poissonian spike trains. All of the BG nuclei in the model receive AMPAergic background inputs, with possible cortical and thalamic origins. GPe neurons also receive GABAergic background inputs representing additional inputs from the striatum. The synaptic efficacies and Poisson rates used for the background inputs in our model are (in unit of nS and kHz, respectively): 4.0 and 1.6 for the Str; 2.0 and 4.0 for GPe AMPA receptors; 2.0 and 2.0 for GPe $GABA_A$ receptors; 1.6 and 4.0 for the STN; 6.0 and 0.8 for the SNr, 2.0 and 3.2 for the Th. Note that in Figure 4 (B, D), for fair comparison when $g_{GPe-Str}$ or $g_{Ark-Str}$ is varied, the spontaneous firing rate of the SNr is fixed by adjusting the external Poisson input rate υ_{ext} to it. For $g_{GPe-Str}$ ($g_{Ark-Str}$) taking the values 0, 0.02, 0.03 and 0.04 nS, the values of υ_{ext} used are 1.1, 0.94, 0.8 and 0.76 kHz, respectively.