J Neurophysiol 114: 650–661, 2015.
First published May 20, 2015; doi:10.1152/jn.00845.2013.

Speed-accuracy tradeoff by a control signal with balanced excitation and inhibition

Chung-Chuan Lo,[1,2]* Cheng-Te Wang,[1] and Xiao-Jing Wang[3]*

[1]*Institute of Systems Neuroscience, National Tsing Hua University, Hsinchu, Taiwan;* [2]*Brain Research Center, National Tsing Hua University, Hsinchu, Taiwan; and* [3]*Center for Neural Science, New York University, New York, New York*

Submitted 28 November 2013; accepted in final form 14 May 2015

Lo CC, Wang CT, Wang XJ. Speed-accuracy tradeoff by a control signal with balanced excitation and inhibition. *J Neurophysiol* 114: 650–661, 2015. First published May 20, 2015; doi:10.1152/jn.00845.2013.—A hallmark of flexible behavior is the brain's ability to dynamically adjust speed and accuracy in decision-making. Recent studies suggested that such adjustments modulate not only the decision threshold, but also the rate of evidence accumulation. However, the underlying neuronal-level mechanism of the rate change remains unclear. In this work, using a spiking neural network model of perceptual decision, we demonstrate that speed and accuracy of a decision process can be effectively adjusted by manipulating a top-down control signal with balanced excitation and inhibition [balanced synaptic input (BSI)]. Our model predicts that emphasizing accuracy over speed leads to reduced rate of ramping activity and reduced baseline activity of decision neurons, which have been observed recently at the level of single neurons recorded from behaving monkeys in speed-accuracy tradeoff tasks. Moreover, we found that an increased inhibitory component of BSI skews the decision time distribution and produces a pronounced exponential tail, which is commonly observed in human studies. Our findings suggest that BSI can serve as a top-down control mechanism to rapidly and parametrically trade between speed and accuracy, and such a cognitive control signal presents both when the subjects emphasize accuracy or speed in perceptual decisions.

decision making; speed-accuracy tradeoff; top-down control; balanced input

THE ABILITY TO DYNAMICALLY adjust speed vs. accuracy is a salient feature of decision-making (Bogacz et al. 2010; Gold and Shadlen 2002; Heitz 2014; Wang 2008; Wickelgren 1977): if "to get it right" is the priority, we slow down to gather more information and gain a better performance. On the other hand, if time is at a premium (e.g., when detecting a predator), we make a quick judgment at the potential cost of accuracy. Speed-accuracy tradeoff (SAT) can result from a trial-by-trial learning process, which is believed to depend on synaptic plasticity and reward information in the cortex and basal ganglia (Balci et al. 2010; Gold and Shadlen 2002; Lo and Wang 2006; Wang et al. 2013). However, adjustment can often be made quickly, "on the fly," based on either task instruction, a changing environment, or a subject's own will (Edwards 1965; Forstmann et al. 2008; Luce 1986; Palmer et al. 2005; Wickelgren 1977).

SAT is commonly considered in terms of adjusting a decision threshold of an integrator (Bogacz et al. 2006; Edwards 1965), such as that described by the drift diffusion model (DDM) (Luce 1986; Ratcliff 1978). In this framework, under speed emphasis the decision threshold is decreased, so it can be reached faster to trigger a response, whereas accuracy emphasis is instantiated by an increase of the decision threshold to integrate more information before a decision is made. The idea of changing decision threshold is intuitively appealing and consistent with fitting behavioral data with DDM in a number of experiments (reviewed in Bogacz et al. 2010). However, recent studies of SAT (Hanks et al. 2014; Heitz and Schall 2012) have revealed neuronal responses that are more complex than predicted by the threshold-tuning hypothesis. In particular, the studies showed that the ramping rates of the neuronal activity were modulated by the task conditions that emphasize speed or accuracy. Latest analysis also suggested that the behavioral and neuronal data in SAT experiments can be explained by changing the threshold and the rate of evidence accumulation in the system-level models, such as the drift diffusion and the linear ballistic accumulator models (Cassey et al. 2014; Rae et al. 2014).

Given the emerging evidence on the modulation of ramping rates by SAT, we ask how can a spiking neural network model be modified to accommodate such evidence with neuronal level mechanisms? In the present study, we used a leading attractor-based neural network model (Wang 2002), which captures the nature of nonlinearity of neural dynamics and has been shown to reproduce behavioral performance as well as neuronal responses in decision-making tasks (Deco and Rolls 2006; Furman and Wang 2008; Liu and Wang 2008; Soltani and Wang 2006; Wang 2002; Wong and Wang 2006). Recent studies suggested that, when considering the biological constraints, the attractor-based circuit models out-perform an integrator when accuracy is more important than speed (Miller and Katz 2013), and that robust and optimal decision-making can be realized by neuronal gain modulation in the attractor network model (Niyogi and Wong-Lin 2013). Furthermore, we have shown that the attractor dynamics and the decision behavior of the spiking neural network model can be rapidly changed by a top-down signal with balanced excitation and inhibition (Lo and Wang 2009; Wang et al. 2013). This balanced synaptic input (BSI) has been observed in vivo in various nervous systems, including the frontal cortex (Haider et al. 2006; Shu et al. 2003), primary visual cortex (Mariño et al. 2005) and spinal cord (Berg et al. 2007). Computational and experimental studies have shown that BSI can affect the response property of neurons (Anderson 2000; Burkitt et al. 2003; Hô and Destexhe 2000), and a key insight is that BSI provides a plausible mechanism for gain modulation (Abbott and Chance 2005;

* C.-C. Lo and X.-J. Wang contributed equally to this study.
Address for reprint requests and other correspondence: X.-J. Wang, Center for Neural Science, New York Univ., 4 Washington Pl., New York, NY 10003 (e-mail: xjwang@nyu.edu).

 0022-3077/15 Copyright © 2015 the American Physiological Society

Ayaz and Chance 2008; Brozovic et al. 2008; Chance et al. 2002; Salinas 2000; Vogels and Abbott 2009).

In this paper, we propose and show that emphasis on speed or accuracy can be rapidly achieved by BSI as a top-down control signal in the circuit model of decision making (Fig. 1). In the model, emphasis on accuracy requires a top-down signaling which, interestingly, reduces the gain of an integrator. Conversely, emphasis on speed correlates with an increased gain. Interestingly, this gain modulation can be realized in the model by changing the condition (strength and/or ratio of excitation and inhibition) of BSI. Moreover, our model predicts that the ramping rate of neural integrators for information accumulation and the baseline neural activity are lower with accuracy emphasis, which is consistent with recent studies in which single neurons in the lateral intraparietal area and the frontal eye field (FEF) of behaving monkeys displayed a reduction both in the ramping rate and the baseline activity under accuracy compared with speed emphasis (Hanks et al. 2014; Heitz and Schall 2012).

Therefore, the neural circuit model has identified a specific neural circuit mechanism for SAT. Our results suggest that BSI can serve as a top-down control mechanism to rapidly and parametrically modulate a perceptual decision. This finding suggests a specific functional role of long-distance top-down projection onto GABAergic inhibitory neurons, which has recently been documented for pathways from the prefrontal cortex to other parts of the brain (Barbas et al. 2005; Bunce and Barbas 2011; Medalla et al. 2007).

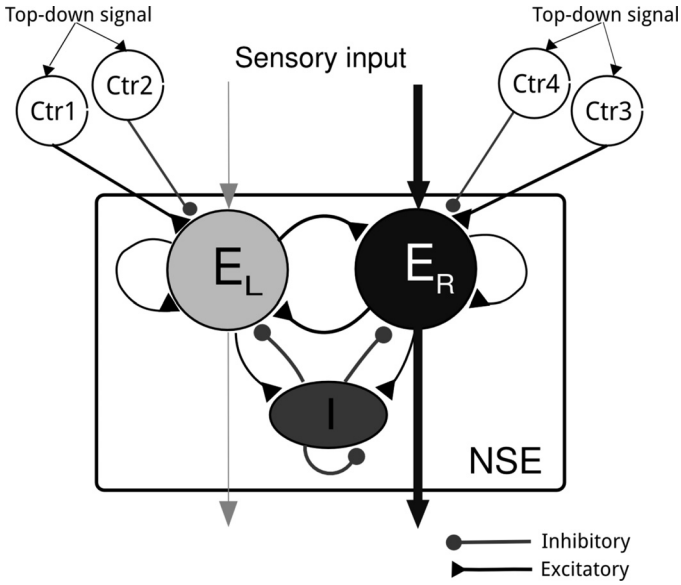

Fig. 1. A cortical circuit model of perceptual decision with the balanced synaptic input (BSI) as a top-down control. The decision circuit consists of two populations (E_L and E_R) of excitatory (E) neurons selective for choice alternatives, a population of inhibitory (I) interneurons and a population of nonselective excitatory neurons (NSE). The neural populations E_L and E_R receive sensory inputs and BSI and compete against each other; which of the two reaches a threshold first determines the network's choice in a simulated random-dot direction discrimination task (see MATERIALS AND METHODS). The top-down circuit consists of two pairs of E and I neural populations (Ctr1 and Ctr2 for E_L; Ctr3 and Ctr4 for E_R). Each pair provides BSI to one of the decision populations (E_L or E_R).

MATERIALS AND METHODS

Perceptual decision-making task. Our decision-making model of spiking neurons was described previously (Hsiao and Lo 2013; Lo and Wang 2006; Wang 2002; Wang et al. 2013; Wong et al. 2007; Wong and Wang 2006) and has been applied to different types of decision processes (Deco et al. 2009; Wang 2008). For the sake of concreteness, here we will focus on model simulations of a visual direction-discrimination task, the random-dot motion task (Roitman and Shadlen 2002). In the task, a subject is shown a display of randomly moving dots. A small portion (called coherence level or motion strength, c') of the dots move coherently toward one of the two possible directions, e.g., right or left. The subject is required to determine the direction of coherent motion. The subject has to indicate the direction by a saccadic eye movement as soon as a decision is reached. In our model, two inputs representing the amount of rightward and leftward random-dot motion directions [presumably from middle temporal area (MT) as previously reported (Britten et al. 1993)] are fed into two competing neural populations (E_R and E_L) in the decision circuit model, respectively (Fig. 1). The mean spike rate μ of each input depends on the motion strength of the stimulus linearly and follows the equations:

$$\mu = \mu_0 + \mu_A c' \quad (1)$$

for the direction of the coherent motion and

$$\mu = \mu_0 + \mu_B c' \quad (2)$$

for the opposite direction. μ_0 (=40 Hz) is the baseline input for the purely random motion, c' is the coherent motion strength with a value between 0 and 1, and μ_A (=120 Hz, unless otherwise stated) and μ_B (=40 Hz, unless otherwise stated) are factors of proportionality. It was reported that the population average of the slope of MT neuron response function is 3.5 times higher in the preferred direction than in the nonpreferred direction (or $\mu_A \sim 3.5\ \mu_B$) (Britten et al. 1993). Given the fact that μ_A and μ_B for individual MT neurons follow broad distributions (Britten et al. 1993), our assumption of $\mu_A = 3\ \mu_B$ is not substantially different from the observation. In the present study, we used six levels of motion strength: 0%, 3.2%, 6.4%, 12.8%, 25.6% and 51.2%. The decision time is defined as the time interval between the start of the sensory input and the time when the population firing rate of either of the populations E_L or E_R reaches a preset threshold (30 Hz). This preset threshold remained unchanged throughout the study. When comparing the model with behavioral data, we also calculated the response time which is the summation of the decision time and a nondecision time (t_{nd}), which represents the neural latency contributed by sensory and motor processes. A correct trial is defined as the trial in which the excitatory population (E_L or E_R) which receives the stronger input ($\mu_0 + \mu_A \times c'$) reaches the decision threshold first.

Neural circuit model. The computational model was described previously (Brunel and Wang 2001; Wang 2002; Wang et al. 2013). Briefly, the network model consists of four interconnected neural populations E_R, E_L, I and NSE (Fig. 1). Each of E_R and E_L contains 240 excitatory neurons and receives inputs that represent the random-dot moving toward right and left, respectively. The populations compete against each other through the inhibitory population I, which contains 400 inhibitory neurons. The nonselective population NSE contains 1,100 excitatory neurons, mimicking neurons that are selective for directions other than the two choice alternatives or to other stimuli that are irrelevant to the present study. In the model, NSE neurons do not receive stimulus input and maintain a baseline activity (several Hertz). The circuit model exhibits winner-take-all competition: only one of the excitatory populations (E_R or E_L) can win the competition by ramping up its activity to cross the decision threshold, whereas the other population is eventually suppressed. This behavior resembles neuronal activity observed in the lateral intraparietal area in monkeys when performing the random-dot task (Roitman and Shadlen

2002; Shadlen and Newsome 2001). In the present study, the decision populations E_R and E_L received additional input with balanced excitation and inhibition as top-down signals from populations Ctr_1 to Ctr_4 (Fig. 1) (Wang et al. 2013).

Synaptic strength and background noise. All synaptic connections between neural populations and within a neural population (recurrent connections) are all to all, i.e., every neuron in the source population makes synaptic connections to every neuron in the target population. The values of synaptic efficacy g (in nS) are as follows (for excitatory connections, values are given as g_{AMPA}/g_{NMDA} for currents mediated by AMPA and NMDA receptors): $g^{E_R-E_R} = g^{E_L-E_L} = 0.09/0.297$, $g^{E_R-I} = g^{E_L-I} = 0.04/0.13$, $g^{E_R-NSE} = g^{E_L-NSE} = 0.05/0.165$, $g^{E_R-E_L} = g^{E_L-E_R} = 0.04294/0.1417$, $g^{NSE-E_R} = g^{NSE-E_L} = 0.04294/0.1417$, $g^{NSE-I} = 0.04/0.13$, $g^{NSE-NSE} = 0.05/0.165$, $g^{I-I} = 1.075$ and $g^{I-E_R} = g^{I-E_L} = g^{I-NSE} = 1.3975$.

Each neuron in the network receives external inputs with effective Poisson spike rate of 2,400 Hz serving as the background noise. The background input is applied through AMPA receptor-mediated currents with a synaptic conductance of 2.1 nS for all excitatory neurons and 1.62 nS for all inhibitory neurons. The excitatory component of BSI is applied by the populations Ctr_1 and Ctr_3 to the decision populations E_L and E_R, respectively, through AMPA receptors. The inhibitory component of BSI is applied by the populations Ctr_2 and Ctr_4 to the decision populations E_L and E_R, respectively, through $GABA_A$ receptors.

Single-neuron and synapse models. Each neuron in the circuit model is simulated using the leaky integrate-and-fire model. The membrane potential $V(t)$ for each neuron obeys the following equation:

$$C_m \frac{dV(t)}{dt} = -g_L\left[V(t) - V_L\right] - I_{syn}(t) \quad (3)$$

where C_m is the membrane capacitance, g_L is the leak conductance, V_L is the resting potential, and I_{syn} is the total synaptic current. When the membrane potential $V(t)$ of each neuron reaches a threshold $V_{threshold} = -50$ mV, a spike is emitted, and $V(t)$ is set to the reset potential $V_{reset} = -55$ mV for a refractory period $T_r = 2$ ms. For inhibitory neurons, we used the following parameters: $C_m = 0.2$ nF, $g_L = 20$ nS and $V_L = -70$ mV. For excitatory neurons, we used $C_m = 0.5$ nF, $g_L = 25$ nS and $V_L = -70$ mV.

The synaptic current $I_{syn(t)}$ includes inputs from visual stimulus ($I_{stimulus}$), other neurons in the circuit [recurrent ($I_{recurrent}$)], background noise (I_{noise}) and BSI (I_{BSI}):

$$I_{syn}(t) = I_{stimulus}(t) + I_{recurrent}(t) + I_{noise}(t) + I_{BSI}(t) \quad (4)$$

where the background noise is applied to all neuronal populations, and visual stimulus and BSI are only applied to the populations E_R and E_L. We modeled three types of receptors for synapses: AMPA, NMDA and $GABA_A$. They are described by:

$$\text{Synaptic current} = g_{AMPA}s_{AMPA}(t)\left[V(t) - V_E\right] + \frac{g_{NMDA}s_{NMDA}(t)\left[V(t) - V_E\right]}{1 + \left[Mg^{2+}\right]e^{-0.062V(t)/3.57}} + g_{GABA}s_{GABA}(t)\left[V(t) - V_I\right] \quad (5)$$

where V_E (=0) and V_I (= −70 mV) are the reversal potentials, $[Mg^{2+}]$ (=1.0 mM) is the extracellular magnesium concentration, g is the synaptic efficacy, and s is the gating variable. Subscripts in g and s denote the receptor type. The gating variables of the three receptors obey

$$\frac{ds_{AMPA}(t)}{dt} = \sum_k \delta(t - t^k) - \frac{s_{AMPA}}{\tau_{AMPA}}$$

$$\frac{ds_{NMDA}(t)}{dt} = \alpha\left[1 - s_{NMDA}(t)\right]\sum_k \delta(t - t^k) - \frac{s_{NMDA}}{\tau_{NMDA}} \quad (6)$$

$$\frac{ds_{GABA}(t)}{dt} = \sum_k \delta(t - t^k) - \frac{s_{GABA}}{\tau_{GABA}}$$

where the decay constants $\tau_{AMPA} = 2$ ms, $\tau_{NMDA} = 100$ ms and $\tau_{GABA} = 5$ ms; $\alpha = 0.63$; $\delta(t - t^k)$ is the Dirac delta function; and t^k is the time of the kth presynaptic spike. We note that we did not model the rise-time dynamics in NMDA-mediated synapses because the rise time (~1–2 ms) is much smaller than the decay time (=100 ms), which dominates the dynamics of the network (Wong and Wang 2006). Therefore, neglecting the rise time of NMDA receptors did not significantly alter the dynamics of the system.

BSI. The setting of BSI generally follows the feedforward BSI mechanism described previously (Wang et al. 2013). We emphasize that the "balance" in BSI is not defined by any single ratio of excitatory and inhibitory input, but depends on the membrane potential of the recipient neurons. Briefly speaking, given a ratio between the strengths of the excitatory and inhibitory components of BSI, the induced depolarizing and hyperpolarizing currents can cancel each other (balanced) at a certain level of the membrane potential, V_B, of the recipient neurons. Taking the decision population E_L for example, BSI can be formalized as

$$I_{BSI} = -g_{AMPA}^{Ctr_1-E_L}s_{AMPA}^{Ctr_1-E_L}\left(V_B - V_E\right) - g_{GABA}^{Ctr_2-E_L}s_{GABA}^{Ctr_2-E_L}\left(V_B - V_I\right) = 0 \quad (7)$$

Assuming that BSI provides a steady input with a mean firing rate r, it can be easily shown that the steady input results in a mean gating variable $s = \tau r$. Let the spike rates of the excitatory and inhibitory components of BSI be r_e and r_i, respectively. Hence we obtain

$$-g_{AMPA}^{Ctr_1-E_L}\tau_{AMPA}r_e\left(V_B - V_E\right) - g_{GABA}^{Ctr_2-E_L}\tau_{GABA}r_i\left(V_B - V_I\right) = 0 \quad (8)$$

which leads to

$$\frac{r_i g_{GABA}^{Ctr_2-E_L}}{r_e g_{AMPA}^{Ctr_1-E_L}} = \frac{\tau_{AMPA}\left(V_B - V_E\right)}{-\tau_{GABAi}\left(V_B - V_I\right)} = -0.4\frac{V_B}{\left(V_B + 70\right)} \quad (9)$$

where $\tau_{GABA} = 5$ ms and $\tau_{AMPA} = 2$ ms. The left side of the equation, $R \equiv r_i g_{GABA}^{Ctr_2-E_L}/r_e g_{AMPA}^{Ctr_1-E_L}$, represents the ratio R between the strength of inhibitory and excitatory components of BSI and is hence defined as BSI ratio. The strength S of BSI is defined as $S \equiv 0.3 r_e g_{AMPA}^{Ctr_1-E_L}$. The purpose of multiplying $r_e g_{AMPA}^{Ctr_1-E_L}$ by 0.3 was to bring the value of the maximum working BSI strength to about 1. We note that all parameters are symmetric between the left side (E_L, Ctr_1 and Ctr_2) and the right side (E_R, Ctr_3 and Ctr_4) of the neural circuit; therefore, BSI ratio and strength for the two decision populations E_L and E_R are identical. The effect of BSI is that it drives the membrane potential toward V_B which acts as an effective reversal potential. When the membrane potential is higher than V_B, BSI produces a hyperpolarized current. If the membrane potential drops below V_B, the BSI current becomes depolarizing.

In a previous study (Wang et al. 2013), we discovered that, depending on the BSI ratio, increasing BSI strength exhibited different influences on speed and accuracy of decision making. When the BSI ratio $R \geq 1.2$ (denoted as I > E), the speed reduces and the accuracy increases with BSI strength. In contrast, when the BSI ratio $R \leq 1.156$ (denoted as I < E), we observed an opposite trend. In the present study, we analyzed the behavioral performance and neuronal activity in I > E and I < E regions, discussed how SAT can be realized with respect to the different BSI settings and compared the results with experimental observations.

In the present study, we mainly tested the BSI ratio between $R = 1.156$ and $R = 1.297$, which correspond to the V_B of −52 mV and −53.5 mV, respectively. This range of membrane potential is below the spike threshold (−50 mV) and is close to the baseline membrane potential (approximately −53 mV) of the neurons when the network is in its baseline state (no random-dot stimulus and no BSI).

Model fits. We fitted the model-predicted psychometric functions to the behavioral data in Palmer et al. (2005) by a maximum likelihood method similar to that used in the same study. Specifically, we maximized a log likelihood (L) function

$$\ln(L) = \sum_{c'} \left[\ln L_P(c') + \ln L_T(c') \right] \quad (10)$$

where

$$L_P(c') = \frac{n!}{k!\,(n-k)!} P(c')^k \left[1 - P(c')\right]^{n-k} \quad (11)$$

represents the likelihood of the observed performance (k correct trials out of n total trials) given the predicted performance $P(c')$ at a coherent level c' and

$$L_T(c') = \frac{1}{w\sqrt{2\pi}} e^{-[t(c') - T(c')]^2/2w^2} \quad (12)$$

represents the likelihood of the observed mean response time $t(c')$, given the predicted mean response time $T(c')$. w is a weight factor for the fitting and is typically given by the standard error of the model-predicted mean response time. However, the magnitude of the predicted standard error was much smaller in the high than in the low coherence level; therefore the fitting biased toward the high coherence levels, i.e., curves in the large c' region fitted better than in the small c' regions. To address the issue, we used a fixed weight ($w = 0.1$) which balanced the fitting between different coherence levels.

RESULTS

Ramping rate of population activity. We first checked the effect of BSI with different values of ratio and strength on an isolated excitatory neuron. We found that, in the I > E regime, BSI exhibits effects that are consistent with the previous finding (Chance et al. 2002) in which BSI reduces the response gain of single neurons, as indicated by a shallower slope of the input-output function [frequency-current (*f-I*) curve] (Fig. 2, *top left*). In contrast, in the I < E regime we observed that increasing BSI strength increases the response gain (Fig. 2, *top left*). We next tested how BSI modulates the decision behavior of the recurrent network model in a simulated random-dots direction discrimination task (see MATERIALS AND METHODS). BSI with an I > E setting significantly slows down the ramping activity of the winning population, while BSI with an I < E setting speeds up the ramping activity (Fig. 2, *top right*). The change of the response gain of single neurons is magnified at the circuit level: while the slope of the initial segment (below 10 Hz in ordinate) of the single neuron *f-I* curve is changed by 10–15% when the BSI strength is varied from 0 to 0.5 for both I > E and I < E settings (Fig. 2, *bottom left*), the mean decision time of the decision circuit is changed by 90% or more (Fig. 2, *bottom right*). The mean ramping rate was estimated by dividing the decision threshold (30 Hz) by the mean decision time (the time it takes from the onset of the random-dots stimulus to the decision threshold crossing). The dramatic difference between the effect of BSI on single neurons and on the population activity originates from the strongly recurrent (attractor) network dynamics, which amplifies considerably any slight change in the single neuron response gain. Therefore, BSI provides a very efficient mechanism for modulating the neural integrator circuit.

Exponential tail of decision time distribution. Interestingly, under the influence of BSI, the circuit model exhibits a large trial-to-trial variability which results in a broad and skewed decision time distribution, as observed in a number of human behavioral studies (Heathcote et al. 1991; Hervey et al. 2006; Luce 1986; Sigman and Dehaene 2005; Usher and McClelland 2001). We plotted the decision time distribution (the probability density function) for each r_{BSI} and found that, when the mean decision time was increased by BSI, it did not simply result from a shift of the peak of the decision time distribution, but also was due to a change in the shape of the distribution (Fig. 3). When BSI with an I > E setting ($S = 0.5$, $R = 1.247$)

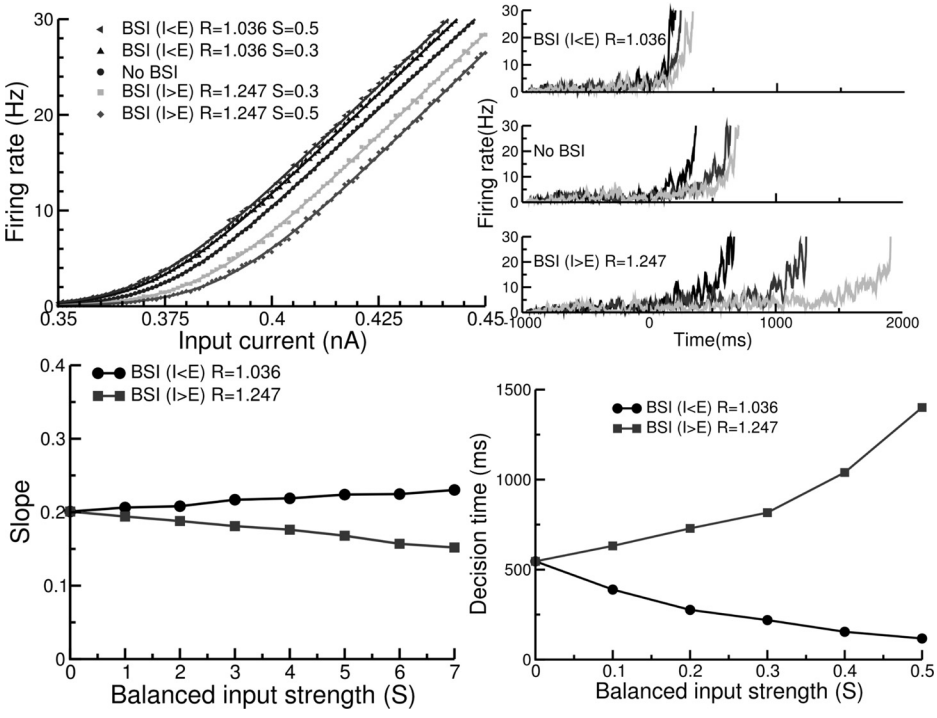

Fig. 2. Effect of BSI on single neurons and on the cortical circuit model. *Top left*: single-neuron input-output relationship [frequency-current (*f-I*) curve] for different BSI strengths (S) and ratios (R). *Top right*: sample traces of neural ramping activity with different BSI strengths and ratios during the decision task. A high BSI ratio (more inhibition) reduces the ramping rate of the population activity which results in a longer mean decision time, while a low BSI ratio (more excitation) increases the ramping rate which causes a smaller mean decision time. The decision time is defined by the threshold (30 Hz) crossing by the ramping neural firing rate (r), and $t = 0$ indicates the onset of the stimulus. *Bottom left*: mean slope of the initial segment (below 10 Hz) of the single neuron *f-I* curve. *Bottom right*: mean decision time as functions of BSI strength for high BSI ratio (squares) and low BSI ratio (circles). The change in the slope of the *f-I* curve at the single neuron level is magnified at the circuit level due to strongly recurrent (attractor) dynamics. The stimulus motion coherence c' is 3.2% for *top right* and *bottom* panels.

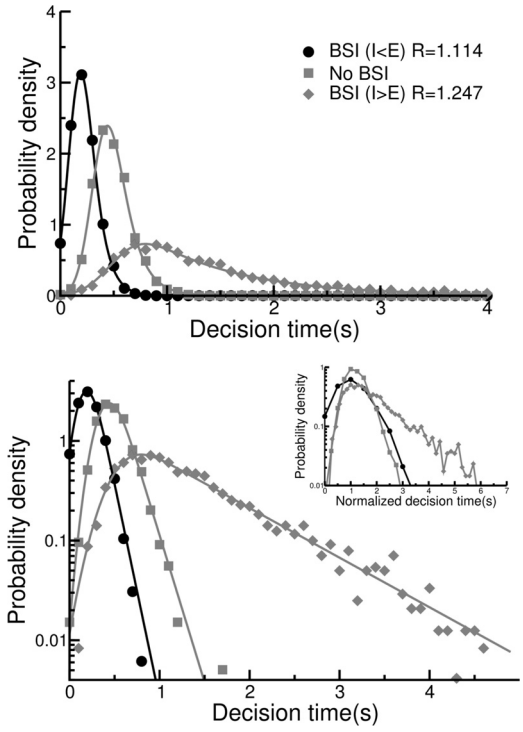

Fig. 3. BSI produces a skewed decision time distribution with a long exponential tail. *Top*: decision time distributions with BSI ($S = 0.5$) of different ratios or without BSI with $c' = 3.2\%$. High ratio BSI increases the mean decision time by shifting the peak, as well as by producing a long right tail. The simulated data are fitted using ex-Gaussian probability density function (solid curves). *Bottom*: same distributions and curve fittings as in *A* but plotted in logarithmic scale along the *y*-axis. Strong BSI produces a long tail following the exponential form, in contrast to the short tail with a more symmetric form in the case of no BSI ($S = 0$). *Inset*: to further demonstrate that the high ratio balanced input increases the skewness of the decision time distribution, we horizontally rescale the distributions so that the peak positions of the distributions are all equal to 1.

was applied to the circuit model, the distribution shifted to the right, while at the same time developed a long exponential tail compared with the no BSI condition (Fig. 3). In contrast, when BSI with an I < E setting ($S = 0.5$, $R = 1.114$) was applied to the model, the distribution shifted to the left with a more symmetric shape.

We further found that the simulated distributions could be well fitted by the ex-Gaussian function, which is an exponentially modified Gaussian function often used to describe response time distributions in various human decision tasks (Ratcliff 1978, 1993; Ratcliff and Rouder 1998). The ex-Gaussian function is the convolution of the Gaussian and the exponential functions and has three parameters: μ for the mean, σ for the standard deviation of the Gaussian component, and τ for the time constant of the exponential component. The Gaussian component forms the peaked distribution with a symmetric shape, while the exponential component gives rise to a long tail. The three parameters (μ, σ, τ) were (0.120, 0.099, 0.101), (0.345, 0.123, 0.147) and (0.500, 0.218, 0.873), and the χ^2 errors of the fittings (divided by the number of bins) are 9.6×10^{-4}, 2.0×10^{-3} and 4.1×10^{-3} for the BSI (I > E), No BSI and BSI (I < E) conditions, respectively. We found that the time constant τ of the exponent component increases with the BSI ratio. We quantified the relative contributions of the exponential and the Gaussian components to the overall shape of the response time distribution by calculating τ/σ, which increases from 1.02 for I < E to 4.00 for I > E conditions. We note that a larger contribution from the exponential component produces a more skewed decision time distribution with a long tail. To visualize the skewness of the decision time distributions, we rescaled the distribution as follows. Let $P(t)$ denote the distribution of decision time t for a given BSI condition, and t_{peak} be the decision time at the peak of the distribution. We rescaled $P_r(t)$ by calculating $P(t/t_{peak}) \times t_{peak}$, so that the peak positions for different BSI conditions are aligned, while the total area of each distribution remains unchanged. The result shows that the decision time distributions for different BSI conditions have different shapes (Fig. 3, *bottom*, *inset*), with a larger BSI ratio giving rise to a longer exponential tail.

SAT: behavioral performance. Our previous study (Wang et al. 2013) showed that the speed and accuracy can be simultaneously adjusted by changing the BSI strength and/or the ratio (Fig. 4, *A* and *B*). This gives a subject the flexibility to perform SAT with a wide range of physiological settings. For example, speed can be traded for accuracy if we change the BSI strength while keeping the BSI ratio as a constant (*path 1* in Fig. 4, *A* and *B*). Alternatively, one can also change the BSI ratio while keeping the BSI strength as a constant (*path 2* in Fig. 4, *A* and *B*).

We next compared the two BSI strategies (*paths 1* and *2*) with the DDM. Note that, in our model, we kept the thresholds symmetric (same for both E_L and E_R populations) and unchanged (30 Hz). We plotted the performance and the decision time as functions of stimulus motion strength for both BSI strategies. For each strategy, we selected three settings that produce long, medium and short mean decision times (Fig. 5, *A* and *B*). We found that the performance and decision time curves of both BSI strategies could be well fitted by DDM in a way that resembled what has been observed in a study of human SAT based on the same random-dot task (Palmer et al. 2005). In the study, the authors fitted the behavioral performance by DDM and found that the normalized decision bound θ varies significantly if the subjects were instructed to perform the task at different speeds, while the other fitting parameters, sensitivity (k) and the residual time (t_R), only changed slightly. The t_R (or the t_{nd}) represents the neural latency that is not part of the ramping activity. In the DDM, the performance P and the mean decision time T depend on θ and k by the following equations (Palmer et al. 2005):

$$P(c') = \frac{1}{1 + e^{-2k\theta c'}}$$

$$T = \frac{\theta}{kc'}\tanh(\theta k c') + t_R$$

where c' is the coherence level of the random dot stimulus. The drift rate in the model is given by kc'. The normalized bound θ is derived by dividing the decision bound by the standard deviation σ of the noise. The normalization turns the model from three parameters (decision bound, σ and the t_R) into two (θ and t_R). To compare our model with the DDM, we fit the two equations shown above to the simulated behavior produced by our spiking network model with the two BSI strategies. Interestingly, we found that both BSI strategies produce behavioral

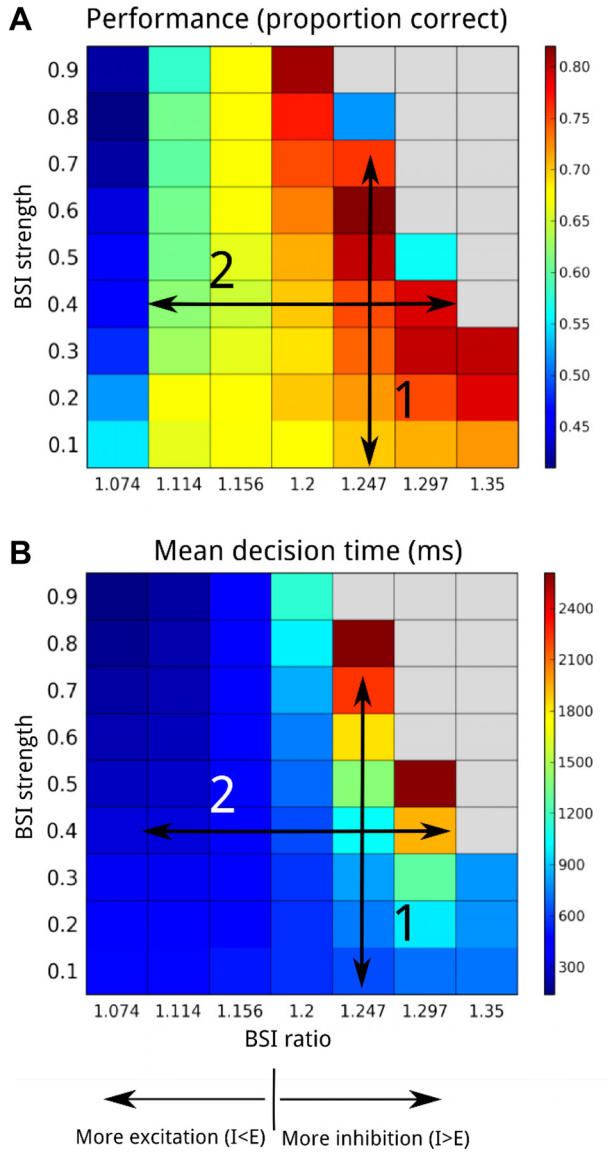

Fig. 4. Speed-accuracy tradeoff (SAT) can be realized in the model in different ways. By changing BSI ratio and strength, we can achieve various levels (indicated by color) of performance (*A*) and mean decision time (*B*). We can divide the parameter space (strength and ratio) into two regimes. In the "more excitation" regime (I < E), increasing BSI strength reduces the performance but increases the speed, while in the "more inhibition" region (I > E), increasing BSI strength improves the performance but reduces the speed. We investigated two specific BSI strategies: taking the *path 1* (the up-down arrow) by changing BSI strength with fixed ratio, or the *path 2* (the left-right arrow) by changing BSI ratio with fixed strength. [Adapted from Wang et al. 2013].

effects that correspond to a significant change of θ, while k and t_R only changed slightly (Fig. 5*C*). For the constant ratio case ($R = 1.247$), three different values of strength $S = 0.1, 0.4, 0.6$ correspond to DDM with parameters $\theta = 0.74, 1.01, 1.46$; $k = 16.2, 15.2, 15.0$; and $t_R = 136$ ms, 148 ms, 154 ms, respectively. For the constant strength case ($S = 0.4$), three different values of ratio $R = 1.156, 1.247, 1.297$ correspond to DDM with parameters $\theta = 0.581, 1.01, 1.51$; $k = 13.5, 15.2, 13.3$; and $t_R = 101$ ms, 148 ms, 122 ms, respectively. Therefore, at the behavioral level, changing the strength of BSI with a constant ratio or changing the ratio with a constant strength in the neural circuit model produce similar effects with varying the decision bound in the DDM. In other words, showing that the behavioral data in SAT can be fitted by the DDM with varying θ and constant k does not necessarily mean that the accumulation rate of the underlying system dynamics is invariant and the decision threshold changes under SAT.

We took a further step by fitting our model to the behavioral data of the subject JP in Palmer et al. (2005) with the maximum likelihood method (*Eqs. 10–12*). For each task condition (fast, neutral or accurate), we maximized the log likelihood function (*Eq. 10*) by tuning the BSI strength (S), BSI ratio (R) and the t_{nd}, while leaving all the rest neuronal and network parameters fixed. In addition, we tuned μ_A and μ_B (*Eqs. 1* and *2*) under the constraint that they are invariant across task conditions. These two parameters describe the relationship between the visual stimuli and the actual inputs to the model and, therefore, affect the shape of the psychometric functions. For each combination of the tuning parameters, we performed 3,000 simulations trials for every coherence levels and then calculated the performance (proportion correct) and mean response time, which were used to calculate the likelihood functions. We found that our model fitted the observed performances and mean response times reasonably well (Fig. 6*A*). The best-fit values of (S, R, t_{nd}) are (0.1, 1.114, 30 ms), (0.1, 1.407, 20 ms) and (0.225, 1.350, 0 ms) for the fast, neutral and accurate conditions, respectively. The best values for μ_A and μ_B were found to be 114 Hz and 76 Hz, respectively, for all three conditions. The difference in the BSI settings between the fast and the neutral conditions mainly lies in the ratio, while both strength and ratio changed between the neutral and the fast conditions. Intriguingly, we found that, in addition to the mean response times, the response time distributions of the model also match very well to the data, in particular for the difficult tasks (low coherence levels) (Fig. 6*B*).

SAT: neural activity. We have shown that adjusting BSI settings produces SAT at the behavioral level. Next, we asked whether the simulated neuronal activity resembles those observed in primate SAT tasks. A recent monkey study (Heitz and Schall 2012) with single-unit recordings may shed light on the question. In the study, the authors measured activity of FEF neurons of monkeys in a visual search task with cues for fast, accurate or neutral responses. While the perceptual decision task was different from the random-dot paradigm, neurons in FEF also exhibit decision and SAT-related neural activity. In particular, averaged firing rate of movement neurons exhibit variable ramping rate across different SAT conditions, and some (29%) movement neurons display SAT-dependent baseline shifts (Fig. 7*C*). We analyzed the simulated neuronal activity in our model and discovered that, while both BSI mechanisms exhibited modulated ramping rate as in the observed data, changing the BSI ratio resulted in a much larger modulation of the baseline firing rate (Fig. 7*A*) than changing the BSI strength did (Fig. 7*B*). Therefore, changing the BSI ratio with a constant strength leads to the modulatory effects that resemble observed neuronal activity (Fig. 7*C*).

Top-down modulation with pure excitation or inhibition. We have demonstrated that BSI with various settings of strength and ratio resembles the observed responses under different speed instructions. The result poses an important question: if the SAT can be realized in our model by slightly "unbalanced" (I > E or I < E) BSI, can we achieve the same effect by simply

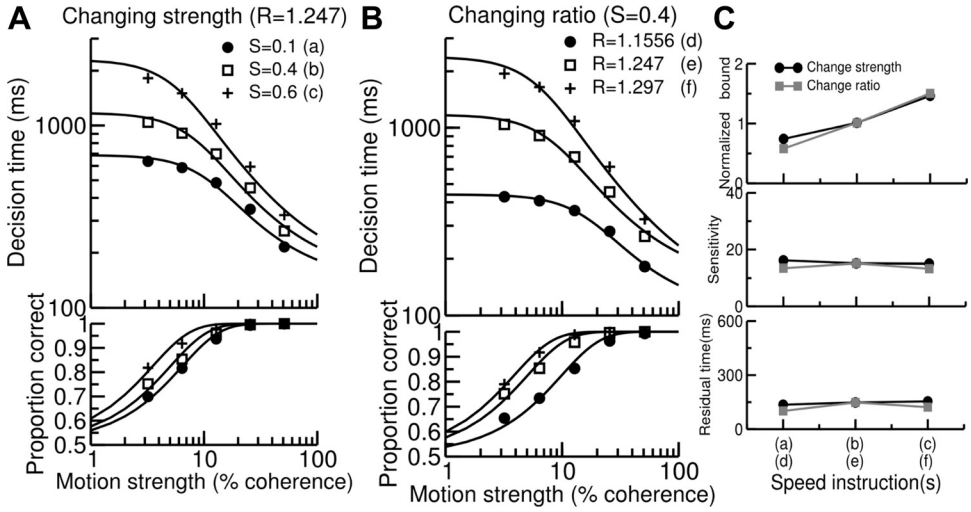

Fig. 5. Varying BSI strength or ratio in the neural circuit model produces the same behavioral effect as changing the decision bound (threshold) in the drift diffusion model (DDM). *A*: the simulated performance (*top*) and mean decision time (*bottom*) with a constant BSI ratio ($R = 1.247$) but different strengths ($S = 0$, 0.4 and 0.6). The simulated performance and mean decision time were fitted by DDM (solid curves) using Levenberg-Marquardt algorithm (Nocedal and Wright 2006) with weighting of $1/y$, where y is the ordinate of the data. *B*: same as in *A* but with a constant strength ($S = 0.4$) and different ratio ($R = 1.156$, 1.247, 1.297). *C*: Three fitting parameters in the DDM, normalized decision bound (*top*), sensitivity (*middle*) and residual time (*bottom*), as functions of BSI strength (black) and ratio (gray). The result showed that changing BSI settings in the cortical circuit model led to significant decision bound changes in DDM, while the other two parameters only change slightly.

applying a small amount of input with pure excitation or inhibition? To address the question, we replaced BSI by synaptic input with pure excitation or inhibition. To compare with BSI, we quantified the strength of the excitatory or inhibitory input on the same basis: first, the synaptic weight of the pure excitation or inhibition was the same with that used for BSI (0.1 nS), and, second, following how we quantified the BSI strength, we also defined the strength of the input as $S \equiv 0.3rg$, where r is the input firing rate, and g is the synaptic conductance. Our tests showed that the network model is very sensitive to the change in the pure excitatory or inhibitory input (Fig. 8). Although a small amount of excitation ($S = 0.05$) or inhibition ($S = 0.03$) did improve the performance and increase the decision time, slightly stronger input (0.1 for excitation or 0.05 for inhibition) dramatically suppressed the ability of the network to make a decision and led to poor performance and increased number of nondecision trials (as labeled in Fig. 8*B*). The nondecision trials were characterized by two trial outcomes: *1*) failed winner-take-all competition in which both decision populations had their firing rates ramp up simultaneously and sometimes prematurely; and *2*) failed decisions in which none of the decision neural populations reached the decision threshold by the end of the trials. The sensitivity of the network shows that purely excitatory or inhibition input does not provide a suitable mechanism for fine control of SAT in the noisy neural environment.

DISCUSSION

In the present study, we investigated how a top-down control signal, implemented as BSI, modulates perceptual decisions in a cortical circuit model, and our results are threefold. First, by changing BSI ratio and/or strength, we can alter the slope or rate of neural population ramping activity, resulting in faster

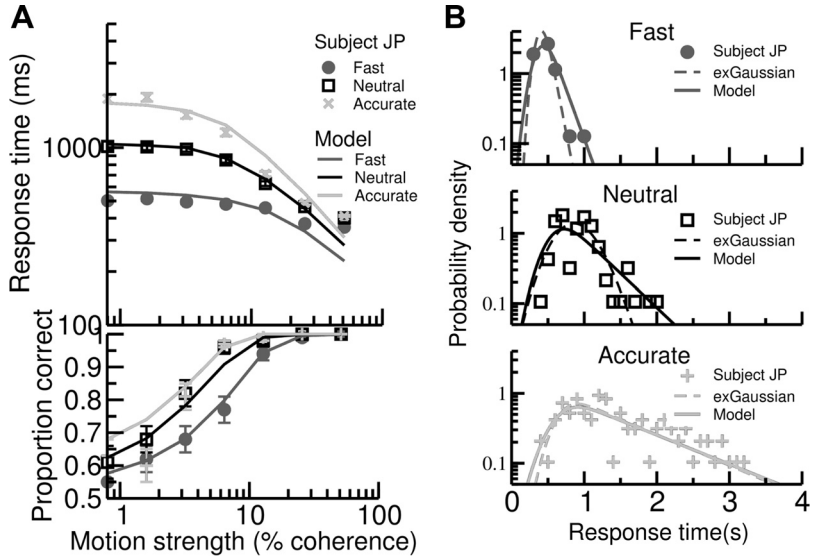

Fig. 6. The cortical circuit model fitted reasonably well to the behavioral data in a human SAT experiment (Palmer et al. 2005). *A*: We varied BSI ratio, BSI strength, nondecision time, and parameters of stimulus input (μ_A and μ_B) and then selected the settings that maximized a log likelihood function for each of the three task conditions (fast, neutral and accurate) for the subject JP. *B*: we overlapped the simulated response time distributions with those from the data (same subject) for $c' = 3.2\%$ and found that they matched very well in all three conditions. We also fitted the ex-Gaussian curves (dashed lines) to the data for comparison.

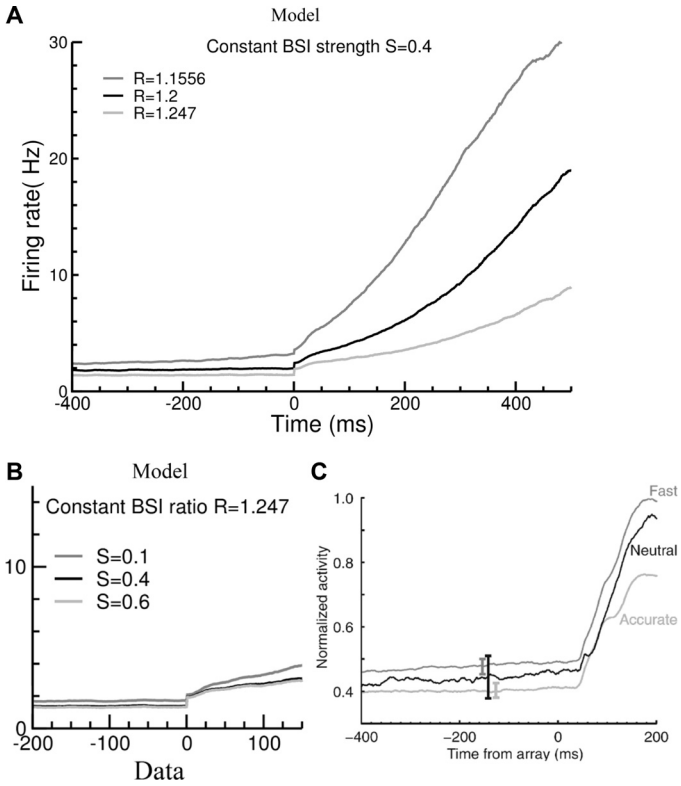

Fig. 7. SAT via changing BSI ratio modulates baseline neural activity in a way that is consistent with a recent primate study (Heitz and Schall 2012). *A*: we calculated the trial-average *r* of the winning decision populations in the correct trials for the two BSI mechanisms. By changing the BSI ratio while keeping the strength as a constant ($S = 0.4$), we observed higher baseline levels evoked by lower BSI ratio. *B*: the change in the baseline level is much smaller (<1 Hz) when SAT was implemented by changing BSI strength with a constant ratio ($R = 1.247$). *C*: a similar trend as in *A* was observed for movement neurons in frontal eye field in the primate experiment (Heitz and Schall 2012). Note that the baseline adjustment was also observed for some movement neurons, but the effect is obscured here by averaging across neurons with and without the effect (Heitz and Schall 2012). *C* is adapted from Heitz and Schall 2012 [reprinted with permission from Elsevier].

but less accurate, or slower but more accurate decisions. Second, SAT can be realized by different strategies: changing either strength or ratio, or both. Third, the model-predicted behavior can be well fitted by only varying the decision threshold in the DDM, thereby demonstrating that behavioral data fitting by the DDM are insufficient to conclude that the ramping slope of neuronal activity remains the same under speed vs. accuracy emphasis.

Our results are consistent with recent empirical studies which observed SAT-dependent ramping rate (Hanks et al. 2014; Heitz and Schall 2012). Furthermore, another two recent studies implemented more sophisticated approaches in model selection and demonstrated that SAT could be accounted for by varying the rate of information accumulation in the drift diffusion and the linear ballistic accumulator models (Cassey et al. 2014; Rae et al. 2014). Therefore, our neuron-level model does not contradict those system-level models, but rather provides more insights into neural circuit mechanisms of SAT and paves the way for establishing links between models at the system and neuronal levels. Although we intend to reproduce various features of neural activity observed in SAT experiments, with a few homogeneous populations, our model was only able to qualitatively reproduce SAT modulated ramping rate and baseline activity, such as those observed in FEF movement neurons (Heitz and Schall 2012). It remains to be studied how our model can be expanded to reproduces diverse activity exhibited by different neuron types, including visual and visuomovement neurons.

We noted that our model fits better to the response time data at the low- than at the high-coherence levels (Fig. 6*A*). We suspect that this is due to different top-down control exerted by the subjects at different coherence levels. In the experiments (Palmer et al. 2005), the subjects were instructed to match the target response time for the most difficult condition (lowest motion coherence). Therefore, the subjects experienced less speed stress at high coherence levels and might implement different top-down control strategies. New experimental designs and more data are necessary for identifying the potential differences in the top-down control across stimulus conditions.

We noted that the threshold in the proposed circuit model acts more like a movement threshold which determines the time for the downstream circuits, such as the basal ganglia, to trigger a saccade (Hsiao and Lo 2013; Lo and Wang 2006). Nevertheless, we would like to emphasize that our proposal of

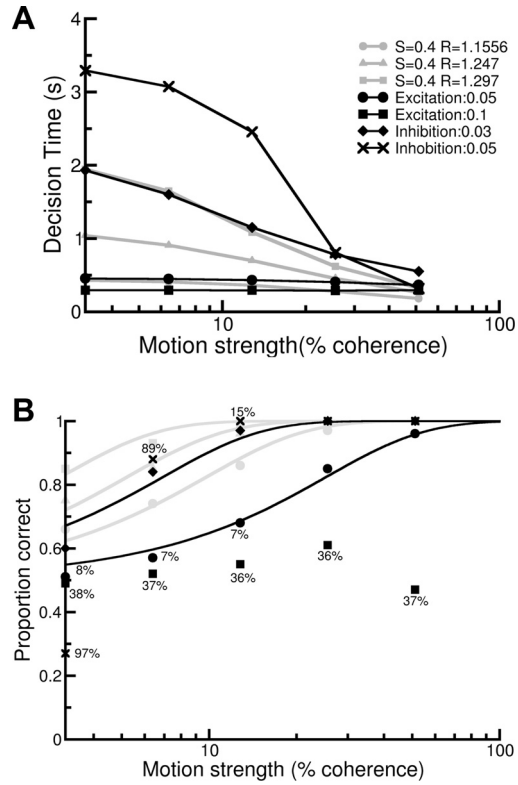

Fig. 8. Purely excitatory or inhibitory input does not provide a better mechanism of SAT than BSI due to the sensitivity of the decision circuit to the input. *A*: mean decision time as functions of stimulus motion strength for different common input conditions. The mean decision times increased dramatically when we added a weak inhibitory input to the decision neural populations (circles and squares). In contrast, an opposite trend was observed when we added a weak excitatory input (diamonds and x's). For comparison, gray lines were added to indicate the mean decision times for the three BSI settings as in Fig. 5*B*. *B*: performance as functions of stimulus motion strength for the same input conditions as in *A*. Adding weak excitatory or inhibitory input dramatically increases the percentage of nondecision trials, indicated by the numbers next to the symbols for those with a large percentage (>5%) of nondecision trials. For comparison, gray lines were added to indicate the performance for the three BSI settings as in Fig. 5*B*.

BSI does not necessarily exclude "threshold tuning" as a mechanism of SAT (Hsiao and Lo 2013; Lo and Wang 2006). In fact, two recent studies (Heitz and Schall 2012; Jantz et al. 2013) observed a variable threshold in movement-related neuron across task conditions. It would be interesting to include downstream premotor circuits in our model for threshold tuning in future studies.

In addition to BSI and decision threshold tuning, several theoretical studies have suggested that the behavioral performance can also be controlled by various mechanisms, including locus coeruleus modulated gain transients (Shea-Brown et al. 2008), the urgency or timing signals (Churchland et al. 2008; Hanks et al. 2014; Standage et al. 2011, 2013), and common excitatory input to the decision neural populations (Furman and Wang 2008; Roxin and Ledberg 2008; Standage et al. 2011). The urgency signal plays a top-down modulatory role that is similar to the BSI mechanism. BSI modulates the accelerated ramping activity exhibited by the attractor network, while the urgency signal introduces variable acceleration to the linear ramping activity of DDM; likewise, the Ornstein-Uhlenbeck process with an acceleration term instead of a leakage term (Usher and McClelland 2001) may also exhibit similar SAT properties. It is worth to investigate, under a rigorous mathematical framework, the correlation between these mechanisms and whether the proposed BSI mechanism can be viewed as a biological implementation of the concept of urgency signals and the accelerating Ornstein-Uhlenbeck process. The common excitatory input can be treated as a limiting case under the framework of the BSI in which the BSI ratio approaches 0. When varying the common excitatory input around a critical value that characterizes a bifurcation, the system is very sensitive to the magnitude of the common input; hence one can easily make a significant alteration to the system dynamics by just a slight change in the input level. The drawback is, as we have shown, that it is difficult to make a precise control for the SAT if the system has a large intrinsic noise, as in most neural systems. On the other hand, in a system that has the capability of precise input control with small noise, the common excitatory input may work as an alternative mechanism of SAT. The proposal of the timing signal in Standage et al. (2013) described a mechanism that encodes elapsed time of a trial by a hypothetical timing network. The mechanism can be viewed as an urgency signal in the form of common excitatory input. We suggest that, by integrating the BSI mechanism into the timing network, one can greatly improve the controllability of this model in terms of SAT.

We have previously shown that BSI alters the dynamics of the system and increases or decreases the local stability ("the crater") at the center of the "energy landscape" under the high ratio ($I > E$) or the low ratio ($I < E$) settings, respectively (Wang et al. 2013) (Fig. 9). The changes in the landscape explain the changes of decision behavior as well as the changes in the neuronal ramping activity under different BSI settings (Wang et al. 2013). Interestingly, if we consider the landscape before the onset of the stimulus, BSI-modulated local stability can also explain the increased baseline firing rate with the speed instruction, as observed in our model and in monkey experiments (Fig. 9, *A* and *C*). Moreover, after the stimulus onset, in the case where the local stability still exists (Fig. 9*D*), the firing rate of the decision neurons exhibits a two-phase

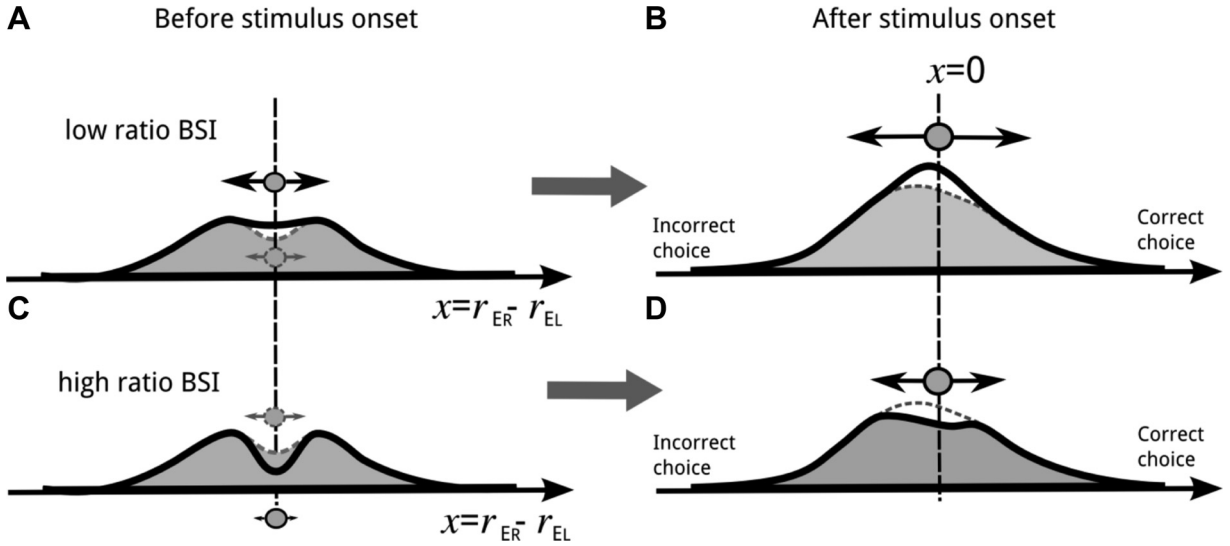

Fig. 9. Schematic plot illustrating how neural dynamics are changed due to BSI-modulated alternation of the potential surface. The *x*-axis represents the difference between the *r* of the two decision populations ($r = r_{ER} - r_{EL}$). From the dynamic system point of view, decision making can be described as a process of falling into one of the two basins (decision attractors) for a particle that starts from the center of an effective "energy landscape" (Wang et al. 2013; Wong and Wang 2006). The energy landscape can be estimated by calculating $\Delta r/\Delta t$ as a function of *r* across a large number of simulation trials. See Wang et al. 2013 for details. *A*: in the no-BSI condition, there is a stable baseline state presented by a deep "crater," or a local minimum (dashed curve), at the center of the landscape prior to the stimulus onset. A general effect of adding a low ratio BSI is to increase the excitability of the network, represented by increased energy level at the baseline state ($r_{ER} - r_{EL}$), or a shallower center energy well (black curve) (Wang et al. 2013). As a result, the network becomes less stable, and the particle moves in a wider range (larger *x*), which leads to higher baseline *r* for both decision populations. *B*: after the stimulus onset, BSI produces a sharper peak of the potential surface. As a result, the particle falls into one of the basins quickly and leads to a shorter mean response time. *C*: in contrast, if a high ratio BSI is applied to the network before the onset of the stimulus, the system becomes more stable and the center energy well is deepened. As a result, the particle moves in a narrower range during the baseline activity, and smaller mean baseline *r* are observed. *D*: after the onset of the stimulus, the high ratio BSI again lowers the energy level at the center, causing a local energy minimum which greatly stabilizes the network. As a result, it prolongs the time required for the particle to fall into one of the basins and thus increases the mean response time. In *A* and *C*, gray arrows indicate the movement ranges under no-BSI condition, and black arrows are for those under the BSI condition.

ramping activity (Fig. 2, *top right*, *bottom* panel): a slow ramping phase (inside the crater) followed by a fast ramping phase (out of the crater). This property shows an interesting parallel between BSI and the gated inhibition mechanism in which an inhibitory control gates the activity of accumulator units (Purcell et al. 2012). The duration of the early "slow" phase in our model depends on the size of the crater and is, therefore, determined by the property of BSI. The start of the fast ramping phase in our model may correspond to the delay onset time of movement neurons found in the FEF (Pouget et al. 2011). Further investigations are needed to identify the correlation between the different mechanisms.

As demonstrated in other theoretical studies of attractor networks, changing between different attractor state greatly affects response times (Deco et al. 2009; Deco and Rolls 2006; Eckhoff et al. 2011; Miller and Katz 2010). Furthermore, we have previously shown (Wang et al. 2013) that the existence of the local stability at the center of the energy landscape is associated with the long tail of the response time distribution. This is because the local energy minimum temporally traps the system in the beginning of each trial and hence partially contributes to prolonged response times. In the present study, we show that the long tail can be well fitted by an exponential function. The result is intriguing, considering that highly skewed decision time distributions with exponential tails have been observed in various human behavioral tasks (Heathcote et al. 1991; Hervey et al. 2006; Luce 1986; Niwa and Ditterich 2008; Ratcliff 1978; Sigman and Dehaene 2005; Usher and McClelland 2001). On the other hand, in the random-dots motion discrimination experiment, decision time distributions of behaving monkeys often appear to be more Gaussian-like (Ditterich 2006; Roitman and Shadlen 2002). Such differences might arise for several reasons, such as the details of task design, the amount of training, and differences across species. Apart from these factors, our finding reported in this paper suggests that the skewness of decision time distribution could, in principle, be explained by a varying degree of top-down control in the form of BSI. Further studies could be designed to test our hypothesis.

The attractor dynamics of our model is characterized by stronger influence of early dynamics than the later dynamics on an ultimate choice. The property might not seem to be optimal for a decision neural network because this implies that the system does not integrate the signal in the entire course of a trial. However, this property has been confirmed in a monkey experiment (Huk and Shadlen 2005; Wong et al. 2007). Two other important characteristics of the attractor dynamics have also been observed in experiments: *1*) the performance cannot be improved indefinitely by increasing the stimulus viewing time, i.e., raising the decision bound (Kiani et al. 2008); and *2*) the ramping neuronal activity exhibits a two-phase dynamics with an early slow-ramping phase followed by a late rapid-ramping phase (Roitman and Shadlen 2002).

In summary, our work suggests that BSI can serve as a mechanism of top-down control that performs a rapid modulation on a neural circuit of perceptual decision. One may note that our proposal of the top-down control in the form of balanced excitation and inhibition seems to be different from one of our laboratory's earlier studies (Lo et al. 2009) in which a planned movement is suppressed by a purely excitatory top-down control to a population of "stop" neurons. However, this difference can be unified under a general framework by considering our recent study (Wang et al. 2013) in which BSI can be formed by a pure excitation that targets both excitatory and inhibitory local neural populations. Therefore, whether the goal is to stop an action or is to perform SAT, the top-down control can be realized by excitatory projections to different part of the circuit. Indeed, various studies have suggested that the prefrontal cortex plays a critical role in top-down executive control (Asplund et al. 2010; Isoda and Hikosaka 2007; Miller and Cohen 2001; Ridderinkhof 2004; Tomita et al. 1999). Long-range excitatory projections from the prefrontal cortex to posterior cortical areas target both pyramidal cells and inhibitory interneurons (Barbas et al. 2005; Bunce and Barbas 2011; Medalla et al. 2007). The recipient inhibitory neurons could in turn target pyramidal cells and thus instantiate the inhibitory part of a top-down BSI (Vogels and Abbott 2009). Our work underscores the importance of understanding the physiological properties of such long-distance pathways, in microcircuit details, in particular the excitation and inhibition balance of top-down signaling by the prefrontal cortex.

ACKNOWLEDGMENTS

We thank S. Fusi for the early work in the neural simulator used in this study and the National Center for High-Performance Computing, Taiwan, for providing computational resources. We thank Drs. John Palmer, Alexander Huk, and Michael Shadlen for providing data used in Palmer et al. (2005).

GRANTS

This work was supported by National Institute of Mental Health Grant R01-MH062349 (X.-J. Wang), the Swartz Foundation (C.-C. Lo and X.-J. Wang), and Ministry of Science and Technology (Taiwan) (C.-C. Lo and C.-T. Wang).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the author(s).

AUTHOR CONTRIBUTIONS

Author contributions: C.-C.L. and X.-J.W. conception and design of research; C.-C.L. and C.-T.W. analyzed data; C.-C.L. and X.-J.W. interpreted results of experiments; C.-C.L. and C.-T.W. prepared figures; C.-C.L. drafted manuscript; C.-C.L. and X.-J.W. edited and revised manuscript; C.-C.L. and X.-J.W. approved final version of manuscript; C.-T.W. performed experiments.

REFERENCES

Abbott LF, Chance FS. Drivers and modulators from push-pull and balanced synaptic input. *Prog Brain Res* 149: 147–155, 2005.

Anderson JS. The contribution of noise to contrast invariance of orientation tuning in cat visual cortex. *Science* 290: 1968–1972, 2000.

Asplund CL, Todd JJ, Snyder AP, Marois R. A central role for the lateral prefrontal cortex in goal-directed and stimulus-driven attention. *Nat Neurosci* 13: 507–512, 2010.

Ayaz A, Chance FS. Gain modulation of neuronal responses by subtractive and divisive mechanisms of inhibition. *J Neurophysiol* 101: 958–968, 2008.

Balci F, Simen P, Niyogi R, Saxe A, Hughes JA, Holmes P, Cohen JD. Acquisition of decision making criteria: reward rate ultimately beats accuracy. *Atten Percept Psychophys* 73: 640–657, 2010.

Barbas H, Medalla M, Alade O, Suski J, Zikopoulos B, Lera P. Relationship of prefrontal connections to inhibitory systems in superior temporal areas in the rhesus monkey. *Cereb Cortex* 15: 1356–1370, 2005.

Berg RW, Alaburda A, Hounsgaard J. Balanced inhibition and excitation drive spike activity in spinal half-centers. *Science* 315: 390–393, 2007.

Bogacz R, Brown E, Moehlis J, Holmes P, Cohen JD. The physics of optimal decision making: a formal analysis of models of performance in two-alternative forced-choice tasks. *Psychol Rev* 113: 700–765, 2006.

Bogacz R, Wagenmakers EJ, Forstmann BU, Nieuwenhuis S. The neural basis of the speed-accuracy tradeoff. *Trends Neurosci* 33: 10–16, 2010.

Britten KH, Shadlen MN, Newsome WT, Movshon JA. Responses of neurons in macaque MT to stochastic motion signals. *Vis Neurosci* 10: 1157–1169, 1993.

Brozovic M, Abbott LF, Andersen RA. Mechanism of gain modulation at single neuron and network levels. *J Comput Neurosci* 25: 158–168, 2008.

Brunel N, Wang XJ. Effects of neuromodulation in a cortical network model of object working memory dominated by recurrent inhibition. *J Comput Neurosci* 11: 63–85, 2001.

Bunce JG, Barbas H. Prefrontal pathways target excitatory and inhibitory systems in memory-related medial temporal cortices. *Neuroimage* 55: 1461–1474, 2011.

Burkitt AN, Meffin H, Grayden DB. Study of neuronal gain in a conductance-based leaky integrate-and-fire neuron model with balanced excitatory and inhibitory synaptic input. *Biol Cybern* 89: 119–125, 2003.

Cassey P, Heathcote A, Brown SD. Brain and behavior in decision-making. *PLoS Comput Biol* 10: e1003700, 2014.

Chance FS, Abbott LF, Reyes AD. Gain modulation from background synaptic input. *Neuron* 35: 773–782, 2002.

Churchland AK, Kiani R, Shadlen MN. Decision-making with multiple alternatives. *Nat Neurosci* 11: 693–702, 2008.

Deco G, Rolls ET. Decision-making and Weber's law: a neurophysiological model. *Eur J Neurosci* 24: 901–916, 2006.

Deco G, Rolls ET, Romo R. Stochastic dynamics as a principle of brain function. *Prog Neurobiol* 88: 1–16, 2009.

Ditterich J. Evidence for time-variant decision making. *Eur J Neurosci* 24: 3628–3641, 2006.

Eckhoff P, Wong-Lin K, Holmes P. Dimension reduction and dynamics of a spiking neural network model for decision making under neuromodulation. *SIAM J Appl Dyn Syst* 10: 148–188, 2011.

Edwards W. Optimal strategies for seeking information: models for statistics, choice reaction times, and human information processing. *J Math Psychol* 2: 312–329, 1965.

Forstmann BU, Dutilh G, Brown S, Neumann J, von Cramon DY, Ridderinkhof KR, Wagenmakers EJ. Striatum and pre-SMA facilitate decision-making under time pressure. *Proc Natl Acad Sci U S A* 105: 17538–17542, 2008.

Furman M, Wang XJ. Similarity effect and optimal control of multiple-choice decision making. *Neuron* 60: 1153–1168, 2008.

Gold JI, Shadlen MN. Banburismus and the brain. *Neuron* 36: 299–308, 2002.

Haider B, Duque A, Hasenstaub AR, McCormick DA. Neocortical network activity in vivo is generated through a dynamic balance of excitation and inhibition. *J Neurosci* 26: 4535–4545, 2006.

Hanks T, Kiani R, Shadlen MN. A neural mechanism of speed-accuracy tradeoff in macaque area LIP. *eLife* 3: e02260, 2014.

Heathcote A, Popiel SJ, Mewhort DJ. Analysis of response time distributions: an example using the Stroop task. *Psychol Bull* 109: 340–347, 1991.

Heitz RP. The speed-accuracy tradeoff: history, physiology, methodology, and behavior. *Front Neurosci* 8: 150, 2014.

Heitz RP, Schall JD. Neural mechanisms of speed-accuracy tradeoff. *Neuron* 76: 616–628, 2012.

Hervey AS, Epstein JN, Curry JF, Tonev S, Eugene Arnold L, Keith Conners C, Hinshaw SP, Swanson JM, Hechtman L. Reaction time distribution analysis of neuropsychological performance in an ADHD sample. *Child Neuropsychol* 12: 125–140, 2006.

Hô N, Destexhe A. Synaptic background activity enhances the responsiveness of neocortical pyramidal neurons. *J Neurophysiol* 84: 1488–1496, 2000.

Hsiao PY, Lo CC. A plastic corticostriatal circuit model of adaptation in perceptual decision making. *Front Comput Neurosci* 7: 178, 2013.

Huk AC, Shadlen MN. Neural activity in macaque parietal cortex reflects temporal integration of visual motion signals during perceptual decision making. *J Neurosci* 25: 10420–10436, 2005.

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

Jantz JJ, Watanabe M, Everling S, Munoz DP. Threshold mechanism for saccade initiation in frontal eye field and superior colliculus. *J Neurophysiol* 109: 2767–2780, 2013.

Kiani R, Hanks TD, Shadlen MN. Bounded integration in parietal cortex underlies decisions even when viewing duration is dictated by the environment. *J Neurosci* 28: 3017–3029, 2008.

Liu F, Wang XJ. A common cortical circuit mechanism for perceptual categorical discrimination and veridical judgment. *PLoS Comput Biol* 4: e1000253, 2008.

Lo CC, Boucher L, Pare M, Schall JD, Wang XJ. Proactive inhibitory control and attractor dynamics in countermanding action: a spiking neural circuit model. *J Neurosci* 29: 9059–9071, 2009.

Lo CC, Wang XJ. Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. *Nat Neurosci* 9: 956–963, 2006.

Lo CC, Wang XJ. Functional tuning of a decision neural network by top-down balanced synaptic input produces skewed reaction time distributions with a long tail. Program no. 803.2. In: *2009 Neuroscience Meeting Planner*. Chicago, IL: Society for Neuroscience, 2009.

Luce RD. *Response Times*. New York: Oxford University Press, 1986.

Mariño J, Schummers J, Lyon DC, Schwabe L, Beck O, Wiesing P, Obermayer K, Sur M. Invariant computations in local cortical networks with balanced excitation and inhibition. *Nat Neurosci* 8: 194–201, 2005.

Medalla M, Lera P, Feinberg M, Barbas H. Specificity in inhibitory systems associated with prefrontal pathways to temporal cortex in primates. *Cereb Cortex* 17, *Suppl* 1: i136–i150, 2007.

Miller EK, Cohen JD. An integrative theory of prefrontal cortex function. *Annu Rev Neurosci* 24: 167–202, 2001.

Miller P, Katz DB. Stochastic transitions between neural states in taste processing and decision-making. *J Neurosci* 30: 2559–2570, 2010.

Miller P, Katz DB. Accuracy and response-time distributions for decision-making: linear perfect integrators versus nonlinear attractor-based neural circuits. *J Comput Neurosci* 35: 261–294, 2013.

Niwa M, Ditterich J. Perceptual decisions between multiple directions of visual motion. *J Neurosci* 28: 4435–4445, 2008.

Niyogi RK, Wong-Lin K. Dynamic excitatory and inhibitory gain modulation can produce flexible, robust and optimal decision-making. *PLoS Comput Biol* 9: e1003099, 2013.

Nocedal J, Wright S. *Numerical Optimization* (2nd Ed.). New York: Springer, 2006.

Palmer J, Huk AC, Shadlen MN. The effect of stimulus strength on the speed and accuracy of a perceptual decision. *J Vis* 5: 376–404, 2005.

Pouget P, Logan GD, Palmeri TJ, Boucher L, Paré M, Schall JD. Neural basis of adaptive response time adjustment during saccade countermanding. *J Neurosci* 31: 12604–12612, 2011.

Purcell BA, Schall JD, Logan GD, Palmeri TJ. From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. *J Neurosci* 32: 3433–3446, 2012.

Rae B, Heathcote A, Donkin C, Averell L, Brown S. The hare and the tortoise: emphasizing speed can change the evidence used to make decisions. *J Exp Psychol Learn Mem Cogn* 40: 1226–1243, 2014.

Ratcliff R. A theory of memory retrieval. *Psychol Rev* 85: 59–108, 1978.

Ratcliff R. Methods for dealing with reaction time outliers. *Psychol Bull* 114: 510–532, 1993.

Ratcliff R, Rouder JN. Modeling response times for two-choice decisions. *Psychol Sci* 9: 347–356, 1998.

Ridderinkhof KR. The role of the medial frontal cortex in cognitive control. *Science* 306: 443–447, 2004.

Roitman JD, Shadlen MN. Response of neurons in the lateral intraparietal area during a combined visual discrimination reaction time task. *J Neurosci* 22: 9475–9489, 2002.

Roxin A, Ledberg A. Neurobiological models of two-choice decision making can be reduced to a one-dimensional nonlinear diffusion equation. *PLoS Comput Biol* 4: e1000046, 2008.

Salinas E. Gain modulation: a major computational principle of the central nervous system. *Neuron* 27: 15–21, 2000.

Shadlen MN, Newsome WT. Neural basis of a perceptual decision in the parietal cortex (area LIP) of the rhesus monkey. *J Neurophysiol* 86: 1916–1936, 2001.

Shea-Brown E, Gilzenrat MS, Cohen JD. Optimization of decision making in multilayer networks: the role of locus coeruleus. *Neural Comput* 20: 2863–2894, 2008.

Shu Y, Hasenstaub A, McCormick DA. Turning on and off recurrent balanced cortical activity. *Nature* 423: 288–293, 2003.

Sigman M, Dehaene S. Parsing a cognitive task: a characterization of the mind's bottleneck. *PLoS Biol* 3: e37, 2005.

Soltani A, Wang XJ. A biophysically based neural model of matching law behavior: melioration by stochastic synapses. *J Neurosci* 26: 3731–3744, 2006.

Standage D, You H, Wang DH, Dorris MC. Gain modulation by an urgency signal controls the speed-accuracy trade-off in a network model of a cortical decision circuit. *Front Comput Neurosci* 5: 7, 2011.

Standage D, You H, Wang DH, Dorris MC. Trading speed and accuracy by coding time: a coupled-circuit cortical model. *PLoS Comput Biol* 9: e1003021, 2013.

Tomita H, Ohbayashi M, Nakahara K, Hasegawa I, Miyashita Y. Top-down signal from prefrontal cortex in executive control of memory retrieval. *Nature* 401: 699–703, 1999.

Usher M, McClelland JL. The time course of perceptual choice: the leaky, competing accumulator model. *Psychol Rev* 108: 550–592, 2001.

Vogels TP, Abbott LF. Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nat Neurosci* 12: 483–491, 2009.

Wang CT, Lee CT, Wang XJ, Lo CC. Top-down modulation on perceptual decision with balanced inhibition through feedforward and feedback inhibitory neurons. *PLoS One* 8: e62379, 2013.

Wang XJ. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* 36: 955–968, 2002.

Wang XJ. Decision making in recurrent neuronal circuits. *Neuron* 60: 215–234, 2008.

Wickelgren WA. Speed-accuracy tradeoff and information processing dynamics. *Acta Psychol (Amst)* 41: 67–85, 1977.

Wong KF, Huk AC, Shadlen MN, Wang XJ. Neural circuit dynamics underlying accumulation of time-varying evidence during perceptual decision making. *Front Comput Neurosci* 1: 6, 2007.

Wong KF, Wang XJ. A recurrent network mechanism of time integration in perceptual decisions. *J Neurosci* 26: 1314–1328, 2006.