

*Please note that this draft may not be identical with the published version.*

Logan, G. D., Yamaguchi, M., Schall, J. D., & Palmeri, T. J. (2015). Inhibitory control in mind and brain 2.0: Blocked-input models of saccadic countermanding. *Psychological Review*, 122, 115-147. doi: 10.1037/a0038893

Inhibitory control in mind and brain 2.0:  
Blocked-input models of saccadic countermanding

Gordon D. Logan<sup>1</sup>, Motonori Yamaguchi<sup>1,2</sup>, Jeffrey D. Schall<sup>1</sup>, Thomas J. Palmeri<sup>1</sup>

<sup>1</sup>Vanderbilt University, <sup>2</sup>Edge Hill University

Running Head: Blocked-input models of saccadic countermanding

Address correspondence to:

Gordon D. Logan,

Department of Psychology,

Vanderbilt University,

Nashville TN 37203

Phone: 615-322-2529

Fax: 615-343-8449

E-mail: [gordon.logan@vanderbilt.edu](mailto:gordon.logan@vanderbilt.edu)

In press, *Psychological Review*

**Abstract**

The interactive race model of saccadic countermanding assumes that response inhibition results from an interaction between a go unit, identified with gaze-shifting neurons, and a stop unit, identified with gaze-holding neurons, in which activation of the stop unit inhibits the growth of activation in the go unit to prevent it from reaching threshold. The interactive race model accounts for behavioral data and predicts physiological data in monkeys performing the stop-signal task. We propose an alternative model that assumes that response inhibition results from blocking the input to the go unit. We show that the blocked-input model accounts for behavioral data as accurately as the original interactive race model and predicts aspects of the physiological data more accurately. We extend the models to address the steady-state fixation period before the go stimulus is presented and find that the blocked-input model fits better than the interactive race model. We consider a model in which fixation activity is boosted when a stop signal occurs and find that it fits as well as the blocked input model but predicts very high steady-state fixation activity after the response is inhibited. We discuss the alternative linking propositions that connect computational models to neural mechanisms, the lessons to be learned from model mimicry, and generalization from countermanding saccades to countermanding other kinds of responses.

The ability to inhibit thought and action is an important component of cognitive control. It improves over childhood and declines in old age. It is strong in healthy adults and weak in people with psychiatric and neurological disorders. It varies between individuals with different personalities and cognitive abilities. It is often studied in the *stop-signal paradigm*, in which people are asked to inhibit a response they are about to execute (for reviews, see Logan, 1994; Verbruggen & Logan, 2008). The inhibitory process in the stop-signal paradigm is not directly observable, so it must be assessed by applying a mathematical model to the data. For 25 years stop-signal behavior was explained in terms of Logan and Cowan's (1984) *independent race model*, which assumes that stop-signal performance depends on the outcome of a race between a go process that produces an overt response and a stop process that inhibits it. The independent race model provides estimates of the latency of the unobservable response to the stop signal (*stop-signal response time* or *SSRT*), which is the primary measure of inhibitory control in stop-signal studies of development, aging, psychopathology, and neuropathology (also see Logan, Van Zandt, Verbruggen & Wagenmakers, 2014). The independent race model addresses whether and when a response is inhibited but does not address how the response is inhibited. It describes the processes that run in the race; it does not describe what happens at the end of the race when the stop process wins. Recently, Boucher, Palmeri, Logan and Schall (2007) proposed an *interactive race model* that describes what happens when the stop process wins: They assumed that a stop unit inhibits the growth of activation in a go unit to prevent it from reaching a threshold that triggers the response (also see Lo, Boucher, Paré, Schall & Wang, 2009; Ramakrishnan, Sureshbabu & Murthy, 2012; Wong-Lin, Eckhoff, Holmes & Cohen, 2010; cf. Salinas & Stanford, 2013). Boucher et al. showed that the interactive race model accounts for behavior as well as the independent race model and goes beyond it to predict important properties of the underlying neurophysiology.

The purpose of this article is to evaluate alternatives to the interactive race model that provide different explanations of how responses are stopped. We focus on *blocked-input* models that stop responses by blocking the input to the go unit instead of inhibiting the growth of activation in the go unit (Band & van Boxtel, 1999; Logan, 1983; Logan & Cowan, 1984). Blocked-input models assume that actions can be stopped by disabling the

mapping between the perceptual system and the motor system (Logan, 1983) or by deleting the goals that enable responding (Logan & Cowan, 1984). We show that the blocked-input model fits behavioral data as well as the interactive race model and predicts aspects of neurophysiological data more accurately.

We extend both models to account for steady-state activity in the fixation period before the trial begins to provide stronger constraints on the model parameters. Both of the new models predict neurophysiological data more accurately, and the blocked-input model fits the behavioral data better than the interactive race model. We extend the interactive race model by adding a top-down boost in fixation activity, and find that this *boosted fixation* model fits behavior as well as the blocked input model but predicts very high steady-state fixation activity after successful inhibition. Our results call for a revision of the *linking propositions* Boucher et al. (2007) proposed to connect the interactive race model to a network of mutually-inhibitory gaze-holding and gaze-shifting neurons in frontal eye fields and superior colliculus, which identified the stop process exclusively with inhibition from gaze-holding neurons. The blocked-input and boosted fixation models provide new perspectives on the neural instantiation of the stop processes, showing that a stop process located outside the gaze control network tips the balance between gaze-holding and gaze-shifting neurons to stop a response.

### **The Stop-Signal Paradigm**

The stop-signal paradigm engages subjects in a *go task* whose purpose is to produce overt responses to a *go stimulus* on a majority of trials. The go stimulus is usually visual and the go response is usually manual, but the same patterns of behavior are observed with other go stimuli and other go responses, including saccadic eye movements (e.g., Hanes & Schall, 1995; Logan & Irwin, 2000), wrist and arm movements (e.g., Brunamonti, Ferraina & Paré, 2012; Scangos & Stuphorn, 2010), head movements (Corneil & Elsley 2005), and speech (e.g., Xue, Aron & Poldrack, 2008). Occasionally, a *stop signal* is presented, which instructs subjects to inhibit the go response on that trial. The stop signal is usually auditory, but the same patterns of results are observed with visual (e.g., Lappin & Eriksen, 1966) and tactile stop signals (e.g., Akerfelt, Colonius & Diederich, 2006). There are two possible outcomes on stop-signal trials: Subjects may inhibit the go response, producing

what is called a *signal-inhibit* or *canceled* trial, or they may fail to inhibit the go response, producing what is called a *signal-respond* or *non-canceled* trial with a response time (RT) that is called *signal-respond RT*. The probability of inhibiting the go response given a stop signal ( $p(\text{inhibit})$ ) or the complementary probability of responding ( $p(\text{respond}|\text{signal})$ ) and signal-respond RT are the major dependent variables. The most important independent variable is the delay between the onset of the go stimulus and the onset of the stop signal, which is called *stop-signal delay* (SSD). When SSD is short, subjects often succeed at inhibiting; when SSD is long, subjects often fail to inhibit. The plot of  $p(\text{inhibit})$  or  $p(\text{respond}|\text{signal})$  against stop-signal delay is called the *inhibition function*, and a major goal of models of the stop-signal paradigm is to explain the shape of the inhibition function.

Stop-signal delay also affects signal-respond RT. Signal-respond RT is short when SSD is short and long when SSD is long. It is typically shorter than RT on trials on which no stop signal is presented (called *no-stop-signal* trials), approaching no-stop-signal RT as SSD increases. Cumulative distributions of signal-respond RTs typically begin at the same point as cumulative distributions of no-stop-signal RTs but reach asymptote at progressively earlier times the shorter the SSD. Another major goal of models of the stop-signal paradigm is to explain these relations between signal-respond RT and no-stop-signal RT and the way they change with SSD.

Figure 1 presents behavioral data from a seminal stop signal experiment by Hanes, Patterson, and Schall (1998), in which behavioral and neural data were collected on the same trials from monkeys performing a saccadic stop signal task. Boucher et al. (2007) fit models to these data, and so do we. On each trial, monkeys were presented with a central fixation point. At a variable time after they fixed their eyes on this point, it extinguished and was replaced by a peripheral target, which was a point of the same size that fell in the receptive field of the cell they were recording from or on the opposite side of the display. Monkeys were rewarded for executing a saccade to the target as quickly as possible. On a minority of trials, the fixation point was re-illuminated as a stop signal, and monkeys were rewarded for inhibiting their saccades on those trials. The monkeys performed this task well, producing behavioral data that were typical of the stop-signal paradigm. Their inhibition functions, plotted in Figure 1, showed the characteristic increase in  $p(\text{respond}|\text{signal})$  with SSD. Their signal-respond RTs were faster than their no-stop-

signal RTs and the distributions, also plotted in Figure 1, were ordered appropriately according to SSD.

### The Independent Race Model

Logan and Cowan (1984) developed the independent race model to account for the major results in the stop-signal paradigm and to provide methods for estimating SSRT. The independent race model assumes that a go process races with a stop process on stop-signal trials and performance is determined by the winner of the race. The stop process wins the race if  $SSRT + SSD < go\ RT$ ; the response is inhibited and a signal-inhibit trial occurs. The go process wins the race if  $go\ RT < SSRT + SSD$ ; the response is executed and a signal-respond trial occurs. Go RT and SSRT are assumed to be independent random variables, so the outcome of the race is stochastic. Increasing SSD handicaps the race in favor of the go process, so the stop process wins less often. Thus,  $p(\text{inhibit})$  decreases and  $p(\text{respond}|\text{signal})$  increases as SSD increases, producing the inhibition function. The stochastic nature of the race also explains the relation between signal-respond RT and no-stop-signal RT and how it changes with SSD. When SSD is short, only the fastest go RTs are faster than  $SSRT + SSD$ , so signal-respond RT is very short, reflecting the lower tail of the go-RT distribution. As SSD increases, more go RTs are fast enough to win the race, so signal-respond RT increases. If SSD is long enough, all go RTs will win the race and signal-respond RT will approach no-stop-signal RT.

The independent race model provides several related methods for estimating the unobserved SSRT from the observed  $p(\text{respond}|\text{signal})$ , the observed signal-respond RT distribution, and the observed no-stop-signal go RT distribution (see Logan, 1994; Logan & Cowan, 1984). The *integration method* is commonly used and intuitively clear. It assumes that SSRT is a constant, which is implausible, but it produces accurate estimates of SSRT if the stop process and go process are independent (see Band, van der Molen & Logan, 2003; Colonius, 1990; De Jong, Coles, Logan & Gratton, 1990; Logan & Cowan, 1984; Verbruggen, Chambers & Logan, 2013). If SSRT is constant, then the finishing time of the stop process (i.e.,  $SSRT + SSD$ ) is a single point in time. All go RTs that occur after that point will be inhibited; all go RTs that occur before it will not be inhibited. That point is not directly observable but it can be estimated from the observed distribution of go RTs on no-stop-

signal trials and the observed  $p(\text{respond}|\text{signal})$  at a particular SSD. The probability of responding given a stop signal at a particular SSD equals the percentile of the no-stop-signal go RT distribution at which go RT equals SSRT + SSD, so SSRT can be estimated by subtracting SSD from the no-stop-signal go RT at that percentile. The independent race model provides other methods for estimating SSRT that make the more realistic assumption that SSRT is a random variable (Colonius, 1990; De Jong et al., 1990; Logan & Cowan, 1984; Logan et al., 2014; Matzke, Dolan, Logan, Brown & Wagenmakers, 2013), which usually produce estimates that are similar to estimates from the integration method (but see Verbruggen et al., 2013). These measures of SSRT form the basis of conclusions about changes in inhibitory control with individual differences, development, aging, psychopathology, and neuropathology (see Logan, 1994; Verbruggen & Logan, 2008).

### Interactive Race Model 1.0

The independent race model addresses only the finishing times of the stop and go processes. It does not consider the processes that give rise to them or the process that causes the go response to stop. In the years since the model was developed, tremendous strides have been made in understanding the computational (Logan et al., 2014; Ratcliff & Smith, 2004; Usher & McClelland, 2001) and neural processes (e.g., Hanes et al., 1998; Paré & Hanes, 2003) that underlie the stop and go processes. Boucher et al. (2007) took advantage of these advances and modeled the go process as a stochastic accumulator that integrates activation over time until it reaches a threshold, whereupon a response is executed. They modeled the stop process a stochastic accumulator that stops responses by inhibiting activation in the go accumulator to prevent it from reaching threshold. The architecture of the model, describing the units and the connections between them is presented in Figure 2. A timing diagram, describing the relation between parameter values and trial events, is also presented in Figure 2.

Go and stop units accumulate activation over time according to the following stochastic differential equations (Usher & McClelland, 2001), which specify the change ( $da_{go}$  and  $da_{stop}$ ) in each unit in each time step  $dt$  ( $dt/\tau$  was set to 1.0):

$$da_{go} = \frac{dt}{\tau} [\mu_{go} - k_{go} \cdot a_{go}(t) - \beta_{stop} \cdot a_{stop}(t)] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{go} \quad (1)$$

$$da_{stop} = \frac{dt}{\tau} [\mu_{stop} - k_{stop} \cdot a_{stop}(t) - \beta_{go} \cdot a_{go}(t)] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{stop} \quad (2)$$

where  $a_{go}(t)$  and  $a_{stop}(t)$  are the activation values for the go unit and the stop unit, respectively. The model has eleven parameters:  $\theta_{go}$ , which represents the threshold on the go process,  $D_{go}$  and  $D_{stop}$ , which represent afferent processing time (i.e., non-decision time),  $\mu_{go}$  and  $\mu_{stop}$ , which represent the mean growth in activation,  $k_{go}$  and  $k_{stop}$ , which are leakage terms representing the loss of activation over time,  $\beta_{go}$  and  $\beta_{stop}$ , which reflect the inhibition of the stop process from the go process and the go process from the stop process, and  $\xi_{go}$  and  $\xi_{stop}$ , which are Gaussian noise terms with means of zero and standard deviations of  $\sigma_{go}$  and  $\sigma_{stop}$ , respectively. The leakage parameters  $k_{go}$  and  $k_{stop}$  were set to 0 to reduce the number of parameters because nonzero values did not improve the goodness of fit (Boucher et al., 2007). A go response occurs when activation in the go unit reaches a threshold  $\theta_{go}$ .

The timing diagram in Figure 2 describes the relation between parameter values and trial events. At the beginning of a trial,  $\mu_{go}$  and  $\mu_{stop}$  are set to 0. After a latent period called  $D_{go}$  that represents afferent processing time (i.e., non-decision time),  $\mu_{go}$  becomes greater than 0, so activation begins to accumulate toward the threshold. On stop signal trials,  $\mu_{stop}$  becomes greater than 0 after SSD and a subsequent latent period called  $D_{stop}$ , whereupon the stop unit begins to inhibit the go unit. The response is inhibited if the stop unit becomes active soon enough to prevent the go unit from reaching threshold. Inhibition fails when the stop unit comes on too late (e.g., when SSD is long) or when it is not activated enough to prevent the go unit from reaching threshold.

Boucher et al. (2007) proposed *linking propositions* (Schall, 2004; Teller, 1984) that map the components of the computational model to the underlying neural circuitry, taking advantage of recent investigations of stopping behavior and neurophysiology in monkeys who performed a saccadic stop-signal task (Hanes et al., 1998; Paré & Hanes, 2003). Saccades are generated by a network of mutually inhibitory gaze-holding and gaze-shifting neurons that extends from cortex through basal ganglia, superior colliculus, and thalamus,

to the brainstem. Boucher et al. identified the go unit with gaze-shifting neurons, the stop unit with gaze-holding neurons and the inhibitory connections from gaze-holding to gaze-shifting neurons.

The linking proposition that connects stochastic accumulation in the go unit to movement-related neurons in frontal eye fields and superior colliculus was first expressed by Hanes and Schall (1996) and has become well established: Activity in movement-related neurons rises to a threshold that is constant within conditions and the rate of growth varies randomly within conditions and systematically between conditions, as in stochastic accumulator models fitted to the data (Ding & Gold, 2012; Gold & Shadlen, 2007; Hanes & Schall, 1996; Pouget, Logan, Palmeri, Boucher, Paré & Schall, 2011; Purcell et al., 2010, 2012; Ratcliff, Cherain & Segraves, 2003; Ratcliff, Hasegawa, Hasegawa, Smith & Segraves, 2007; Shadlen & Kiani, 2013). The linking proposition that connects the stop unit to fixation neurons and their inhibitory connections to movement neurons is less well established. A major purpose of this article is to evaluate that linking proposition.

### **Fitting Interactive Race Model 1.0 to Behavioral Data**

Boucher et al. (2007) used Equations 1 and 2 to fit simulations of interactive race 1.0 to the behavioral data from the two monkeys reported by Hanes et al. (1998). The fitting routine minimized Pearson's chi square between observed and predicted frequencies at each stop signal delay for inhibition functions and at each quintile of each signal-respond and no-stop signal RT distribution. In fitting the model, Boucher et al. fixed the threshold ( $\theta_{go} = 1,000$ ), and set  $D_{go}$  to the afferent delay time measured from the activity profile of movement-related neurons (35 ms for monkey C and 80 ms for monkey A). Boucher et al. set  $k = 0$  because they found that variation in the amount of leakage did not affect the goodness of fit. We generated our own fits of interactive race 1.0, following the procedure described in Appendix A, and explored several variations of the model. We compared Boucher et al.'s (2007) assumption that  $\theta_{go}$  was constant and  $\sigma_{go}$  was a free parameter with a more common assumption that  $\theta_{go}$  was a free parameter and  $\sigma_{go}$  was fixed at 1.0 (e.g., Donkin, Heathcote & Brown, 2009; Ratcliff & Smith, 2004). We compared their assumption that  $D_{stop}$  was a constant with variants in which  $D_{stop}$  was stochastic to

allow variability in SSRT. We compared their assumption  $a_{stop}(t)$  grows over time with variants in which  $a_{stop}$  is constant. All variants of the model produced equivalent fits to the behavioral data and equivalent predictions of the physiological data (see Appendix B). Consequently, we chose a variant of the interactive race model in which  $D_{stop}$  had no stochastic component, threshold was a free parameter and  $\sigma_{go}$  was fixed to follow common practice in fitting stochastic accumulator models to data (Ratcliff & Smith, 2004) and to compare with the fits of the blocked input model and other models we consider later in the article.

The fits of interactive race 1.0 with  $\sigma_{go}$  fixed and  $\theta_{go}$  free are shown in Figure 3 (top row). The best-fitting values of the free parameters are presented in Table 1. As Boucher et al. (2007) found,  $\beta_{stop}$  was much larger than  $\beta_{go}$ , suggesting that inhibition from the stop unit to the go unit is more potent than inhibition from the go unit to the stop unit. As Boucher et al. also found,  $D_{stop}$  was quite large and occupied most of SSRT. This, and the large value of  $\beta_{stop}$ , led Boucher et al. to conclude that inhibition from the stop unit to the go unit must be *late and potent*. The stop process is independent of the go process for most of its duration (i.e., for  $D_{stop}$  ms) but then interacts briefly and potently with the go unit to reverse the accumulation of go-unit activation toward the threshold.

### **Predicting Neural Data with Interactive Race Model 1.0**

Interactive race 1.0 uses the dynamics of activation over time to account for behavioral measures of stopping and going. Boucher et al. (2007) argued that the dynamics of the model were predictions of the dynamics of movement neurons whose activity was recorded while the behavioral data were collected. If so, then the model dynamics should match the neural dynamics in important respects. Figure 4 shows the dynamics predicted from fixed parameter values that were estimated separately in behavioral fits. The trends in the simulated activation functions are qualitatively similar to the neural data (see Figure 5). The predicted activation rises from baseline to threshold, modulates around SSRT, and decays after it modulates. Here, we ask whether the predicted growth rate, modulation time, and decay rate match the neural data

quantitatively. As Boucher et al. (2007) correctly emphasized, these predictions are true predictions, in that they are calculated from the parameters that fit the behavioral data best without any further adjustment.

**The neural data.** Boucher et al. (2007) addressed the neurophysiological data that Hanes et al. (1998) collected on the same trials as the behavioral data. The mean normalized firing rates for movement-related neurons on signal-inhibit trials and latency-matched no-stop-signal trials are plotted for different SSDs and monkeys in Figure 5. Each panel shows modulation of firing rate on signal-inhibit trials. The signal-inhibit and no-stop-signal functions follow each other to a point and then diverge. The no-stop-signal function continues to grow but the signal-inhibit function decays to baseline. These results were replicated with monkeys in superior colliculus (Paré & Hanes, 2003), frontal eye fields (Murthy, Ray, Shorter, Schall, & Thompson, 2009), and dorsal premotor cortex (Mirabella, Pani, & Ferraina, 2011) and with rats in basal ganglia (Schmidt, Leventhal, Mallet, Chen, & Berke, 2013). We use the models to predict the rate of growth before the modulation, the time at which the modulation occurs relative to SSRT, and the rate of decay after the modulation. We focus primarily on signal-inhibit trials because they include the crucial neural data—growth, modulation, and decay. We do not model the modulation and decay of neural activity on no-stop-signal trials because they reflect the correlates and consequences of moving the eyes, which are not relevant to stopping an eye movement before it begins. There are no eye movements on signal-inhibit trials, so the effects are due to the mechanism that stops the eye movements.

The activation functions for monkey C look different from the activation functions from monkey A in the initial rise from baseline and in the return to baseline after modulation on signal-inhibit trials. The neurons from monkey C were mostly *visuomovement neurons*, showing a visual response and a movement response, whereas the neurons from monkey A were mostly *movement neurons* (Bruce & Goldberg, 1985; Schall, 1991). This classification is based on a delayed saccade task, in which monkeys must maintain fixation while a peripheral target is flashed and extinguished, and then move their eyes to the remembered location when a go signal is presented. Visual neurons are active during the initial flash but not the movement, movement neurons are active during the movement but not during the initial flash, and visuomovement neurons are active during both. Hanes et al. (1998) analyzed both

movement and visuomovement neurons, and found that both modulated their activity on signal-inhibit trials. Movement neurons provide the strongest test of the model because there is no visual response to contaminate measures of growth rate, modulation time, and decay rate (see Monkey A in Figure 5). Visuomovement neurons provide a weaker test. The visual response occurs early and persists after the response is inhibited, and that obscures measures of growth and decay rates (see Monkey C in Figure 5). However, visuomovement neurons modulate at about the same time as movement neurons (Ray, Pouget & Schall, 2009). Thus, both cell types can be used to test model predictions about modulation time, using the methods of Boucher et al. (2007).

**Predicted growth rate.** The rate of growth to threshold is an essential component of the linking proposition that identifies movement neurons with stochastic accumulator models. In neurons, the rate of growth varies randomly within conditions and systematically between conditions, just as drift rate varies in stochastic accumulator models fitted to the data (Gold & Shadlen, 2007; Hanes & Schall, 1996; Pouget et al., 2011; Purcell et al., 2010, 2012; Ratcliff et al., 2003; Ratcliff et al., 2007; Shadlen & Kiani, 2013). Our models of saccadic countermanding assume there is one accumulator for the go response, and its drift rate is the same on no-stop-signal and stop-signal trials (this is the context independence assumption that all race models make; Colonius, 1990; Logan et al., 2014). From Equation 1, we see that the rate of growth in the go unit on signal-inhibit trials before the point of modulation (i.e., for values of  $t \leq \text{SSD} + D_{\text{stop}}$ ) is  $\mu_{go} - k_{go} \cdot a_{go}(t)$ . The stop process has not engaged yet, so  $\beta_{\text{stop}} \cdot a_{\text{stop}}(t)$  drops out of the equation. Equation 1 predicts that the rate of growth for no-stop-signal trials should be the same as on signal-inhibit trials, because the stop unit is never engaged (i.e., SSD is essentially infinite, so  $t$  is always less than  $\text{SSD} + D_{\text{stop}}$ ).

We compared signal-inhibit growth rates predicted from interactive race 1.0 with growth rates observed in neural data recorded from movement and visuomovement neurons. We analyzed growth rate and decay rate on the same signal-inhibit trials. Growth rate before modulation reflects only the dynamics of the go unit; decay rate after modulation reflects the interaction of the stop and go units. We ran simulations of the model with the best-fitting parameters, and measured the rate of growth of simulated

activation (normalized) using the same methods we used to measure the rate of growth in neural firing rates (normalized; for details, see Appendix C). Observed and predicted values appear in Figure 6. Predicted growth rates fell within the 95% confidence intervals of the observed growth rates in both monkeys, obtained by bootstrapping. This supports the linking proposition that identifies movement neurons with the go unit (Boucher et al., 2007).

**Predicted decay rate.** Predicted and observed decay rates were measured from model simulations and neural data, using the methods we used to measure growth rates (see Appendix C). The predicted decay rate after the point of modulation is given by Equation 1 for values of  $t > SSD + D_{stop}$ . It includes continued growth of go activation from the go stimulus (i.e.,  $\mu_{go} - k_{go} \cdot a_{go}(t)$ ) and growing inhibition from the stop unit (i.e.,  $-\beta_{stop} \cdot a_{stop}(t)$ ) that overcomes go activation. The predicted and observed values, plotted in Figure 6, were quite different. Predicted decay rate fell outside the 95% confidence intervals of the observed decay rate for both monkeys, obtained by bootstrapping. The effect, quantified here, can be seen in the predicted activation functions in Figure 4, where the decay of activation in the go unit after modulation is almost instantaneous. The decay rate reflects the interaction between the stop unit and the go unit. The analyses of growth rates suggest that the model predicted the go unit well, so the strong decay must reflect the influence of the stop unit. The inhibition from the stop unit to the go unit ( $\beta_{stop}$ ) may be too strong. Simulations described in Appendix D suggest that  $\beta_{stop}$  can vary substantially without much effect on the goodness of fit to the behavioral data. This large misprediction was surprising and motivated our investigation of the next set of models, described below (see Models of Fixation and Movement Initiation).

**Predicted cancel times.** Boucher et al. (2007) used model activation functions to predict *cancel time*, which is the time at which go activation modulates on stop-signal trials relative to SSRT. Cancel time is important in neuroscience because it is an essential criterion for determining whether modulation of neural activity happens early enough to participate in response inhibition. Activity modulates on stop signal trials in many areas of the brain, but some modulation happens well after SSRT, so it must be a consequence and not a cause of response inhibition (Hanes et al., 1998). For example, all visual neurons in

FEF (Hanes et al., 1998) and all neurons in supplementary eye field (Stuphorn, Brown & Schall, 2010) and supplementary motor area (Scangos & Stuphorn, 2010) modulate after SSRT when monkeys countermand saccades, suggesting that these neurons are not part of the circuit that actually stops the response. Instead, they may participate in proactive control.

Following Hanes et al. (1998), we defined cancel time as the time at which go unit activation on signal-inhibit trials became significantly less than go unit activation on latency-matched no-stop-signal trials minus SSRT (see Appendix D for details). We generated predicted cancel times by simulating the model with the best-fitting parameters from the fits to the behavioral data.

The mean predicted and observed cancel times are plotted in Figure 7. There was much more variability in observed cancel times in monkey C than in monkey A, but the predicted cancel times fell within the bootstrapped 95% confidence intervals of the observed cancel times for both monkeys. Replicating Boucher et al. (2007), interactive race 1.0 predicts neural cancel times accurately.

### **Blocked-input Model 1.0**

Interactive race 1.0 assumes that the stop unit inhibits go unit activation. Here, we examine *blocked-input* models that assume that the stop unit does not inhibit go activation directly (Band & van Boxtel, 1999; Logan, 1983; Logan & Cowan, 1984; also see Awh, Matsukura, & Serences, 2003). Instead, the stop signal activates a top-down process that turns off go activation, setting  $\mu_{go}$  to 0, when it reaches its threshold, like turning off the ignition to stop a motor. If the input is blocked early enough, go unit activation will not rise to threshold and the response will be inhibited, producing a signal-inhibit trial. If go activation reaches threshold before its input is blocked, inhibition will fail and the go response will be executed, producing a signal-inhibit trial.

Blocked-input models are motivated by cognitive psychology, addressing the arbitrary nature of the voluntary responses that are involved in the stop-signal task and other cognitive paradigms. Humans and monkeys are not compelled to respond as they do by hard-wired reflex-like connections between stimuli and responses. Instead, cognitive

control processes establish temporary connections between stimuli and responses that address the demands of the current task but can be changed whenever task demands change (Logan & Gordon, 2001; Miller & Cohen, 2001). The temporary connections act as *prepared reflexes*, so responses can be triggered automatically when a relevant stimulus appears (Cohen-Kadosh & Meiran, 2009; Hommel, 2000; Logan, 1978), but they can be disabled on a moment's notice when the stimulus is no longer relevant. Blocked-input models assume that temporary connections can be disabled quickly in response to stop signals, disconnecting the input that drives the responses (Band & van Boxtel, 1999; Logan, 1983; Logan & Cowan, 1984).

Indeed, there is evidence that the monkeys whose data we analyzed used temporary connections between stimuli and responses, like prepared reflexes. They performed the stop task as if visually responsive neurons drove movement neurons reflexively (Purcell et al., 2010, 2012), yet visual neurons do not always drive movement neurons. During each session with the stop task, each monkey performed a *memory-guided saccade task* to classify the neurons whose activity was recorded. In this task, monkeys fixate a central point and a peripheral target is flashed. The monkeys must maintain fixation on the central point after the target is extinguished, and then move their eyes to the remembered location of the target when the central point is extinguished. Visual neurons increase their firing rate when the target appears but not when the eyes move; movement neurons increase their firing rate before the eyes move but not when the target appears; and visuomovement neurons respond at both times (Bruce & Goldberg, 1985; Schall, 1991). In the memory-guided saccade task, visual neuron activity does not drive movement neuron activity. Visual neurons respond to the onset of the target without activating movement neurons (some sustain a lower level of activity during the delay period), and movement neurons respond at the offset of the fixation point without a burst in visual neuron activity to drive them. Yet a few minutes later, after the neurons have been classified, monkeys switch to the countermanding task, and visual neurons now appear to drive movement neurons. Some process prevents visual neurons from driving movement neurons in the memory-guided saccade task and allows visual neurons to drive movement neurons in the countermanding task. The blocked-input model assumes that this process is recruited to countermand saccades.

Following Purcell et al. (2010, 2012), we might identify the stop process with a gate between visual neurons that select the target and movement neurons that generate a movement to it. The stop process may raise the gate, blocking input to the movement neurons and thereby preventing them from reaching threshold. This is consistent with the observation that visual neurons do not modulate before SSRT on signal-inhibit trials but movement neurons do. The gate prevents the continued activation in visual neurons from driving the movement neurons to threshold. The gate may also be raised during the memory-guided saccade task to prevent visual neurons from driving movement neurons; when signaled to make a saccade, the gate may be lowered to allow visual neurons with sustained activity to drive movement neurons.

The architecture of blocked input 1.0 is presented in Figure 8. The model assumes there is no direct inhibition between the stop and go units (i.e.,  $\beta_{go} = \beta_{stop} = 0$ ), so the go unit is governed only by its drive ( $\mu_{go}$ ) and leakage. Equation 1 becomes

$$da_{go} = \frac{dt}{\tau} [\mu_{go} - k_{go} \cdot a_{go}(t)] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{go} \quad (3)$$

The afferent stage of the stop process is governed by the following stochastic differential equation:

$$da_{stop} = \frac{dt}{\tau} [\mu_{stop} - k_{stop} \cdot a_{stop}(t)] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{stop} \quad (4)$$

When the afferent stage is finished,  $\mu_{go}$  is set to 0.

The timing diagram, relating model events to stimulus events, is also presented in Figure 8. At the beginning of each trial,  $\mu_{go} = \mu_{stop} = 0$ . After an afferent processing time of  $D_{go}$  ms,  $\mu_{go}$  becomes greater than 0, and go activation begins to accumulate toward the threshold  $\theta_{go}$ . On stop signal trials,  $\mu_{stop}$  becomes greater than 0 at SSD plus the nondecision time for the stop process ( $ND_{stop}$ ), and stop activation begins to accumulate toward the threshold  $\theta_{stop}$ . When stop activation reaches threshold, the stop process blocks the input to the go process, setting  $\mu_{go} = 0$ , which stops or reverses the growth of go activation toward threshold. If  $k_{go} = 0$  (i.e., if there is no leakage), go activation will hover around the level it reached before  $\mu_{go}$  was set to 0 as noise adds and subtracts small amounts of activation. If  $k_{go} > 0$ , then go activation will decay. The greater the value of  $k_{go}$ ,

the more rapidly activation is lost. In either case, go activation will no longer grow toward threshold and the response will be inhibited unless the noise is substantial.

### **Fitting Blocked Input Model 1.0 to Behavioral Data**

We fit two versions of blocked-input 1.0 to the behavioral data. Both assumed  $\sigma_{go} = \sigma_{stop} = 1.0$  and both allowed  $\theta_{go}$  and  $\theta_{stop}$  to vary freely. One assumed a constant afferent component of the stop process and the other assumed a stochastic afferent component. Both blocked-input models fit the data as well as any of the interactive race models we considered (see Appendix B, Table B1). The values of the best-fitting parameters and measures of goodness of fit for the model that assumed a stochastic afferent stop unit appear in Table 1. In this model, the afferent stage of the stop process has two components, a non-stochastic phase representing non-decision time (characterized by the constant  $ND_{stop}$ ) and a stochastic phase representing growth of activation to a threshold (characterized by  $\mu_{stop}$  and  $\theta_{stop}$ ). The predicted inhibition functions and distributions of signal-respond and no-signal RTs from this model are presented in Figure 3 (bottom row). Blocked input 1.0 fit the data as well as interactive race 1.0. The  $\chi^2$  values were numerically smaller but very similar, within 3 units. Thus, we conclude that the fits were equivalent.

The fact that blocked input 1.0 fit as well as interactive race 1.0 is important because it indicates that direct inhibition of the stop process on the go process is not necessary to account for stop-signal behavior. Models that assume no direct inhibition from the stop process account for the data just as well. This calls for a re-evaluation of the linking proposition in the original interactive race model that identifies the stop process with fixation neurons and their inhibitory connections with movement neurons. Other linking propositions may provide equally satisfactory accounts of the data (also see Schall, 2004).

Boucher et al. (2007, p. 392) fit a blocked-input model with noise variable and threshold constant at 1,000 and found that it did not fit as well as the interactive race model. Their poor fit may have stemmed from the regions of the parameter space that were sampled. The starting values we used may have been better tailored to the data. As described in Appendix A, unlike Boucher et al., we first fit the no-stop-signal RTs; then we

fixed those parameters and fit the stop-signal data; finally, we used the parameters from the first two fits to generate a range of starting values for fits in which all parameters were free to vary. This procedure may have led us to a better-fitting region of the parameter space. The important point is that our current fits of blocked input 1.0 were as good as our fits of interactive race 1.0.

### Predicting Neural Data with Blocked-Input Model 1.0

We generated predicted activation functions for blocked input 1.0 for both monkeys by simulating the model with the best-fitting parameters. The predicted activation functions, shown above in Figure 4, show the same qualitative effects as the neural activation functions: growth to a threshold, modulation around SSRT, and decay. Here, we evaluate the predictions quantitatively.

**Growth rate.** From Equation 3, the predicted growth rate before the point of modulation (i.e., for  $t \leq SSD + ND_{stop} + \theta_{stop}/\mu_{stop}$ ) is  $\mu_{go} - k_{go} \cdot a_{go}(t)$ . This is exactly the same growth rate predicted by interactive race 1.0 (see above), except for differences in parameter values. Predicted growth rates for blocked input 1.0, plotted in Figure 6, fell within the 95% confidence intervals of the observed growth rates for both monkeys.

**Decay rate.** From Equation 3, the predicted decay rate after the point of modulation (i.e., for  $t > SSD + ND_{stop} + \theta_{stop}/\mu_{stop}$ ) is  $-k_{go} \cdot a_{go}(t)$ . After that point,  $\mu_{go} = 0$ , so  $\mu_{go}$  does not contribute to the decay rate. The predicted decay rate for interactive race 1.0 is more complex, including terms for continued growth of go activation (i.e.,  $\mu_{go} > 0$ ) and inhibition from the stop unit to the go unit (i.e.,  $-\beta_{stop} \cdot a_{stop}(t)$ ). Predicted decay rates for blocked input 1.0, also plotted in Figure 6, fell within the 95% confidence intervals of the observed growth rates for both monkeys.

**Cancel time.** Predicted cancel times are plotted in Figure 7. The predicted cancel times fell within the 95% confidence intervals of the observed cancel times for both monkeys.

### Interim Summary and Evaluation 1.0

Interactive race 1.0 and blocked input 1.0 fit the behavioral data equally well. This was surprising because Boucher et al. (2007) found that a blocked input model fit worse

than the interactive race model. We found equivalent fits across several model variants (see Appendix B), suggesting substantial mimicry. Following Boucher et al., we turned to neurophysiological data to break the mimicry. Interactive race 1.0 predicted cancel times and growth rates accurately but mispredicted decay rates for both monkeys. We attributed the misprediction to the strong inhibitory connections between stop and go units (i.e., large  $\beta_{stop}$ ). Blocked input 1.0 predicted growth rates, decay rates, and cancel times accurately. Thus, the neurophysiological data break the mimicry in favor of blocked input 1.0.

We sought constraints on inhibition and leakage parameters that might improve the accuracy of neurophysiological predictions without compromising goodness of fit to the behavioral data. We found useful constraints by extending the models to account for the interaction between gaze-shifting and gaze-holding neurons in the fixation period before the movement begins. We show below that steady-state activation in the fixation period depends on the ratio of growth rate to leakage (i.e.,  $\mu/k$ ), which constrains leakage. Inhibition from stop to go (i.e.,  $\beta_{stop}$ ) cannot be so large that it suppresses all growth in go activation.

### **Models of Fixation and Movement Initiation**

Boucher et al. (2007) developed the interactive race model to explain response inhibition in terms of the network of mutually inhibitory gaze-holding and gaze-shifting neurons that controls saccadic eye movements, which extends from cerebral cortex to the brainstem. They identified the stop unit exclusively with the activation of gaze-holding neurons and their inhibitory connections to gaze-shifting neurons, which prevent the activation of gaze-shifting neurons from reaching the threshold required to elicit an eye movement. A key assumption of their model was that response inhibition occurred within the gaze control network. Blocked input 1.0 assumes that response inhibition occurs outside the gaze control network: Responses are stopped by blocking the inputs that drive the activation of gaze-shifting neurons toward threshold, like stopping an engine by turning off the ignition. The fits to behavioral data showed substantial mimicry between interactive race 1.0 and blocked input 1.0. We sought to resolve the mimicry by modeling more of the dynamics of the network of gaze-shifting and gaze-holding neurons, extending

the model back in time to include the period in which the eyes are fixated at the beginning of a trial. If stop and go units tip the balance in the gaze control network, we should model the processes that achieve that balance. As we show below, modeling steady-state fixation constrains the interaction between gaze-holding and gaze-shifting units, and those constraints affect how the gaze control network can implement response inhibition.

We consider two models that explain response inhibition within the network and two models that explain response inhibition outside the network. The first within-network model is the *baseline model*, which assumes there is symmetrical inhibition between fixation and movement units ( $\beta_{stop} = \beta_{go}$ ) and stopping and going are driven by stimulus events (the onsets of the go signal and the stop signal) that tip the balance in the gaze control network. True to its title, this model serves as a baseline for evaluating the additional processes in the other models. The second within-network model is *interactive race 2.0*, in which the inhibition between fixation and movement units is asymmetrical ( $\beta_{stop} > \beta_{go}$ ). This model makes the same linking propositions as interactive race 1.0 (Boucher et al., 2007), identifying the stop process with fixation units in the network and identifying the mechanism of inhibition with strong asymmetrical inhibition from fixation units to movement units.

The first model that explains response inhibition outside the gaze control network is *blocked input 2.0*, in which a top-down stop process blocks the input to the movement units (setting  $\mu_{move}$  to 0) when it reaches threshold. The second model that explains response inhibition outside the gaze control network is *boosted fixation model 1.0*, in which a top-down stop process increases the activation of fixation unit beyond the level required to maintain fixation (multiplying  $\mu_{fix}$  by a constant  $b > 1$ ) when it reaches threshold.

For this set of models we define the units more neutrally as fixation (*fix*) and movement (*move*) units, linking them generally to gaze-holding and gaze-shifting neurons in the network extending from cerebral cortex to the brain stem and specifically to fixation and movement neurons in frontal eye fields from which the neural measures we analyze were taken. We do not presume that the fixation unit is the stop unit. That is an assumption and a linking proposition in the baseline model and interactive race 2.0, but not in blocked input 2.0 or boosted fixation 1.0. The model fits will determine the plausibility

of that assumption. In all four models, activation in movement and fixation units is described by the following stochastic differential equations:

$$da_{move} = \frac{dt}{\tau} [\mu_{move} - k_{move} \cdot a_{move}(t) - \beta_{fix} \cdot a_{fix}(t)] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{move} \quad (5)$$

$$da_{fix} = \frac{dt}{\tau} [\mu_{fix} - k_{fix} \cdot a_{fix}(t) - \beta_{move} \cdot a_{move}(t)] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{fix} \quad (6)$$

The move unit accumulates activation until it reaches a threshold,  $\theta_{go}$ , whereupon an eye movement is executed. If move-unit activation fails to reach threshold, the response is withheld.

All models address the steady-state activation in the fix and move units during the fixation period before the go stimulus appears. During the fixation period,  $\mu_{go} = 0$  and  $\mu_{fix} > 0$ . Consequently,  $a_{move}(t)$  will be driven to 0 and remain there because we assume activation cannot be negative. Fixation activity  $a_{fix}(t)$  will reach a maximum value that is determined by the ratio of  $\mu_{fix}$  to  $k_{fix}$ , which can be derived from Equation 6 by setting  $da_{fix} = 0$  and solving for  $a_{fix}(t)$ . During steady-state fixation  $\beta_{go} \cdot a_{go}(t) = 0$ , so

$$da_{fix} = \mu_{fix} - k_{fix} \cdot a_{fix}(t) = 0.$$

Solving for  $a_{fix}(t)$  yields

$$a_{fix}(t) = \frac{\mu_{fix}}{k_{fix}} \quad (7).$$

The maximum value of  $a_{fix}(t)$  is constant over time because the  $\mu_{fix}$  and  $k_{fix}$  parameters that determine it are constant over time. Thus, we can model steady-state fixation activity with no new parameters.

Modeling steady-state fixation activity imposes important constraints on the activation and inhibition in the move and fix units. Eye movements can only occur if activation in the move unit ( $\mu_{go}$ ) and inhibition from the move unit to the fix unit ( $\beta_{go}$ ) are large enough to overcome steady-state activation in the fix unit and if inhibition from the fix unit to the move unit ( $\beta_{stop}$ ) is not large enough to suppress move activation entirely.

We fit the four models to the data from the two monkeys using the procedure described in Appendix A. In all model fits,  $\sigma_{fix} = \sigma_{move} = 1.0$  (Donkin et al., 2009; Ratcliff & Smith, 2004). Trials started with go-stimulus onset. Move unit activation was set to 0 and

fix unit activation was set to the steady-state value,  $\mu_{fix}/k_{fix}$ . We performed two sets of fits: a *constrained* set, in which parameters were forced to be equal for the fix and move units (i.e.,  $\mu_{fix} = \mu_{move}$ ,  $k_{fix} = k_{move}$ , and  $D_{fix} = D_{move}$  for all models;  $\beta_{fix} = \beta_{move}$  for baseline, blocked input 2.0, and boosted fixation 1.0), and an *unconstrained* set, in which parameters were allowed to vary freely between fix and move units. In the constrained model fits, the baseline model is nested in the other three models, so we can test the significance of the effects of adding the additional mechanisms expressed in their parameters by evaluating differences in  $\chi^2$ . In the unconstrained fits, interactive race 2.0 is nested in blocked input 2.0 and boosted fixation 1.0, so we can test the significance of the effects of adding mechanisms outside the network of mutually-inhibitory units that control saccades. The unconstrained baseline model is the same as interactive race 2.0, so we only fit three unconstrained models.

### Baseline Model

The baseline model assumes that performance is driven entirely by the stimuli. The architecture is presented in Figure 9A and the timing diagram is presented in Figure 10A. In the model,  $\mu_{fix}$  is set to the best-fitting value when the fixation point is present, set to zero when the fixation point disappears and the target appears, and set to the best-fitting value again  $D_{fix}$  ms after the stop signal appears.  $\mu_{move}$  is set to zero during fixation and remains at zero for  $D_{go}$  ms after the target appears, whereupon it is set to the best-fitting value.

**Fits to behavior.** The predicted inhibition functions and RT distributions for the baseline model are plotted along with the observed values in Figure 11 for monkey C and Figure 12 for monkey A. The best-fitting parameters and measures of goodness of fit appear in Table 2. The baseline model did not fit the behavioral data very well. It mispredicted the inhibition functions and signal-respond RTs for both monkeys, suggesting that some factor other than stimulus presentation is necessary to account for response inhibition.

**Neural predictions.** The mean predicted activation functions for move and fix units appear in Figure 13. The predicted functions show growth, modulation, and decay that

resemble neural activation qualitatively. Predicted growth rates and decay rates fell within the 95% confidence intervals of the observed rates (see Figure 14). Predicted cancel times were longer than observed in monkey C and shorter than observed in monkey A (see Figure 15).

### Interactive Race Model 2.0

Interactive race model 2.0 is the same as the baseline model, except that it assumes asymmetrical inhibition between fixation and movement units, so  $\beta_{fix} > \beta_{move}$ , in keeping with the original interactive race model (Boucher et al., 2007). Its architecture is presented in Figure 9. Its timing diagram is presented in Figure 10.

**Fits to behavior.** The predicted inhibition functions and RT distributions from the unconstrained fits are plotted in Figure 11; the best-fitting parameters and measures of goodness of fit from the constrained and unconstrained fits appear in Tables 2 and 3, respectively. In the constrained fits, interactive race 2.0 fit better than the baseline model for both monkeys:  $\chi^2(1) = 291.68$  and  $143.73$  for monkeys C and A, respectively, both  $p$ 's  $< .001$ . However, it predicted longer signal-respond RTs than observed for both monkeys. The unconstrained fits, which allowed  $\mu_{move}$  and  $\mu_{fix}$  to differ and  $D_{move}$  and  $D_{fix}$  to differ, were better than the constrained fits,  $\chi^2(3) = 10.09$ ,  $p < .02$ , and  $54.48$ ,  $p < .001$  for monkeys C and A, respectively.

The constraint of accounting for steady-state fixation activity required the leakage parameter to be greater than 0 and required a smaller asymmetry in the inhibition between fix and move units. For monkeys C and A, respectively,  $\beta_{fix}$  was 2.2 and 3.5 times as large as  $\beta_{move}$  in the constrained fits and 2.4 and 4.3 times as large in the unconstrained fits. These ratios were much smaller than the corresponding ratios in Boucher et al.'s (2007) fits, where  $\beta_{fix}$  was 43 and 22 times as large as  $\beta_{move}$  for monkeys C and A, respectively, and much smaller than the ratios in our fits of interactive race 1.0, where  $\beta_{fix}$  was 100 and 379 times as large as  $\beta_{move}$  for monkeys C and A, respectively. When we used the  $\beta_{fix}$  and  $\beta_{move}$  parameters from those models in interactive race 2.0, we could not generate any go responses. The inhibition from the fix units was too strong to allow the move units to accumulate activation. The reduction in  $\beta_{fix}$  that was necessary to allow

interactive race 2.0 to produce responses resulted in poorer fits to the data (compare  $\chi^2$  values in Tables 2 and 3 with those in Table 1).

**Neural predictions.** The mean activation functions, which appear in Figure 13, show growth, modulation, and decay that resemble neural activation functions qualitatively. Predicted growth and decay rates (Figure 14) and cancel times (Figure 15) fell within the 95% confidence intervals of the observed values for both monkeys. Apparently, the constraints that arose in modeling the fixation period improved the accuracy of the neurophysiological predictions.

### Blocked Input Model 2.0

Blocked-input model 2.0 is the same as the baseline model except that it includes a top-down process outside the gaze control network that responds to the stop signal. Its architecture is presented in Figure 9 and its timing diagram is presented in Figure 10. The top-down process changes  $\mu_{move}$  to 0 at  $SSD + D_{control}$ . If the change occurs early enough, the response is inhibited. In the constrained fits,  $\mu_{move} = \mu_{fix}$ ,  $\beta_{move} = \beta_{fix}$ ,  $k_{move} = k_{fix}$ , and  $D_{move} = D_{fix}$ . In the unconstrained version, they were allowed to differ. For simplicity, both constrained and unconstrained fits assumed that  $D_{control}$  is a constant. Appendix B shows that stochastic and constant values of  $D_{control}$  produced equivalent fits of blocked input 1.0.

**Fits to behavior.** The predicted inhibition functions and RT distributions for the unconstrained version are plotted in Figure 11 for monkey C and Figure 12 for monkey A. The best-fitting parameters and measures of goodness of fit for constrained and unconstrained versions appear in Tables 2 and 3, respectively. Both versions fit well; both fit better than blocked input 1.0 considered earlier (cf. interactive race 2.0, which produced worse fits than interactive race 1.0). In the constrained fits, blocked input 2.0 fit better than the baseline model for both monkeys:  $\chi^2(1) = 314.37$  and  $217.76$  for monkeys C and A, respectively, both  $p$ 's  $< .001$ . The unconstrained fit was not significantly better than the constrained fit for monkey C,  $\chi^2(4) = 5.37$ ,  $p = .251$ , but it was for monkey A,  $\chi^2(4) = 31.58$ ,  $p < .001$ . More importantly, in the unconstrained fits, blocked-input 2.0 fit better than interactive race 2.0 for both monkeys:  $\chi^2(1) = 17.97$  and  $51.13$  for monkeys C and A,

respectively, both  $p$ 's  $< .001$ . Modeling fixation activity before the target appears to have broken the mimicry between the blocked-input model and the interactive race model.

**Neural predictions.** The mean activation functions, plotted in Figure 13, show growth, modulation, and decay, like the neurons. Predicted growth and decay rates (Figure 14) and predicted cancel times (Figure 15) fell within the 95% confidence intervals of observed values for both monkeys. Again, the constraint of modeling fixation activity appears to have improved the accuracy of neurophysiological predictions.

### Boosted Fixation Model 1.0

We considered a fourth model in which a top-down process outside the gaze control network boosts the activity in the fixation unit after the afferent stage of the stop process reached threshold. Its architecture is presented in Figure 9 and its timing diagram is presented in Figure 10. Boosted fixation 1.0 is the same as blocked input 2.0 except that the external, top-down process multiplies  $\mu_{fix}$  by a constant,  $b$  (for "boost"),  $D_{control}$  ms after the stop signal appears. In the constrained fits,  $\mu_{move} = \mu_{fix}$ ,  $\beta_{move} = \beta_{fix}$ ,  $k_{move} = k_{fix}$ , and  $D_{move} = D_{fix}$ ; in the unconstrained fits, they were allowed to differ. Both constrained and unconstrained fits assumed that  $D_{control}$  is a constant.

**Fits to behavior.** The predicted inhibition function and RT distributions for the unconstrained version of the model are plotted in Figure 11 for monkey C and Figure 12 for monkey A. The best-fitting parameter values and measures of goodness of fit are presented in Tables 2 and 3, respectively. In the constrained fits, boosted fixation 1.0 fit better than the baseline model for both monkeys:  $\chi^2(2) = 308.77$  and  $242.84$  for monkey C and A, respectively, both  $p$ 's  $< .001$ . The unconstrained fits were not significantly better than the constrained fits for either monkey:  $\chi^2(4) = 8.26$ ,  $p = .082$ , and  $\chi^2(4) = 4.09$ ,  $p = .394$  for monkey C and A, respectively. However, in the unconstrained fits, boosted fixation 1.0 fit better than interactive race 2.0 for both monkeys:  $\chi^2(4) = 15.26$  and  $48.72$  for monkey C and A, respectively, both  $p$ 's  $< .001$ .

Boosted fixation 1.0 did not fit substantially better than blocked-input 2.0. The models are not nested, so we cannot compare the fits with inferential statistics. For monkey C, the  $\chi^2$  values were numerically larger for boosted fixation 1.0 in both

constrained and unconstrained fits, but the differences were small. For monkey A,  $\chi^2$  for boosted fixation 1.0 was smaller in the constrained fits but larger in the unconstrained fits. Thus, we conclude that the fits were equivalent.

The equivalence of the fits of boosted fixation 1.0 and blocked input 2.0 lead us to conclude that some top-down process that interprets the reappearance of the fixation point as a stop signal seems necessary to tip the balance between gaze-holding and gaze-shifting neurons. Boosting activation in fix units and reducing activation in move units have essentially the same effect on predicted behavior.

**Neural predictions.** The mean activation functions for movement units, plotted in Figure 13, show growth, modulation, and decay that resemble the neural data. Predicted growth and decay rates (Figure 14) and predicted cancel times (Figure 15) fell within the 95% confidence intervals of observed values for both monkeys. However, the mean activation functions for fixation units, plotted in Figure 13, show a very steep rise in fixation activity that extends off the top of the scale after the boost parameter takes effect. This is not observed in fixation neurons (Hanes et al., 1998; Paré & Hanes, 2003) or in the other models, where fixation activity returns to the steady-state pre-movement level.

The steep rise and off-scale limit results from the very large boost parameters the model required to achieve good fits: For the constrained fits,  $b = 3.71$  and  $17.96$  for monkeys C and A, respectively; for the unconstrained fits,  $b = 2.99$  and  $17.69$  for monkeys C and A, respectively. These boost parameters result in predicted post-trial steady-state fixation activities after successful inhibition of 159 and 955 for monkeys C and A, respectively, compared to pre-trial steady-state fixation activities of 53 and 54, respectively. The post-inhibition fixation activities are very high. We simulated no-stop-signal trials following signal inhibit trials using these steady-state fixation activities to assess their predictions about slowing after successful inhibition, and found mean RTs of 395 and 719 ms for monkeys C and A, respectively. These values are much longer than the observed mean no-stop-signal RTs (254 and 271 ms for monkeys C and A, respectively), suggesting extremely large post-stop-signal slowing (114 and 448 ms for monkeys C and A, respectively; cf. Bissett & Logan, 2011; Emeric et al., 2007; Nelson, Boucher, Logan, Palmeri,

& Schall, 2010). The observed post-stop-signal slowing in these monkeys averaged 18 ms (Pouget et al., 2011).

Boosted fixation 1.0 will require some modification to capture post-stop-signal slowing and the dynamics of fixation neuron activity more accurately. Perhaps boosting fixation for a short period would be sufficient to capture behavior and neural dynamics. However, that would add at least one other parameter to the model (the duration of the boost), which increases the complexity of the model without much potential for improving goodness of fit. Blocked input 2.0 accounts for the behavioral data and neural data as well with fewer parameters.

### **Interim Summary and Evaluation 2.0**

Extending the models to account for steady-state fixation added constraints that improved the accuracy of neurophysiological predictions: predicted growth rates, decay rates, and cancel times were similar to observed values for all models. The constraints produced differences in goodness of fit to the behavioral data: The baseline model fit significantly worse than all other models. Interactive race 2.0 fit significantly worse than blocked input 2.0, breaking the mimicry between the 1.0 versions of these models. Boosted fixation 1.0 fit better than interactive race 2.0 and fit just as well as blocked input 2.0. However, it predicted very high steady-state activity in fixation cells after successful inhibition. Rectifying this prediction would add complexity to boosted fixation 1.0, so parsimony would favor blocked input 2.0. However, parsimony is a weak criterion. Both models require additional top-down processes that have not yet been specified computationally or identified with specific neurons. Without implementing the additional processes for both models, it is not possible to say which is simpler.

We compared models that assumed the stop process was inside the gaze control network (baseline and interactive race 2.0) with models that assumed the stop process was outside the gaze control network (blocked input 2.0 and boosted fixation 1.0). Baseline and interactive race 2.0 fit behavior significantly worse than blocked input 2.0 and boosted fixation 1.0, suggesting that a top-down stop process outside the network is necessary to explain response inhibition. This conclusion calls into question the linking proposition that identifies the stop process with fixation neurons and their inhibitory connections to

movement neurons (Boucher et al., 2007). Instead, the inhibition is mutual and maintains a balance between fixation and movement that is tipped in favor of inhibition by an outside process that blocks the input or generates a boost in fixation activity.

## Discussion

The main goal of this article was to evaluate models of saccadic countermanding that explain how responses are stopped, asking whether the stop process resides inside or outside the network of mutually-inhibitory gaze-holding and gaze-shifting neurons that extends from cortex to brainstem. Here, we discuss model mimicry, the implications for linking propositions that connect the computational models to neurons, and the potential for generalizing our models beyond saccades to other responses.

### Model Mimicry

An important result of our investigation is extensive model mimicry. Blocked input 1.0 mimicked interactive race 1.0, and boosted fixation 1.0 mimicked blocked input 2.0. Mimicry is common in fits to behavioral data (e.g., Ratcliff & Smith, 2004). Following Boucher et al. (2007), we found that neurophysiological data could break some of the mimicry among models (also see Hanes & Schall, 1996; Purcell et al., 2010, 2012). We found that requiring the models to account for steady-state fixation led to mimicry in neurophysiological predictions (see Figures 14-15). Following Newell (1990), we think that model mimicry is an informative result: It reveals fundamental constraints on an important computational problem that appear in many attempts to solve it.

**Mimicry reveals fundamental constraints.** The independent race model identified fundamental constraints on the computational problem of inhibiting responses with a stop process that races against a go process: the stop process has to beat the go process to inhibit the response (Logan & Cowan, 1984). The model was designed to produce mimicry: It was formulated in terms of generic distributions of finishing times for the stop and go processes without specifying the underlying processes, so every model of the underlying processes must make the same predictions (see Logan et al., 2014). Thus, the relationships between finishing time distributions identified in the independent race model represent fundamental constraints on all race models of stopping.

The models we considered in this article address the computational problem of stopping responses that are generated by a stochastic accumulation process. In all of the models, the go unit accumulates activation until it reaches a threshold for a response. The fundamental constraint in these models is that responses must be stopped by preventing go unit activation from reaching threshold (also see Salinas & Stanford, 2013). This can be done by raising the threshold or by reducing the rate of accumulation. The physiological data rule out raising the threshold (Hanes et al., 1998; Hanes & Paré, 2003; also see De Jong et al., 1990, 1995). All of the models we considered stop responses by reducing the rate of accumulation. Equation 5, which describes the rate of accumulation in the go unit, allows three possibilities: blocking the input (reducing  $\mu_{go}$ ), increasing inhibition from the stop unit (increasing  $a_{stop}(t)$  or  $\beta_{stop}$ ), or increasing leakage (increasing  $k_{go}$ ). All models that change these parameters without further constraint should produce equivalent changes in behavior, though they may predict different neural dynamics. Changing the rate can stop responses only if the change occurs before the go unit hits threshold, so go RT will be faster on signal-respond trials than on trials with no stop signals, as observed. Changing the rate can stop responses only if the change occurs before SSRT, so cancel-times will be zero or less, as observed.

**Breaking mimicry with neurophysiological data.** In our approach, successful models had to fit behavioral data well and predict neural data accurately (Boucher et al., 2007; Purcell et al., 2010, 2012). Neural data broke the mimicry between interactive race 1.0 and blocked input 1.0. Interactive race 1.0 predicted decay rates that were much steeper than the rates observed in movement neurons. Neural data did not distinguish between models of fixation and movement initiation. All models predicted cancel times and growth and decay rates that were similar to each other and similar to the observed measures from movement neurons. However, the constraints that corrected the models' neural predictions broke the mimicry in the behavioral fits. Baseline fit worse than any other model, and interactive race 2.0 fit worse than blocked input 2.0 and boosted fixation 1.0. Neural data may break the mimicry between blocked input 2.0 and boosted fixation 1.0: After successful inhibition, boosted fixation 1.0 predicts a rapid rise in fixation unit activity to a steady state that is 3-17 times higher than the steady state activity before the

movement. Actual fixation neurons return to pre-movement firing rates after successful inhibition (Hanes et al., 1998).

**Breaking mimicry with converging computational constraints.** Our models suggest the importance of converging computational constraints on the search for underlying mechanisms. The purported mechanism should serve more than one computational function, so the constraints required to implement one function constrain the implementation of the other. The requirement to maintain fixation constrained the interaction between fix and move units in a way that made it difficult for interactive race 2.0 to account for response inhibition accurately. This principle of converging computational constraints is similar to the principle of converging operations (Garner, Hake & Eriksen, 1956) that guided the search for mental operations in the early days of cognitive psychology. Psychology and neuroscience have advanced a lot in the meantime, but the principle of converging constraints remains valid and useful.

**Inevitable mimicry of blocking and inhibiting?** Our modeling exercise can be viewed pessimistically as a never-ending competition between inhibition and blocking as mechanisms of stopping, in which we constantly revise each model to address challenges from the other. Interactive race 1.0 assumes that inhibition causes stopping. Blocked input 1.0 assumes blocking and works as well or better. Blocked input 2.0 fit better than interactive race 2.0 but no better than boosted fixation 1.0, which invokes inhibition through another mechanism (boost, not  $\beta_{stop}$ ). Indeed, blocking could be accomplished by removing input (e.g., by moving the eyes or closing them) or by inhibiting it, leading to another series of models and revisions. Much of the research on distinguishing other forms of inhibition from non-inhibitory alternatives has followed a similar course of revision without resolution (see MacLeod, Dodd, Sheard, Wilson, & Bibi, 2003). Historically, cognitive research on broad binary distinctions has followed the same course, leading Newell (1973) to claim, "You can't play 20 questions with nature and win" (p. 283).

We view our modeling effort more optimistically. Broad concepts like inhibition and blocking input cannot be tested rigorously unless they are instantiated in specific architectures described by specific equations (also see Newell, 1990; Townsend, 1990; Townsend & Ashby, 1983). From this perspective, each of the models we examined is a

different instantiation of the general principle of inhibition or blocking, and our model fits and neural predictions allowed us to rule out some specific instantiations of the principles and guide us toward the more effective instantiations. We found problems with each of the inhibition models (interactive race 1.0 over-predicted decay rates; interactive race 2.0 misfit the behavioral data; boosted fixation 1.0 over-predicted post-inhibition steady-state fixation rates) and no problems with the blocking models (blocked input 1.0 and 2.0 both fit the behavioral data well and predicted the neural data accurately). It is hard to say whether blocking or inhibition is superior in general, but it was clear in these particular cases. Evaluating models case by case is the road to progress.

Our models may be less susceptible to mimicry because of our core assumptions that link model components to specific neurons and the connections between them. Inhibition and excitation may be similar mathematically and produce similar effects on behavior, but they are very different neurophysiologically. Inhibitory and excitatory neurons have different morphology and are mediated by different transmitter substances. Animal models allow pharmacological interventions, stimulation, and surgery that target inhibitory or excitatory systems to distinguish the alternatives (e.g., Eagle, Bari, & Robbins, 2008). Recordings from single neurons provide important converging data that can break mimicry.

In the spirit of competitive hypothesis testing, we have treated blocking and inhibition as mutually exclusive acts of control, but there is no strong reason to do so. Equation 5 shows that blocking and inhibition modulate separate components of growth rate: blocking changes  $\mu_{move}$  and inhibition changes  $\beta_{fix} \cdot a_{fix}(t)$ . Nothing in the mathematics or the theory prevents both components from changing when a stop signal occurs. Indeed, there may be much to be learned from asking whether people and animals can use both mechanisms. The possibility of two mechanisms might give new insights into inhibitory deficits in clinical, developmental, neurological, and neuroscientific research. Deficits may differ depending on the inhibitory mechanism that is affected. Sensory or attentional challenges may affect blocking while motor challenges may affect inhibition.

## Linking Propositions

A main goal of this research is to link computational models of response inhibition to the underlying physiology. All of the models we considered assume that the computation underlying the go process is stochastic accumulation to a threshold, and all of the models link this computation to gaze-shifting neurons in frontal eye fields and superior colliculus. This linking proposition is well established in previous research (Hanes & Schall, 1996; Pouget et al., 2011; Purcell et al., 2010, 2012; Ratcliff et al., 2003; also see Gold & Shadlen, 2007; Shadlen & Kiani, 2013) and consistent with our current findings. Our models differ in their assumptions about the computational mechanism underlying the stop process (inhibition, blocking, or boosting) and make different linking propositions that connect them to the underlying physiology. Interactive race 1.0 and 2.0 link the stop process to exclusively fixation neurons in frontal eye fields and superior colliculus and the inhibitory connections from fixation neurons to movement neurons (also see Boucher et al., 2007). Blocked input 1.0 and 2.0 and boosted fixation 1.0 emphasize the contribution of stop processes outside the network of fixation and movement neurons that tip the balance in the network in different ways.

The model fits call into question the linking propositions in the interactive race models that identify the stop process with fixation neurons and their connections to movement neurons. Blocked input 1.0 fit the behavioral data as well as interactive race 1.0, suggesting that inhibition from fixation neurons to movement neurons is not necessary to account for response inhibition. Inhibition may be sufficient, but blocking the input is also sufficient. Interactive race 2.0 did not fit as well as blocked input 2.0 or boosted fixation 1.0, suggesting that the stop process lies outside the network of fixation and movement neurons. The stop process shifts the balance in the network (Schall, 2004) but lies outside it. However, we have not specified the linking propositions that connect blocking the input and boosting fixation to actual neurons or neural networks. That is an important goal for future work.

These conclusions suggest the importance of distinguishing between the *source* and the *site* of response inhibition (Band & van Boxtel, 1999). The site is the locus at which go responses are actually inhibited – the point immediately before the point of no return (De Jong et al., 1990; Logan, 1981; Osman et al., 1986, 1990). The source is the locus at which the significance of the stop and go stimuli is apprehended and from which an effect on the

site is propagated. All of the models assume that the site of response inhibition is the network of gaze-holding and gaze-shifting neurons that ranges from cortex to brainstem. They differ in their assumptions about the source. Interactive race models assume it is inside the gaze control network, while blocked input and boosted fixation models assume it is outside the gaze control network.

Functional magnetic resonance imaging (fMRI) studies reveal an extensive network of brain areas involved in stopping responses, including motor cortex, premotor cortex, supplementary motor area, anterior cingulate cortex, inferior frontal cortex, basal ganglia, and subthalamic nucleus (Aron et al., 2007; Ridderinkhof, van den Wildenberg, Segalowitz & Carter, 2004; Swick, Ashley & Turken, 2011; Zandbelt, Bloemendaal, Hoogendam, Kahn, & Vink, 2013). The site of inhibition is likely motor cortex for keypress responses and frontal eye fields for eye movements (Curtis, Cole, Rao & D'Esposito, 2004). The source of inhibition could be in one or more of the other areas. However, activation in these areas may be a consequence of inhibition rather than the cause of it, and distinguishing consequences from causes requires more temporal precision than fMRI can provide. Electrophysiology has the required temporal resolution and may help disambiguate fMRI results. Single cell recordings from supplementary eye fields (Stuphorn et al., 2000, 2010) and anterior cingulate cortex (Ito et al., 2003) reveal neurons whose activity modulates on stop-signal trials, but the modulation occurs well after SSRT, suggesting it is a consequence of inhibition rather than a cause (also see Chen, Scangos, & Stuphorn, 2010; Scangos Aronberg, & Stuphorn, 2013; Scangos & Stuphorn, 2009; Schmidt, Leventhal, Mallet, Chen, & Berke, 2013).

Ultimately, specifying the source and the site will require specifying the connections between them. Further developments will include specifying the connections between perceptual inputs and the decision mechanism in the source. Imaging studies are developing methods for identifying connectivity between brain areas in humans (Aron, Behrens, Smith, Frank, & Poldrack, 2007; Duann, Ide, Luo, & Ray Li, 2009; Zandbelt et al., 2013). Animal studies afford more direct measures. In our modeling of saccadic inhibition, we assume visually responsive neurons are the inputs to movement related neurons in frontal eye fields (Purcell et al., 2010; 2012). Specifying the inputs to the visually responsive neurons is an important goal for our future research.

## Beyond Saccades

The models addressed in this article focus on behavior and physiology in countermanding saccadic eye movements. We believe eye movements provide an ideal model system in which to link computation, behavior, and neurophysiology. The computational constraints and the underlying neurophysiology are well understood, and that allows us to formulate and test hypotheses about linking propositions. Eye movements are simple—there are basically only two degrees of freedom to control, according to Listing's law. Body movements involve many more degrees of freedom, some of which are coupled dynamically to perform the task at hand. However, the stop signal studies we hope to inform use keypress responses, which involve simple flexion movements of single fingers. We believe our models will generalize to keypresses and other movements.

Saccades may be special: Saccadic SSRTs are faster than manual SSRTs (Boucher, Stuphorn, Logan, Schall & Palmeri, 2007; Logan & Irwin, 2000) and are affected by different variables (Logan & Irwin, 2000). Manual stop-signal tasks pervade the countermanding literature (Logan, 1994; Verbruggen & Logan, 2008), and manual SSRTs are not different from vocal SSRTs (van den Wildenberg & Christoffels, 2010; Xue, Aron & Poldrack, 2008) or foot SSRTs (if measured from the onset of electromyographic activity; Tabu, Mima, Aso, Takahashi & Fukuyama, 2012). Brunamonti et al. (2012) found no difference in SSRTs with finger, wrist, and arm movements. Thus, saccadic SSRTs may be outliers. Does this challenge our assumption that eye and hand movements are stopped by the same mechanism (Brunamonti et al., 2012; Logan & Irwin, 2000)?

From a strict anatomical perspective, the answer must be “yes.” Stopping a response modulates activity in motor cortex, among other loci, and eye and hand movements involve different regions of motor cortex: the frontal eye fields, which control eye movements, are anterior to the motor strip, which controls hand movements. However, we believe the models we considered can be extended naturally to countermanding of other responses. The circuits that underlie eye, hand, and body movements in cortex, basal ganglia, and thalamus are more similar than different. Stopping a manual response requires modulation of activity in motor and premotor cortex

(Mirabella et al. 2011) and basal ganglia (Schmidt et al. 2013), while stopping a saccade requires modulation of activity in frontal eye field, superior colliculus and basal ganglia circuits.

From a functional perspective, the answer is “no.” Differences in SSRT do not challenge our generalization from eye movements to other kinds of movements. Eye movements show the same qualitative and quantitative effects as hand movements, as predicted by the independent race model: monotonic inhibition functions and signal-respond RTs that are faster than no-stop-signal RTs (Boucher, Stuphorn, et al., 2007; Hanes & Carpenter, 1999; Hanes & Schall, 1995; Logan & Irwin, 2000; Middlebrooks & Schall, 2014). The same models may apply but with different parameters to accommodate the differences in RT.

From a theoretical perspective, the answer is also “no.” There are two components to our theory: The stochastic accumulator models and the neurons they are mapped onto by our linking propositions. The stochastic accumulator models address behavior, and so could be fit to behavioral data from any stop-signal task, regardless of the response. If we disregard the process that stops the response (blocking or inhibition) and assume there is no leakage, then our models are examples of the “special independent race models” that Logan et al. (2014) developed for keypress responses. In these models, one stop diffusion races against  $N$  go diffusions, one for each possible go response. For saccades, choice errors are rare, so  $N = 1$ . Thus, the mathematics and computations of our models could be extended easily to other responses.

Extending our models to other responses would require us to change the specific linking propositions that connect components of the stochastic accumulators to particular neurons. After all, keypresses are not controlled by the frontal eye fields. However, other responses invite parallel sets of linking propositions that connect model components to the neurons and neural circuits that control those responses. We hope that our models and analyses might guide researchers in searching for the neurons that link to the computations.

## **The Independent Race Model Revisited**

The current models invite comparisons with independent race models (Logan & Cowan, 1984; Logan et al., 2014). All of the models predict inhibition functions, signal-respond RT, and SSRT. The current models go beyond independent race models in asking how response inhibition occurs. Our demonstrations of model mimicry reveal several possible mechanisms for stopping responses but do not allow clear conclusions about which mechanism best accounts for stop-signal data. Inhibiting go activation works as well as blocking the input (i.e., interactive race 1.0 fit as well as blocked input 1.0, and boosted fixation 1.0 fit as well as blocked input 2.0). All of our models assume growth of activation to a threshold, and any mechanism that stops or reverses that growth seems sufficient to explain behavioral and physiological data. Thus, it is reasonable to ask what we gain from modeling the process that stops the response.

The current model fits cannot be compared with fits of the original independent race model because that model is non-parametric. The independent race model makes predictions for all finishing time distributions, regardless of their form, and its predictions are not evaluated by fitting the model to data (Logan & Cowan, 1984). Boucher et al. (2007) developed a version of the independent race model in which stop and go units were stochastic accumulators and the race ended when the first process hit its threshold (cf. Logan et al., 2014). They compared this model to the interactive race model and found equivalent fits (see Boucher et al., 2007, Table 1). Our own comparisons of the same independent race model to interactive race 1.0 also produced equivalent fits (see Appendix B, Table B1).

An important strength of independent race models is that they provide measures of SSRT (Colonius, 1990; De Jong et al., 1990; Logan & Cowan, 1984; Logan et al., 2014; Matzke et al., 2013), which have been useful in clinical science, developmental science, neurology, and neuroscience (Verbruggen & Logan, 2008). The measures of SSRT are easy to calculate (no model fitting is required), so investigators who are primarily interested in SSRT might find it more convenient to use the Logan and Cowan (1984) independent race model (see Band et al., 2003; Verbruggen et al., 2013).

An important strength of the models we consider here is that they provide measures of the components of SSRT, separating the afferent processing of the stop signal from the interactive processing that reverses go activation and stops the response: SSRT = afferent

time + interactive time (Boucher et al., 2007). Separating components may be important beyond our immediate goal of linking models to neurons. It may provide important insight into SSRT deficits in clinical science, developmental science, neurology, and neuroscience, distinguishing deficits in afferent processing from deficits in executing the interaction. Populations with sensory or attentional challenges may show prolonged afferent processing. Populations with motor challenges may show a prolonged interactive period. These measures are harder to obtain than SSRT (model fitting is required), but the potential gain in understanding may be well worth the effort.

An important point, emphasized by Boucher et al. (2007), is that the afferent times are large, relative to SSRT, in all of the models we analyzed. Thus, the stop process is independent of the go process for much of its duration. The interaction between stopping and going that reverses the trajectory of go activation is late and potent, as Boucher et al. (2007) observed. Thus, the models approximate the independent race model. Much of their ability to predict behavior may follow from the closeness of this approximation. The principles underlying the independent race model may be doing most of the work.

## **Conclusions**

The Boucher et al. (2007) interactive race model was an important step forward in understanding saccadic countermanding because it specified the mechanism that stops responses, and it provided an accurate account of behavioral and physiological data. The present article takes another step forward by proposing and testing alternative, blocked input models that specify a different mechanism for stopping responses. We showed that the blocked input model accounts for behavioral and physiological data as well as or better than the interactive race model. The blocked input model provides a different perspective on the linking propositions that relate computational and physiological mechanisms.

## References

- Ackerfelt, A., Colonius, H., & Diederich, A. (2006). Visual-tactile saccadic inhibition. *Experimental Brain Research, 169*, 554-563.
- Aron, A. R., Behrens, T. E., Smith, S., Frank, M. J., & Poldrack, R. A. (2007). Triangulating a cognitive control network using diffusion-weighted magnetic resonance imaging (MRI) and functional MRI. *Journal of Neuroscience, 27*, 3743-3752.
- Aron, A. R., Duston, S., Eagle, D. M., Logan, G. D., Stinear, C. M., & Stuphorn, V. (2007). Converging evidence for a fronto-basal-ganglia system for inhibitory control of action and cognition. *Journal of Neuroscience, 27*, 11860-11864.
- Awh, E., Matsukura, M., & Serences, J. T. (2003). Top-down control over biased competition during covert spatial orienting. *Journal of Experimental Psychology: Human Perception and Performance, 29*, 52-63.
- Band, G. P. H., & van Boxtel, G. J. M. (1999). Inhibitory motor control in stop paradigms: Review and reinterpretation of neural mechanisms. *Acta Psychologica, 101*, 179-211.
- Band, G. P. H., van der Molen, M. W., & Logan, G. D. (2003). Horse-race model simulations studies of the stop signal procedure. *Acta Psychologica, 112*, 105-142.
- Bissett, P. G., & Logan, G. D. (2011). Balancing cognitive demands: Control adjustments in the stop-signal paradigm. *Journal of Experimental Psychology: Learning, Memory and Cognition, 37*, 392-404.
- Boucher, L., Palmeri, T. J., Logan, G. D., & Schall, J. D. (2007). Inhibitory control in mind and brain: An interactive race model of countermanding saccades. *Psychological Review, 114*, 376-397.
- Boucher, L., Stuphorn, V., Logan, G. D., Schall, J. D., & Palmeri, T. J. (2007). Stopping eye and hand movements: Are the processes independent? *Perception & Psychophysics, 69*, 785-801.
- Bruce, C. J., & Goldberg, M. E. (1985). Primate frontal eye fields. I. Single neurons discharging before saccades. *Journal of Neurophysiology, 53*, 603-635.
- Brown, S.D., & Heathcote, A. (2008) The simplest complete model of choice reaction time: Linear ballistic accumulation. *Cognitive Psychology, 57*, 153-178.
- Brunamonti, E., Ferraina, S., & Paré, M. (2012). Controlled movement processing: Evidence

- for a common inhibitory control of finger, wrist, and arm movements. *Neuroscience*, *215*, 69-78.
- Camalier, C. R., Gotler, A., Murthy, A., Thompson, K. G., Logan, G. D., Palmeri, T. J., & Schall, J. D. (2007). Dynamics of saccade target selection: Race model analyses of double step and search step saccade production in human and macaque. *Vision Research*, *47*, 2187-2211.
- Chen, X., Scangos, K. W., & Stuphorn, V. (2010). Supplementary motor area exerts proactive and reactive control of arm movements. *Journal of Neuroscience* *30*, 14657-14675.
- Cohen, J. Y., Pouget, P., Heitz, R. P., Woodman, G. F., & Schall, J. D. (2009). Biophysical support for functionally distinct cell types in the frontal eye field. *Journal of Neurophysiology*, *101*, 912-916.
- Cohen-Kdoshay, O., & Meiran, N. (2009). The representation of instructions operates like a prepared reflex: Flanker compatibility effects found in first trial following S-R instructions. *Experimental Psychology*, *56*, 128-133.
- Colonius, H. (1990). A note on the stop-signal paradigm, or how to observe the unobservable. *Psychological Review*, *97*, 309-312.
- Corneil, B. D., & Elsley, J. K. (2005). Countermanding eye-head gaze shifts in humans: marching orders are delivered to the head first. *Journal of Neurophysiology*, *94*, 883-895.
- Curtis, C. E., Cole, M. W., Rao, V. Y., & D'Esposito, M. (2005). Canceling planned action: An fMRI study of countermanding saccades. *Cerebral Cortex*, *15*, 1281-1289.
- De Jong, R., Coles, M. G. H., Logan, G. D. & Gratton, G. (1990). Searching for the point of no return: The control of response processes in speeded choice reaction performance. *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 164-182.
- De Jong, R., Coles, M. G. H., & Logan, G. D. (1995). Strategies and mechanisms in nonselective and selective inhibitory motor control. *Journal of Experimental Psychology: Human Perception and Performance*, *21*, 498-511.
- Ding, L., Gold, J.I. (2012) Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field. *Cerebral Cortex*, *22*, 1052-1067.

- Donkin, C., Brown, S. D., & Heathcote, A. (2009). The over-constraint of response time models: Rethinking the scaling problem. *Psychonomic Bulletin & Review*, *16*, 1129-1135.
- Duann, J.-R., Ide, J. S., Luo, X., & Ray Li, C.-S. (2009). Functional connectivity delineates distinct roles of the inferior frontal cortex and presupplementary motor area in stop signal inhibition. *Journal of Neuroscience*, *29*, 10171-10179.
- Eagle, D. M., Bari, A., & Robbins, T. W. (2008). The neuropsychopharmacology of action inhibition: Cross-species translation of the stop-signal and go/no-go tasks. *Psychopharmacology*, *199*, 439-456.
- Emeric, E. E., Brown, J. W., Boucher, L., Carpenter, R. H. S., Hanes, D. P., Harris, R., Logan, G. D., Mashru, R. N., Paré, M., Pouget, P., Stuphorn, V., Taylor, T. L., & Schall, J. D. (2007). Influence of history on countermanding saccade performance in humans and macaque monkeys. *Vision Research*, *47*, 35-49.
- Garner, W. R., Hake, H. W., & Eriksen, C. W. (1956). Operationism and the concept of perception. *Psychological Review*, *63*, 149-159.
- Gold, J. I., & Shadlen, M. N. (2007). The neural basis of decision making. *Annual Review of Neuroscience*, *30*, 535-560.
- Hanes, D. P., & Carpenter, R. H. S. (1999). Countermanding saccades in humans. *Vision Research*, *39*, 2777-2791.
- Hanes, D. P., Patterson, W. F., II, & Schall, J. D. (1998). Role of frontal eye fields in countermanding saccades: Visual, movement, and fixation activity. *Journal of Neurophysiology*, *79*, 817-834.
- Haupt, R. L., & Haupt, S. E. (2004). *Practical genetic algorithms* (2nd ed.). Hoboken, NJ: Wiley.
- Hommel, B. (2000). The prepared reflex: Automaticity and control in stimulus-response translation. In S. Monsell & J. Driver (Eds.), *Control of cognitive processes: Attention and Performance XVIII* (pp. 247-273). Cambridge, MA: MIT Press.
- Ito, S., Stuphorn, V., Brown, J. W., & Schall, J. D. (2003). Performance monitoring by the anterior cingulate cortex during saccade countermanding. *Science*, *302*, 120-122.
- Lappin, J. S., & Eriksen, C. W. (1966). Use of a delayed signal to stop a visual reaction-time response. *Journal of Experimental Psychology*, *72*, 805-811.

- Leite, F. P., & Ratcliff, R. (2010). Modeling reaction time and accuracy of multiple-alternative decisions. *Attention, Perception, & Psychophysics, 72*, 246-273.
- Lo, C.-C., Boucher, L., Paré, M., Schall, J. D., & Wang, X.-J. (2009). Proactive inhibitory control and attractor dynamics in countermanding action: A spiking neural circuit model. *Journal of Neuroscience, 29*, 9059-9071.
- Logan, G. D. (1978). Attention in character classification: Evidence for the automaticity of component stages. *Journal of Experimental Psychology: General, 107*, 32-63.
- Logan, G. D. (1981). Attention, automaticity, and the ability to stop a speeded choice response. In J. Long & A. D. Baddeley (Eds.), *Attention and Performance IX*. Hillsdale, NJ: Erlbaum.
- Logan, G. D. (1983). On the ability to inhibit simple thoughts and actions: I. Stop signal studies of decision and memory. *Journal of Experimental Psychology: Learning, Memory and Cognition, 9*, 585-606.
- Logan, G. D. (1994). On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In D. Dagenbach & T. H. Carr (Eds.), *Inhibitory processes in attention, memory, and language*. (pp. 189-239). San Diego: Academic Press.
- Logan, G. D. & Burkell, J. (1986). Dependence and independence in responding to double stimulation: A comparison of stop, change, and dual-task paradigms. *Journal of Experimental Psychology: Human Perception and Performance, 12*, 549-563.
- Logan, G. D. & Cowan, W. B. (1984). On the ability to inhibit thought and action: A theory of an act of control. *Psychological Review, 91*, 295-327.
- Logan, G. D., & Gordon, R. D. (2001). Executive control of visual attention in dual-task situations. *Psychological Review, 108*, 393-434.
- Logan, G. D., & Irwin, D. E. (2000). Don't look! Don't touch! Inhibitory control of eye and hand movements. *Psychonomic Bulletin & Review, 7*, 107-112.
- Logan, G. D., Van Zandt, T., Verbruggen, F., & Wagenmakers, E.-J. (2014). On the ability to inhibit thought and action: General and special theories of an act of control. *Psychological Review, 121*, 66-95.
- MacLeod, C. M., Dodd, M. D., Sheard, E. D., Wilson, D. E., & Bibi, U. (2003). In opposition to inhibition. In B. H. Ross (Ed.), *The psychology of learning and motivation, 43*, 163-214.

- Marder, E., & Goaillard, J. M. (2006). Variability, compensation and homeostasis in neuron and network function. *Nature Reviews Neuroscience*, *7*, 563-574.
- Matzke, D., Dolan, C. V., Logan, G. D., Brown, S. D., & Wagenmakers, E.-J. (2013). Bayesian parametric estimation of stop-signal reaction time distributions. *Journal of Experimental Psychology: General*, in press.
- McAdams, C. J., & Maunsell, J. H. R. (1999). Effects of attention on orientation-tuning functions in single neurons in macaque cortical area V4. *Journal of Neuroscience*, *19*, 431-441.
- Middlebrooks, P. G., & Schall, J. D. (2014). Response inhibition during perceptual decision making in humans and macaques. *Attention, Perception & Psychophysics*, in press.
- Miller, E. K., & Cohen, J. D. (2001). An integrative theory of prefrontal cortex function. *Annual Review of Neuroscience*, *24*, 167-202.
- Mirabella, G., Pani, P., & Ferraina, S. (2011). Neural correlates of cognitive control of reaching movements in the dorsal premotor cortex of rhesus monkeys. *Journal of Neurophysiology*, *106*, 1454-1466.
- Murthy, A., Ray, S., Shorter, S. M., Schall, J. D., & Thompson, K. G. (2009). Neural control of visual search by frontal eye field: Effects of unexpected target displacement on visual selection and saccade preparation. *Journal of Neurophysiology*, *101*, 2485-2506.
- Nelson, M. J., Boucher, L., Logan, G. D., Palmeri, T. J., & Schall, J. D. (2010). Nonindependent and nonstationary response times in stopping and stepping saccade tasks. *Attention, Perception & Psychophysics*, *72*, 1913-1929.
- Newell, A. (1973). You can't play 20 questions with nature and win: Projective comments on the papers of this symposium. In W. G. Chase (Ed.), *Visual information processing*. (pp. 283-308). New York: Academic Press.
- Newell, A. (1990). *Unified theories of cognition*. Cambridge, MA, Harvard University Press.
- Osman, A., Kornblum, S., & Meyer, D. E. (1990). Does response programming necessitate response execution? *Journal of Experimental Psychology: Human Perception and Performance*, *16*, 183-198.
- Osman, A., Kornblum, S., & Meyer, D. E. (1986). The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. *Journal of Experimental Psychology: Human Perception and Performance*, *12*, 243-258

- Paré, M., & Hanes, D. P. (2003). Controlled movement processing: Superior colliculus activity associated with countermanded saccades. *Journal of Neuroscience*, *23*, 6480-6489.
- Pouget, P., Logan, G. D., Palmeri, T. J., Boucher, L., Paré, M., & Schall, J. D. (2011). Neural basis of adaptive response time adjustment during saccade countermanding. *Journal of Neuroscience*, in press
- Purcell, B. A., Heitz, R. P., Cohen, J. Y., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2010). Neurally constrained modeling of perceptual decision making. *Psychological Review*, *117*, 1113-1143.
- Purcell, B. A., Schall, J. D., Logan, G. D., & Palmeri, T. J. (2012). From salience to saccades: Multiple-alternative gated stochastic accumulator model of visual search. *Journal of Neuroscience*, *32*, 3433-3446.
- Ramakrishnan, A., Sureshbabu, R., & Murthy, A. (2012). Understanding how the brain changes its mind: Microstimulation in the macaque frontal eye fields reveals how saccade plans are changed. *The Journal of Neuroscience*, *32*, 4457-4472.
- Ratcliff, R., Cherian, A., & Segraves, M. (2003). A comparison of macaque behavior and superior colliculus neuronal activity to predictions from models of two-choice decisions. *Journal of Neurophysiology*, *90*, 1392-1407.
- Ratcliff, R., Hasegawa, Y. T., Hasegawa, R. P., Smith, P. L., & Segraves, M. A. (2007). Dual diffusion model for single-cell recording data from the superior colliculus in a brightness-discrimination task. *Journal of Neurophysiology*, *97*, 1756-1774.
- Ratcliff, R., & Smith, P. L. (2004). A comparison of sequential sampling models for two-choice reaction time. *Psychological Review*, *111*, 333-367.
- Ray, S., Pouget, P., & Schall, J. D. (2009). Functional distinction between visuomovement and movement neurons in macaque frontal eye field during saccade countermanding. *Journal of Neurophysiology*, *102*, 3091-3100.
- Ridderinkhof, K. R., van den Wildenberg, W. P. M., Segalowitz, S. J., & Carter, C. S. (2004). Neurocognitive mechanisms of cognitive control: The role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. *Brain and Cognition*, *56*, 129-140.
- Rieger, M., & Gauggel, S. (1999). Inhibitory after-effects in the stop signal paradigm. *British Journal of Psychology*, *90*, 509-518.

- Salinas, E., & Stanford, T. R. (2013). The countermanding task revisited: Fast stimulus detection is a key determinant of psychophysical performance. *Journal of Neuroscience*, *33*, 5668-5685.
- Scangos K. W., Aronberg, R., & Stuphorn, V. (2013). Performance monitoring by presupplementary and supplementary motor area during an arm movement countermanding task. *Journal of Neurophysiology*, *109*, 1928-1939.
- Scangos, K. W., & Stuphorn, V. (2010). Medial frontal cortex motivates but does not control movement initiation in the countermanding task. *Journal of Neuroscience*, *30*, 1968-1982.
- Schachar, R., Logan, G. D., Chen, S., Ickowiz, A., & Barr, C. (2007). Restraint and cancellation: Multiple inhibition deficits in attention deficit hyperactivity disorder. *Journal of Abnormal Child Psychology*, *35*, 229-238.
- Schall, J. D. (1991). Neuronal activity related to visually guided saccades in the frontal eye fields of rhesus monkeys: Comparison with supplementary eye fields. *Journal of Neurophysiology*, *66*, 559-579.
- Schall, J. D. (2004). On building a bridge between brain and behavior. *Annual Review of Psychology*, *55*, 23-50.
- Schall, J. D., & Boucher, L. (2007). Executive control of gaze by the frontal lobes. *Cognitive, Affective & Behavioral Neuroscience*, *7*, 396-412.
- Schmidt, R., Leventhal, D. K., Mallet, N., Chen, F., & Berke, J. D. (2013). Canceling actions involves a race between basal ganglia pathways. *Nature Neuroscience*, *16*, 1118-1124.
- Shadlen, M. N., & Kiani, R. (2013). Decision making as a window on cognition. *Neuron*, *80*, 791-806.
- Stuphorn, V., & Schall, J. D. (2006). Executive control of countermanding saccades by the supplementary eye field. *Nature Neuroscience*, *9*, 925-931.
- Swick, D., Ashley, V., & Turken, A. U. (2011). Are the neural correlates of stopping and not going identical? Quantitative meta-analysis of two response inhibition tasks. *NeuroImage*, *56*, 1655-1665.

- Tabu, H., Mima, T., Aso, T., Takahashi, R., & Fukuyama, H. (2012). Common inhibitory prefrontal activation during inhibition of hand and foot responses. *Neuroimage*, *59*, 3373-3378.
- Teller, D. Y. (1984). Linking propositions. *Vision Research*, *24*, 1233-1246.
- Thompson, K. G., Hanes, D. P., Bichot, N. P., & Schall, J. D. (1996). Perceptual and motor processing stages identified in the activity of macaque frontal eye field neurons during visual search. *Journal of Neurophysiology*, *76*, 4040-4055.
- Townsend, J. T. (1990). Serial vs. parallel processing: Sometimes they look like Tweedledum and Tweedeldee but they can (and should) be distinguished. *Psychological Science*, *1*, 46-54.
- Townsend, J. T., & Ashby, F. G. (1983). *Stochastic modeling of elementary psychological processes*. Cambridge: Cambridge University press.
- Treue, S., & Martinez-Trujillo, J. C. (1999). Feature-based attention influences motion processing gain in macaque visual cortex. *Nature*, *399*, 575-579.
- Usher, M., & McClelland, J. L. (2001). The time course of perceptual choice: The leaky, competing accumulator model. *Psychological Review*, *108*, 550-592.
- van den Wildenberg, W. P. M., & Christoffels, I. K. (2010). STOP TALKING! Inhibition of speech is affected by word frequency and dysfunctional impulsivity. *Frontiers in Psychology*, *1*(145).
- Verbruggen, F., Chambers, C. D., & Logan, G. D. (2013). Fictitious inhibitory differences: How skewness and slowing distort the estimation of stopping latencies. *Psychological Science*, *24*, 352-362.
- Verbruggen, F., & Logan, G. D. (2008). Response inhibition in the stop-signal paradigm. *Trends in Cognitive Sciences*, *12*, 418-424.
- Verbruggen, F., & Logan, G. D. (2009). Proactive adjustments of response strategies in the stop-signal paradigm. *Journal of Experimental Psychology: Human Perception and Performance*, *35*, 835-854.
- Wasserman, L. (2000). Bayesian model selection and model averaging. *Journal of Mathematical Psychology*, *44*, 92-107.
- Wong-Lin, K., Eckhoff, P., Holmes, P., & Cohen, J. D. (2010). Optimal performance in a countermanding saccade task. *Brain Research*, *1318*, 178-187.

Woodman, G. F., Kang, M.-S, Thompson, K., & Schall, J. D. (2009). Visual search efficiency modulates the onset of response preparation: Neurophysiological evidence for discrete flow. *Psychological Science, 19*, 128-136.

Xue, G., Aron, A. R., & Poldrack, R. A. (2008). Common neural substrates for inhibition of spoken and manual responses. *Cerebral Cortex, 18*, 1923-932.

Zandbelt, B. B., Bloemendaal, M., Hoogendam, J. M., Kahn, R. S., & Vink, M. (2013). Transcranial magnetic stimulation and functional MRI reveal cortical and subcortical interactions during stop-signal response inhibition. *Journal of Cognitive Neuroscience, 25*, 157-174.

## Appendix A: Description of Fitting Procedure

### Standard Fitting Method

All models were fitted separately to two sets of behavioral data acquired from two monkeys who performed a saccadic countermanding task (see Hanes et al., 1998, for details of the data collection procedure). Model parameters were optimized by minimizing chi-square between the behavioral data and model predictions, as in Boucher et al. (2007).

The behavioral data consisted of  $p(\text{respond}|\text{signal})$  and RT for signal-respond trials at each SSD and for no-signal trials (see Boucher et al., 2007, for detailed descriptions of the data selection procedure). There were 2,431 trials from monkey C and 2,928 trials from monkey A. For conditions in which the total number of trials was greater than 40, RTs were binned into five intervals, each bounded by a quintile above (i.e., 0-20%, 20-40%, ..., 80-100%). For conditions in which the total number of trials was smaller than 40 (i.e., for short SSDs), all RTs were grouped into a single bin. The frequency ( $f_{ij}$ ) of RTs for each quintile bin and for signal-inhibit trials for each condition served as observed data points to which model predictions were compared (where  $i$  indexes SSD and no-signal conditions and  $j$  indexes quintile bins and signal-inhibit trials).

To generate model predictions, a model was run 1500 trials for each SSD and no-signal conditions. For monkey C, there were four SSDs and one no-stop-signal condition, for a total of 7,500 simulated trials. For monkey A, there were six SSDs and one no-stop-signal condition for a total of 10,500 simulated trials. The proportions ( $p_{ij}$ ) of signal-inhibit trials and those of signal-respond trials that fall within the quintile bins, as defined by the behavioral data, were computed. The chi-square was calculated by

$$\chi^2 = \sum_i \sum_j \frac{(f_{ij} - N_i \cdot p_{ij})^2}{N_i \cdot p_{ij}},$$

where  $N_i$  is the total number of trials for the respective SSD and no-signal conditions for a given subject. The figures showing best fits to the data include all SSDs in the inhibition functions but only plot cumulative distribution functions for SSDs that include 40 or more trials, following Boucher et al. (2007).

A simplex method (Nelder & Mead, 1965) was used to find the global minima. The model was first fitted to the data of the no-signal condition to optimize the parameters of

the go unit (see Table 1) except for the lateral inhibition parameter ( $\beta_{go}$ ), because the parameter had no effect in the no-signal trials throughout which the stop unit is inactive. One hundred random positions were used as the initial position of the simplex. Next, these go parameters were fixed at the best values, and the remaining parameters were optimized by fitting to the data of the stop-signal trials, starting with 400 initial positions. Finally, all parameters were set free and the simplex was implemented, using the best parameter values for the go and stop units obtained in the preceding steps as the starting values.

### Alternative Fitting Methods

We considered two alternatives to our standard fitting method. Both alternatives evaluated the model fits with the Bayesian Information Criterion (BIC) as defined for multinomial data:

$$BIC = -2 \sum_i \sum_j N_i p_{ij} \cdot \ln(\pi_{ij}) + m \cdot \ln \left( \sum_i N_i \right) \quad (A1)$$

where  $p_{ij}$  and  $\pi_{ij}$  are the proportions of predicted and observed RTs, respectively, for quintile bin  $j$  ( $i$  indexes all SSD and no-signal conditions and  $j$  indexes quintile bins and signal-inhibit trials),  $N_i$  is the number of trials in condition  $i$ , and  $m$  is the number of free parameters in the model (see Leite & Ratcliff, 2010; Wasserman, 2000).

First, we replicated our standard fitting method, minimizing the deviance term in Equation A1 instead of  $\chi^2$ . As before, we fit the no-signal data by themselves to generate starting values for the go unit. Then we fixed those values and fit the whole data set, including stop-signal trials, to generate starting values for the stop unit. Then we fit the whole data set using the starting values obtained in the first two steps. The best-fitting parameter values and BIC measures for fits of interactive race 1.0 and blocked input 1.0 for each monkey are presented in Table A1 along with the best-fitting parameter values and  $\chi^2$  measures from our original fits. The BIC measures led to the same conclusion as the  $\chi^2$  values: interactive race 1.0 fit about as well as blocked input 1.0. The best-fitting parameter values were very similar.

The best-fitting parameter values and BIC values for the baseline model, interactive race 2.0, blocked input 2.0, and boosted fixation 1.0 appear in Table A2. The BIC values did

not differ much between the models, ranging from 7256.9 (baseline model) to 7212.7 (blocked input 2.0) for monkey C, and from 7655.7 (blocked input 2.0) to 7597.6 (baseline model) for monkey A. The best-fitting parameter values were generally similar to those from the  $\chi^2$  fits, although there were some tradeoffs:  $\mu_{move}$ ,  $\mu_{fix}$ ,  $D_{move}$ ,  $D_{fix}$ , and  $\theta$  tended to be smaller in the BIC fits than in the  $\chi^2$  fits, suggesting a slower growth of activation that began earlier and reached a lower threshold for both go and stop units.

Second, we used a genetic algorithm (Haupt & Haupt, 2004) to generate starting values instead of following our usual procedure, using BIC as the fit statistic. We used a three-step procedure: First, we ran a search with 200 simulations per condition with three colonies and a population size of 30. Second, we ran a search with 1500 simulations per condition, with one colony and a population size of 30. One set of starting values was the best-fitting parameters from the first search. The other 29 were generated randomly. Third, we used the best-fitting parameters from the second step for a final simplex fit. We ran these three steps 400 times for each model and each monkey.

The best-fitting parameter values and BIC measures for genetic algorithm fits of interactive race 1.0 and blocked input 1.0 for each monkey are presented in Table A1 along with the best-fitting parameter values and  $\chi^2$  measures from our original fits. The BIC values were about the same as the ones obtained with our usual procedure, suggesting there was no strong advantage in using the genetic algorithm. The genetic algorithm produced BIC values that were similar for the two models, supporting the conclusion of model mimicry. The parameter values differed from those from the other fits:  $D_{stop}$  decreased in interactive race 1.0,  $ND_{stop}$  decreased in blocked input 1.0, and  $\mu_{stop}$  and  $\sigma_{stop}$  increased in both models.

We repeated these procedures for fits of the constrained versions of the baseline model, interactive race 2.0, blocked input 2.0, and boosted fixation 1.0. Our initial runs of the genetic algorithm produced threshold values for the go unit that were much lower than the asymptotic activation in the fix unit during steady-state fixation (e.g., for baseline model,  $\theta = 5.66$  and  $\mu_{fix}/k_{fix} = 69$  for monkey C and  $\theta = 0.818$  and  $\mu_{fix}/k_{fix} = 149$  for monkey A; for interactive race 2.0,  $\theta = 0.090$  and  $\mu_{fix}/k_{fix} = 115$  for monkey C and  $\theta = 0.802$  and  $\mu_{fix}/k_{fix} = 90$  for monkey A), so RT was determined much more strongly by the release of

inhibition from fix units than by the growth of activation in move units. By contrast, go thresholds were very similar to asymptotic activation in the fix unit in the  $\chi^2$  fits (e.g., for baseline model,  $\theta = 46$  and  $\mu_{fix}/k_{fix} = 46$  for monkey C and  $\theta = 39$  and  $\mu_{fix}/k_{fix} = 43$  for monkey A; for interactive race 2.0,  $\theta = 46$  and  $\mu_{fix}/k_{fix} = 57$  for monkey C and  $\theta = 46$  and  $\mu_{fix}/k_{fix} = 46$  for monkey A), reflecting the similarity of firing rates of gaze-holding neurons during fixation and gaze-shifting neurons at threshold (Hanes et al., 1998; Paré & Hanes, 2003). Consequently, we re-ran the fits with the constraint that go thresholds could not be less than 50% of the steady-state activation in fix units (i.e.,  $\theta > 0.5 \times \mu_{fix}/k_{fix}$ ). The best-fitting parameter values and fit statistics are presented in Table A2.

The BIC values in the genetic algorithm fits were somewhat larger than the BIC values produced by our usual procedure, suggesting that the genetic algorithm did not find better parameter values. In the genetic algorithm fits, the BIC values did not vary much among the models, ranging from 7262.0 (baseline model) to 7240.1 (blocked input 2.0) for monkey C, and from 7691.3 (baseline model) to 7655.7 (blocked input 2.0) for monkey A. Several parameters changed substantially from the  $\chi^2$  fits:  $\beta_{move}$ ,  $\beta_{fix}$ ,  $k_{move}$ , and  $k_{fix}$  were larger, while  $D_{move}$ ,  $D_{fix}$ , and  $\theta$  were smaller.

These analyses led us to conclude that the fitting procedures did not have much effect on the conclusions we could draw from the fits. The procedures that used BIC did not distinguish as clearly among the models. We preferred to use  $\chi^2$  because it facilitates comparisons with Boucher et al. (2007), who also used it, and it allows us to test the significance of differences in goodness of fit for nested models.

### **Appendix B: Alternative Interactive Race Models**

Here, we evaluate alternative versions of the interactive race model to explore the importance of the ancillary assumptions Boucher et al. (2007) made in fitting the model. The first assumptions we consider are that threshold was fixed at 1,000 and the assumption that noise should be a free parameter. In most fits of stochastic accumulator models, threshold is a free parameter (to account for the speed-accuracy tradeoff) and noise is fixed at some arbitrary value to fix the scale for the other parameters. We fit the interactive race model to the data from monkeys A and C, allowing threshold to vary freely

and fixing noise ( $SD_{go} = 1$ ). The details of the modeling are presented in Appendix A. The values of the best-fitting parameters and measures of goodness of fit are presented in Table B1 along with the parameters and measures of goodness of fit for the original interactive race model. The fits of the new model were just as good as the fits of the original interactive race model. Consequently, we let threshold vary as a free parameter and fixed noise in all subsequent model fits.

Next, we assessed the importance of stochastic variability in the afferent stage of the stop process. Boucher et al. (2007) assumed that processes prior to the interactive stage of the stop process were constant in duration. We compared an interactive race models with deterministic (i.e., constant) and stochastic afferent stages. We assumed the afferent stage consists of a brief non-decision time of  $D_{stop}$  ms and a subsequent stochastic accumulation stage in which activation reflecting the presence of the stop signal increases with rate  $\mu_{stop}$  to a threshold  $\theta_{stop}$ . When the stop unit reaches threshold, the interactive stage begins and the stop process starts to inhibit the go unit.

Finally, we assessed the importance of the temporal dynamics of the activation in the interactive stage, comparing three different models, one in which activation grows from zero, as in Boucher et al. (2007), one in which activation is constant at the value it reaches when it reaches threshold, and one in which activation continues to grow after it reaches threshold.

In the *grow-from-zero* model, the activation in the stop unit begins at zero and grows at the same rate as stop activation grew before reaching threshold (i.e., at rate  $\mu_{stop}$ ). We evaluated models in which the rate of activation could be different before and after stop activation reached threshold and found no substantial improvement in goodness of fit. In the grow-from-zero model, stop activation grows over time according to Equation 2 and inhibits the go unit by an amount proportional to the amount by which stop activation exceeds threshold (i.e.,  $a_{stop}(t) - \theta_{stop}$ ). Thus, if  $a_{stop}(t) > \theta_{stop}$ , then Equation 1 becomes:

$$da_{go}(t) = \frac{dt}{\tau} \left[ \mu_{go} - \beta_{stop} \cdot (a_{stop}(t) - \theta_{stop}) \right] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{go}. \quad (\text{B1a})$$

If  $a_{stop}(t) < \theta_{stop}$ , then Equation 1 becomes:

$$da_{go}(t) = \frac{dt}{\tau} \mu_{go} + \sqrt{\frac{dt}{\tau}} \cdot \xi_{go}. \quad (\text{B1b})$$

This is essentially the same assumption about inhibition that was made in the original interactive race model. This model evaluates the importance of having a stochastic afferent stage in the stop unit. Note that  $k = 0$  so the leakage term drops out of Equations B1a and B1b.

In the *constant-at-threshold* model, the activation in the stop unit equals the amount of activation required to reach threshold (i.e.,  $\theta_{\text{stop}}$ ). In this model, the stop unit inhibits the go unit strongly and steadily as soon as it begins. Thus, when  $a_{\text{stop}}(t)$  reaches threshold, Equation 1 becomes:

$$da_{go}(t) = \frac{dt}{\tau} [\mu_{go} - \beta_{stop} \cdot \theta_{stop}] + \sqrt{\frac{dt}{\tau}} \cdot \xi_{go} \quad (\text{B2})$$

When  $a_{\text{stop}}(t)$  is below threshold, Equation B1b applies. This model evaluates the importance of having inhibition grow over time.

In the *grow-from-threshold* model, the activation in the stop unit continues to grow at rate  $\mu_{\text{stop}}$  after stop activation reaches threshold. In this model, the stop process inhibits the go unit strongly as soon as the interactive phase begins, and the inhibition continues to grow stronger over time. Equation B1b applies until  $a_{\text{stop}}(t)$  reaches threshold, and then Equation 1 applies. In this model,  $a_{\text{stop}}(t) = \theta_{\text{stop}}$  at the beginning of the interactive stage and it continues to grow throughout the duration of the interactive stage. This model evaluates the importance of having activation grow over time if it starts from a large value.

The values of the parameters that produced the best fits and measures of goodness of fit are presented in Table B1. All three models fit the data well and the fits were as good as the fits of the original interactive race model. Thus, the original assumption that  $D_{\text{stop}}$  was a constant does not seem to be necessary to produce good fits to the data. Models that include the more plausible assumption that there is stochastic variability in the afferent stage of the stop process fit just as well. Previous studies assuming SSRT is constant instead of variable showed that the assumption had little effect on estimates of SSRT if stop and go processes were independent (Band et al., 2003; De Jong, et al., 1990; Logan & Cowan, 1984; Verbruggen et al., 2013). The present models extend that conclusion to fits of inhibition functions and distributions of signal-respond and no-signal go RTs.

The three models made very different assumptions about the temporal dynamics of the inhibition in the interactive phase, but these assumptions did not result in substantial

differences in goodness of fit. Inhibition must be strong enough to prevent go activation from reaching threshold, but how its strength unfolds over time is not constrained very much by the data. In all three models, inhibition of the stop process on the go process depends on the  $\beta_{\text{stop}}$  parameter, which reflects the strength of the inhibitory connection between stop and go units. The inhibitory  $\beta_{\text{stop}}$  parameter trades off with the excitatory  $a_{\text{stop}}(t)$  parameter, so the amount of inhibition equals the product of these parameters. Because of this tradeoff, the value of  $\beta_{\text{stop}}$  parameter depends on the assumptions that are made about the temporal dynamics of inhibition.

### **Appendix C: Growth and Decay Rates on Signal-Inhibit Trials**

We calculated growth and decay rates for the neurophysiological data that were collected on the same trials as the behavioral data (see Hanes et al., 1995, and Hanes et al., 1998, for details of data acquisition, and Boucher et al., 2007, for the procedure for selecting trials for analysis). The neural data were converted to spike density functions (SDFs; see Thompson, Hanes, Bichot, & Schall, 1996) by convolving spike trains with a combination of rise and fall parameters that resembled a postsynaptic potential. Thus, the rate over time,  $R(t) = (1 - \exp[-t/\tau_r]) \cdot (\exp[-t/\tau_f])$ , where  $\tau_r$  is the rise parameter (set to 1 ms) and  $\tau_f$  is the fall parameter (set to 20 ms). For each monkey, for each cell, SDFs were normalized to the response threshold for no-signal trials for that cell. Response threshold was defined as the spike density 10 ms prior to the time at which the response occurred (i.e., RT – 10 ms) on no-signal trials. The threshold values were obtained for individual trials and averaged across all no-signal trials for the given cell. There were five cells for Monkey C and twelve cells for Monkey A. For Monkey C, there were 101 signal-inhibit trials at the 68-ms SSD, 65 at the 117-ms SSD, 24 at the 169-ms SSD, and 6 at the 217-ms SSD. We excluded the last two SSD trials due to the small number of trials. For Monkey A, there were 124 signal-inhibit trials at the 84-ms SSD, 61 ms at the 101-ms SSD, 130 at the 134-ms SSD, 93 at the 184-ms SSD, 18 at 201-ms SSD, and 21 at 234-ms SSD. We excluded last two SSDs due to the small number of trials.

The SDFs for signal-inhibit trials were normalized as proportions of the SDF for the same cell at threshold on no-signal trials. The rate of growth in movement cells is usually

calculated by estimating the onset of growth and time at which threshold is reached, and dividing the difference in activity by the difference in time (Hanes & Schall, 1996; Pouget et al., 2011; Purcell et al., 2010, 2012; Woodman, Kang, Thompson, & Schall, 2009). This measure was not practical for analyzing rates on signal-inhibit trials because SDFs did not reach threshold and because it was difficult to define an offset measure analogous to onset to calculate decay rate. Instead, we estimated rates in the middle of the growth and decay periods, focusing on the regions where SDF was between 25% and 75% of the range from its minimum value to its maximum value. Minima and maxima could be calculated easily, and both SDFs and model activation functions look linear in this range. We estimated slopes by dividing the difference in activation at 25% and 75% of the range by the difference in the times at which activation reached these points.

We computed the growth and decay rates of the normalized SDFs by bootstrapping 1000 samples at each SSD. Each sampling step involved the same number of trials as the number of available trials for the SSD, and SDFs were averaged across resampled trials. While SDFs for individual trials were very noisy, averaged SDFs were smooth and had a single peak. To determine the growth rate, the maximum and the minimum spike density of averaged SDF were computed with the constraint that the minimum spike density preceded the maximum spike density in time. Because spike density functions tend to be nonlinear near the minimum and maximum, the slope was computed between 25% and 75% of the range between the minimum and the maximum. Growth rate was defined as the difference between spike density at 75% and 25% of the range between maximum and minimum divided by the difference in time at which spike density reached 75% and 25% of its maximum. The decay rate was determined in a similar manner, except that the minimum was computed with the constraint that it followed the peak spike density in time. This procedure was repeated 1000 times for each SSD, and the resulting slopes were aggregated over SSDs.

The growth and decay rates for models' activation function were computed similarly. Each model was simulated the same number of trials as the number of available trials in each SSD condition, and the same computation was applied to derive the growth and decay slopes. These slopes were aggregated over SSDs.

## Appendix D: Cancel Times

### Cancel times in neural data

Cancel time is the difference between SSRT and the time at which movement-neuron activity on signal-inhibit trials diverges significantly from latency-matched movement neuron activity on no-signal trials. Hanes et al. (1998) defined divergence time as the time at which significant differential activity of a movement cell began during trials for which a saccade toward the target location was successfully canceled in response to a stop signal (i.e., signal-inhibit trials), as compared to the activity of that cell during *latency-matched* no-signal trials (no-signal trials for which the go process is sufficiently slow so that it could have been inhibited;  $RT > SSRT + SSD$ ). Operationally, divergence time is defined as the time point at which the difference between firing rates for signal-inhibit trials and latency-matched no-signal trials exceeds 2 standard deviations of the mean difference during the 600-ms interval before target presentation, given that the difference reaches 6 standard deviations and remains greater than 2 standard deviations in the next 50 ms (see Hanes et al., 1998, pp. 822-823). We used the cancel times that were calculated by Boucher et al. (2007) for each monkey, for each SSD in each neuron, using the same trials we used to calculate growth and decay rates. These were the same trials on which the behavioral data we fitted were collected. We present means across neurons and SSDs for each monkey.

### Cancel times in model simulations

We calculated mean cancel times for each model by performing the same analysis on activation functions produced by simulations of no-signal and signal-inhibit trials. We generated cancel times by simulating the model for 50 trials for each SSD condition and 150 trials for the no-signal condition. SSRT was computed for each SSD condition by using the integration method (see Logan, 1994; Verbruggen & Logan, 2009), and the latency-matched no-signal trials were obtained with the criterion  $RT > SSD + SSRT$ . If there were no latency-matched no-signal trials or signal-inhibit trials in any of the SSD conditions, the simulation was discarded and repeated.

Following Boucher et al.'s (2007) procedure, the difference between the mean simulated activation functions for the latency-matched no-signal trials and signal-inhibit

trials were computed at each millisecond, and the standard deviation from the target onset to SSD was computed. The divergence time was defined as the point at which the difference between the mean activation function exceeded 6 standard deviations and remained greater than that level in the next 50 ms. Cancel time was then computed by subtracting SSRT from the divergence time. This procedure was repeated 100 times.

### **Uniqueness of simulated cancel times**

To examine how unique the predicted cancel times were across model variation, we chose a target parameter ( $\beta_{stop}$  for interactive race models and  $k_{go}$  for blocked input models) that determined the shape of activation function for each model and systematically varied it to see whether variation could produce a range of cancel time distributions without affecting the goodness of fit to the behavioral data. We chose a set of target parameter values arbitrarily and fitted a model with the target parameter fixed at one value while letting other parameters in the model freely vary. The fits used 50 randomly chosen starting values in the vicinity of the best parameter values reported in Table 1. Cancel times were generated with the best parameters with the procedure described above. The model was also simulated 100 times, and chi-squares were computed for each simulation to examine whether the goodness of fit was significantly affected by the values of the target parameters. The mean cancel times varied considerably as the target parameters varied (see Figure C1) but the goodness of fit did not vary substantially (see Figure C2).

### **Author Notes**

Logan, Palmeri, and Schall are at Department of Psychology, Vanderbilt University, Nashville TN 37203. They are affiliated with the Vanderbilt Vision Research Center and the Vanderbilt Center for Integrative and Cognitive Neuroscience. Yamaguchi is now at the Department of Psychology, Edge Hill University, Ormskirk, UK. This research was supported primarily by grant number R01-EY021833 from the National Institutes of Health, and also by grant number BCS 0957074, and SMA 1041755 from the National Science Foundation and R01-MH55806, R01-EY008890, and P30-EY08126 from the National Institutes of Health and by Robin and Richard Patton through the E. Bronson Ingram Chair in Neuroscience. We are grateful to Paul Middlebrooks for help with the spike density functions. Data and models can be found at

<https://sites.google.com/site/cogyamaguchi/>

Address correspondence to Gordon D. Logan, Department of Psychology, Vanderbilt University, Nashville TN 37203. Electronic mail may be addressed to [gordon.logan@vanderbilt.edu](mailto:gordon.logan@vanderbilt.edu).

Table 1. Models of movement initiation. Best fitting parameter values and measures of goodness of fit for the interactive race model and the blocked input model for Monkeys C and A. Values in bold italics were fixed a priori.

	Interactive Race 1.0	Blocked Input 1.0	Interactive Race 1.0	Blocked Input 1.0
	<i>Monkey C</i>		<i>Monkey A</i>	
$\mu_{go}$	0.217	0.307	0.204	0.275
$\sigma_{go}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{go}$	0.007	-	0.002	-
$k_{go}$	-	0.003	-	0.003
$D_{go}$	<b>35</b>	<b>35</b>	<b>80</b>	<b>80</b>
$\mu_{stop}$	0.686	0.995	0.797	1.069
$\sigma_{stop}$	1.626	1.429	1.463	1.656
$\beta_{stop}$	0.703	-	0.758	-
$k_{stop}$	-	0.000	-	0.000
$ND_{stop}$	-	22.51	-	10.80
$D_{stop}$	84.32	-	68.33	-
$\theta$	41.93	47.74	38.99	39.54
<i>Afferent Time</i>	84 <sup>1</sup>	70 <sup>2</sup>	68 <sup>1</sup>	48 <sup>2</sup>
<i>Pearson's <math>\chi^2</math></i>	41.93	39.54	124.96	123.81
<i>Boucher et al. (2007) <math>\chi^2</math></i>	50.64		120.94	

Note: <sup>1</sup>Afferent time =  $D_{stop}$ . <sup>2</sup>Afferent time =  $ND_{stop} + \theta/\mu_{stop}$ .

Table 2. Constrained models of fixation and movement initiation. Best fitting parameter values and measures of goodness of fit for the baseline model, interactive race model, blocked input model, and the boosted fixation model for Monkeys C and A. In all model fits move and fix parameters were constrained to be the same (indicated by underline) unless the model assumes they are different. Values in bold italics were fixed a priori.

Monkey C

	Baseline	Interactive Race 2.0	Blocked Input 2.0	Boosted Fixation 1.0
<i>Monkey C</i>				
$\mu_{move}$	0.694	0.556	0.439	0.458
$\sigma_{move}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{move}$	0.006	0.005	0.004	0.005
$k_{move}$	0.015	0.014	0.009	0.009
$D_{move}$	44.92	37.84	44.77	37.15
$\mu_{fix}$	<u>0.694</u>	<u>0.556</u>	<u>0.439</u>	<u>0.458</u>
$\sigma_{fix}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{fix}$	<u>0.006</u>	0.011	<u>0.004</u>	<u>0.005</u>
$k_{fix}$	<u>0.015</u>	<u>0.014</u>	<u>0.009</u>	<u>0.009</u>
$D_{fix}$	<u>44.92</u>	<u>37.84</u>	<u>44.77</u>	<u>37.15</u>
$\theta$	45.83	38.84	36.68	39.21
$D_{control}$	-	-	88.29	72.90
$b$	-	-	-	3.71
Pearson's $\chi^2$	354.82	63.14	40.45	46.05

### Monkey A

	Baseline	Interactive Race 2.0	Blocked Input 2.0	Boosted Fixation 1.0
<i>Monkey A</i>				
$\mu_{move}$	1.261	0.826	0.362	0.365
$\sigma_{move}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{move}$	0.010	0.004	0.005	0.009
$k_{move}$	0.022	0.018	0.007	0.006
$D_{move}$	55.28	50.49	49.25	41.53
$\mu_{fix}$	<u>1.261</u>	<u>0.826</u>	<u>0.362</u>	<u>0.365</u>
$\sigma_{fix}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{fix}$	<u>0.010</u>	0.014	<u>0.005</u>	<u>0.009</u>
$k_{fix}$	<u>0.022</u>	<u>0.018</u>	<u>0.007</u>	<u>0.006</u>
$D_{fix}$	<u>55.28</u>	<u>50.49</u>	<u>49.25</u>	<u>41.53</u>
$\theta$	45.83	46.31	46.31	26.07
$D_{control}$	-	-	62.33	61.03
$b$	-	-	-	17.96
<i>Pearson's <math>\chi^2</math></i>	311.36	167.63	93.60	68.52

Table 3. Unconstrained models of fixation and movement initiation. Best fitting parameter values and measures of goodness of fit for the interactive race model, blocked input model, and the boosted fixation model for Monkeys C and A. In all model fits move and fix parameters were free to vary. Values in bold italics were fixed a priori.

Monkey C

	Interactive Race 2.0	Blocked Input 2.0	Boosted Fixation 1.0
$\mu_{move}$	0.585	0.417	0.354
$\sigma_{move}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{move}$	0.005	0.004	0.004
$k_{move}$	0.015	0.008	0.007
$D_{move}$	37.89	46.32	42.09
$\mu_{fix}$	0.679	0.331	0.425
$\sigma_{fix}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{fix}$	0.012	0.010	0.009
$k_{fix}$	0.015	0.008	0.008
$D_{fix}$	46.12	47.82	53.18
$\theta$	25.21	27.89	24.90
$D_{control}$	-	89.61	77.69
$b$	-	-	2.99
Pearson's $\chi^2$	53.05	35.08	37.79

Monkey A

	Interactive Race 2.0	Blocked Input 2.0	Boosted Fixation 1.0
$\mu_{move}$	0.613	0.391	0.391
$\sigma_{move}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{move}$	0.003	0.005	0.009
$k_{move}$	0.014	0.006	0.006
$D_{move}$	47.11	41.26	38.29
$\mu_{fix}$	0.625	0.323	0.324
$\sigma_{fix}$	<b>1.000</b>	<b>1.000</b>	<b>1.000</b>
$\beta_{fix}$	0.013	0.009	0.012
$k_{fix}$	0.013	0.006	0.006
$D_{fix}$	39.82	45.85	38.63
$\theta$	38.32	28.98	25.86
$D_{control}$	-	62.02	68.41
$b$	-	-	17.69
Pearson's $\chi^2$	113.15	62.02	64.43

Table A1: Comparing model fits generated with different fit statistics ( $\chi^2$  vs. Bayesian Information Criterion or BIC) and different fitting methods (our standard method vs. Genetic Algorithm or GA). Fixed parameters are in bold italics.

	Monkey C					
	Interactive Race 1.0			Blocked Input 1.0		
	<i>Pearson's</i> $\chi^2$	BIC	GA/BIC	<i>Pearson's</i> $\chi^2$	BIC	GA/BIC
$\mu_{go}$	0.217	0.241	0.229	0.307	0.310	0.299
$\sigma_{go}$	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
$\beta_{go}$	0.007	0.009	0.332	-	-	-
$k_{go}$	-	-	-	0.003	0.003	0.004
$D_{go}$	<b>35</b>	<b>35</b>	<b>35</b>	<b>35</b>	<b>35</b>	<b>35</b>
$\mu_{stop}$	0.686	0.775	2.261	0.995	1.032	8.046
$\sigma_{stop}$	1.626	1.578	9.89	1.429	1.048	8.33
$\beta_{stop}$	0.703	0.780	0.363	-	-	-
$k_{stop}$	-	-	-	0	0.001	0.064
$ND_{stop}$	-	-	-	22.51	18.85	47.59
$D_{stop}$	84.32	83.24	51.78	-	-	-
$\theta$	41.93	51.99	48.19	47.74	49.15	44.42
<i>Fit statistic</i>	41.93	7259.6	7227.8	39.54	7220.6	7302.7

Table A1 continued.

	Monkey A					
	Interactive Race 1.0			Blocked Input 1.0		
	<i>Pearson's</i> $\chi^2$	BIC	GA/BIC	<i>Pearson's</i> $\chi^2$	BIC	GA/BIC
$\mu_{go}$	0.204	0.217	0.202	0.275	0.260	0.333
$\sigma_{go}$	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
$\beta_{go}$	0.002	0.002	0.711	-	-	-
$k_{go}$	-	-	-	0.003	0.002	0.005
$D_{go}$	<b>80</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>80</b>	<b>80</b>
$\mu_{stop}$	0.797	0.973	9.398	1.069	1.468	8.887
$\sigma_{stop}$	1.463	1.722	10.009	1.656	1.383	1.126
$\beta_{stop}$	0.758	0.955	0.949	-	-	-
$k_{stop}$	-	-	-	0	0.001	0
$ND_{stop}$	-	-	-	10.8	11.77	48.83
$D_{stop}$	68.33	66.43	49.64	-	-	-
$\theta$	38.99	41.06	38.84	39.54	40.90	41.44
<i>Fit statistic</i>	124.96	7672.7	7666.9	123.81	7707.7	7707.0

Table A2: Comparing model fits generated with different fit statistics ( $\chi^2$  vs. Bayesian Information Criterion or BIC) and different fitting methods (our standard method vs. Genetic Algorithm or GA). Parameters constrained to be the same for move and fix units are in bold italics.

	Monkey C											
	Baseline			Interactive Race 2.0			Blocked Input 2.0			Boosted Fixation 1.0		
	<i>Pearson's</i>			<i>Pearson's</i>			<i>Pearson's</i>			<i>Pearson's</i>		
	$\chi^2$	BIC	BIC/GA	$\chi^2$	BIC	BIC/GA	$\chi^2$	BIC	BIC/GA	$\chi^2$	BIC	BIC/GA
$\mu_{move}$	0.694	0.426	0.582	0.556	0.417	0.573	0.439	0.368	0.36	0.458	0.356	0.499
$\sigma_{move}$	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
$\beta_{move}$	0.006	0.01	0.342	0.005	0.006	0.503	0.004	0	0.271	0.005	0.008	0.416
$k_{move}$	0.015	0.005	0.015	0.013	0.009	0.014	0.009	0.003	0.02	0.009	0.007	0.024
$D_{move}$	44.92	17.79	18.9	37.84	42.31	2.2	44.77	10.22	77.21	<b>37.15</b>	54.92	65.29
$\mu_{fix}$	<b>0.694</b>	<b>0.426</b>	<b>0.582</b>	<b>0.556</b>	<b>0.417</b>	<b>0.573</b>	<b>0.439</b>	<b>0.368</b>	<b>0.36</b>	<b>0.458</b>	<b>0.356</b>	<b>0.499</b>
$\sigma_{fix}$	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
$\beta_{fix}$	<b>0.006</b>	<b>0.01</b>	<b>0.342</b>	0.011	0.017	0.307	<b>0.004</b>	<b>0</b>	<b>0.271</b>	<b>0.005</b>	<b>0.008</b>	<b>0.416</b>
$k_{fix}$	<b>0.015</b>	<b>0.005</b>	<b>0.015</b>	<b>0.013</b>	<b>0.009</b>	<b>0.014</b>	<b>0.009</b>	<b>0.003</b>	<b>0.02</b>	<b>0.009</b>	<b>0.007</b>	<b>0.024</b>
$D_{fix}$	<b>44.92</b>	<b>17.79</b>	<b>18.9</b>	<b>37.84</b>	<b>42.31</b>	<b>2.2</b>	<b>44.77</b>	<b>10.22</b>	<b>77.21</b>	<b>37.15</b>	<b>54.92</b>	<b>65.29</b>
$\theta$	45.92	20.11	25.12	38.84	19.69	28.53	36.68	57.56	16.26	42.05	23.04	20.78
$D_{control}$	-	-	-	-	-	-	88.29	77.65	64.43	78.22	70.35	24.38
$b$	-	-	-	-	-	-	-	-	-	3.71	2.87	12.27
<i>Fit statistic</i>	354.82	7256.9	7262	63.14	7227.3	7266.1	40.45	7212.7	7240.1	46.05	7224.2	7250.9

Table A2 continued.

	Monkey A											
	Baseline			Interactive Race 2.0			Blocked Input 2.0			Boosted Fixation 1.0		
	<i>Pearson's</i>			<i>Pearson's</i>			<i>Pearson's</i>			<i>Pearson's</i>		
	$\chi^2$	BIC	BIC/GA	$\chi^2$	BIC	BIC/GA	$\chi^2$	BIC	BIC/GA	$\chi^2$	BIC	BIC/GA
$\mu_{move}$	1.261	0.596	0.927	0.826	0.509	0.575	0.362	0.415	0.426	0.365	0.435	0.484
$\sigma_{move}$	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
$\beta_{move}$	0.01	0.015	1.000	0.004	0.001	0.039	0.005	0.004	0.096	0.009	0.007	0.137
$k_{move}$	0.022	0.004	0.015	0.018	0.009	0.012	0.007	0.005	0.014	0.006	0.005	0.015
$D_{move}$	55.28	33.93	0.00	50.49	24.15	0.00	49.25	36.86	62.48	41.53	60.71	36.44
$\mu_{fix}$	<b>1.261</b>	<b>0.596</b>	<b>0.927</b>	<b>0.826</b>	<b>0.509</b>	<b>0.575</b>	<b>0.362</b>	<b>0.415</b>	<b>0.426</b>	<b>0.365</b>	<b>0.435</b>	<b>0.484</b>
$\sigma_{fix}$	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>	<b>1</b>
$\beta_{fix}$	<b>0.01</b>	<b>0.015</b>	<b>1</b>	<b>0.014</b>	0.013	0.101	<b>0.005</b>	<b>0.004</b>	<b>0.096</b>	<b>0.009</b>	<b>0.007</b>	<b>0.137</b>
$k_{fix}$	<b>0.022</b>	<b>0.004</b>	<b>0.015</b>	<b>0.018</b>	<b>0.009</b>	<b>0.012</b>	<b>0.007</b>	<b>0.005</b>	<b>0.014</b>	<b>0.006</b>	<b>0.005</b>	<b>0.015</b>
$D_{fix}$	<b>55.28</b>	<b>33.93</b>	<b>0</b>	<b>50.49</b>	<b>24.15</b>	<b>0</b>	<b>49.25</b>	<b>36.86</b>	<b>62.48</b>	<b>41.53</b>	<b>60.71</b>	<b>36.44</b>
$\theta$	<b>45.83</b>	4.21	40.66	46.31	24.15	30.76	46.31	38.48	22.55	26.07	19.58	27.59
$D_{control}$	-	-	-	-	-	-	62.33	61.62	0.00	61.03	74.72	62.15
$b$	-	-	-	-	-	-	-	-	-	17.96	22.88	43.94
<i>Fit statistic</i>	311.36	7597.6	7691.3	167.63	7630.7	7675.9	93.6	7607	7655.7	68.52	7611.1	7664.2

Table B1. Best parameters and goodness of fit

Model	$\mu_{go}$	$\sigma_{go}$	$\beta_{go}$	$\kappa_{go}$	$\mu_{stop}$	$\sigma_{stop}$	$\beta_{stop}$	$\kappa_{stop}$	$D_{stop}$	$\theta$	$\chi^2$
Threshold fixed:					Monkey C						
<i>Independent Race</i>	4.64	20.26	-	-	17.67	15.58	-	-	29	<b>1000</b>	57.24
<i>Interactive Race</i>	4.63	20.43	0.010	-	4.62	20.41	0.434	-	67	<b>1000</b>	50.64
Variance fixed:											
<i>Independent Race</i>	0.230	<b>1.000</b>	-	-	0.897	2.073	-	-	29.08	48.80	43.87
<i>Interactive Race</i>	0.225	<b>1.000</b>	0.010	-	4.896	1.172	0.651	-	77.23	47.50	48.61
<i>Grow from zero</i>	0.222	<b>1.000</b>	0.010	-	0.590	1.280	0.700	-	0.00	47.87	43.67
<i>Constant at threshold</i>	0.228	<b>1.000</b>	-	-	0.919	1.461	0.525	-	29.17	48.39	41.73
<i>Grow from threshold</i>	0.226	<b>1.000</b>	0.017	-	0.564	1.438	0.465	-	0.00	47.83	46.48
<i>Blocked Input stochastic</i>	0.307	<b>1.000</b>	-	0.003	0.995	1.429	-	0.000	22.51	47.74	39.54
<i>Blocked input constant</i>	0.297	<b>1.000</b>	-	0.003	-	-	-	-	73.49	48.43	47.99
Threshold fixed:					Monkey A						
<i>Independent Race</i>	5.09	26.38	-	-	50.24	40.17	-	-	51	<b>1000</b>	128.80
<i>Interactive Race</i>	5.08	26.24	0.005	-	5.07	26.34	0.111	-	51	<b>1000</b>	120.94
Variance fixed:											
<i>Independent Race</i>	0.217	<b>1.000</b>	-	-	0.684	2.085	-	-	10.43	40.86	126.98
<i>Interactive Race</i>	0.214	<b>1.000</b>	0.002	-	0.752	1.674	0.776	-	70.28	40.37	133.44
<i>Grow from zero</i>	0.211	<b>1.000</b>	0.008	-	0.668	2.021	0.763	-	9.98	40.06	117.43
<i>Constant at threshold</i>	0.219	<b>1.000</b>	-	-	0.738	1.586	0.715	-	12.46	40.14	130.25
<i>Grow from threshold</i>	0.220	<b>1.000</b>	0.009	-	0.629	1.760	0.556	-	0.00	41.00	131.12
<i>Blocked input stochastic</i>	0.275	<b>1.000</b>	-	0.003	1.069	1.656	-	0.000	10.8	39.54	123.81
<i>Blocked input constant</i>	0.252	<b>1.000</b>	-	0.004	-	-	-	-	49.25	36.47	130.23

Note: Fixed parameters in bold italics.

Figure 1. Inhibition functions (first and third panels) and the cumulative distribution functions for no-signal and signal-respond trials (second and fourth panels) for monkeys C and A. No-signal RTs are plotted in black. Signal-respond RTs are plotted in gray with lighter shades for shorter stop signal delays (SSD).

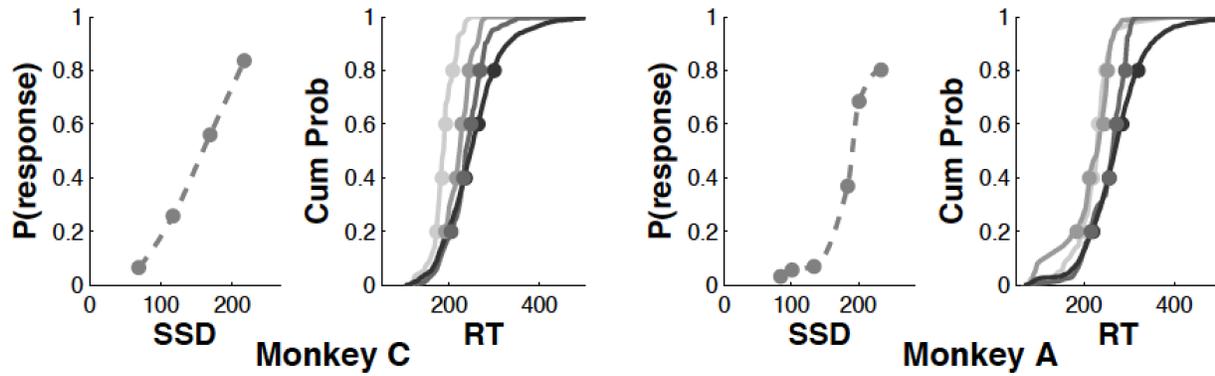
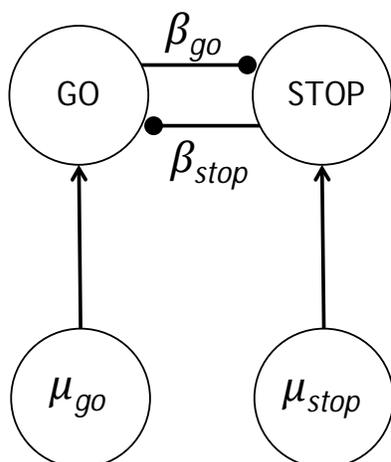


Figure 2. Interactive Race Model 1.0 of movement initiation: Architecture (Panel A) and timing of events on a trial (Panel B).

## Interactive Race 1.0

### A. Architecture



### B. Timing

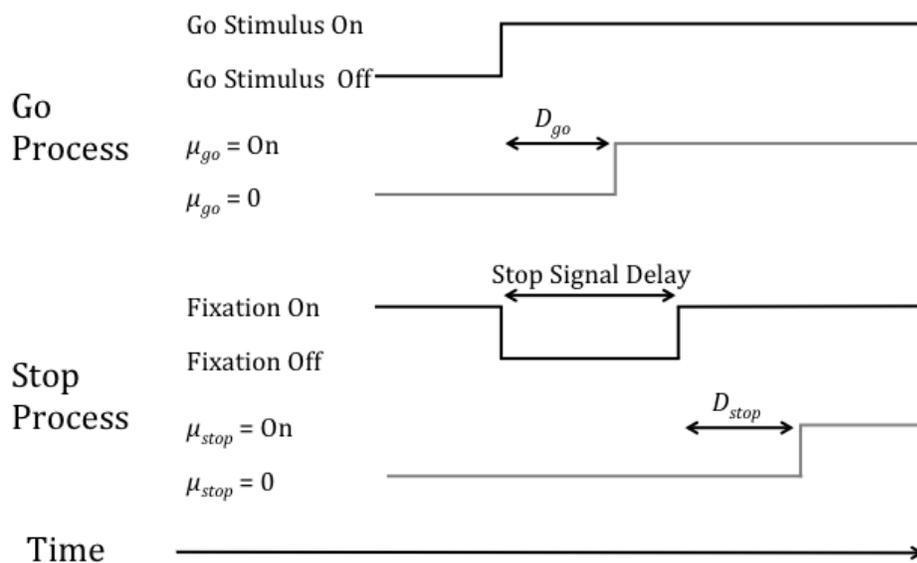


Figure 3. The observed and predicted inhibition function (first and third panels) and the cumulative distribution functions for no-signal and signal-response trials (second and fourth panels) for monkeys C and A. No-signal RTs are plotted in black. Signal-response RTs are plotted in gray with lighter shades for shorter stop signal delays (SSD). The predictions are from models of movement initiation. The interactive race model predictions are in the top panels. The blocked-input model predictions are in the bottom panels.

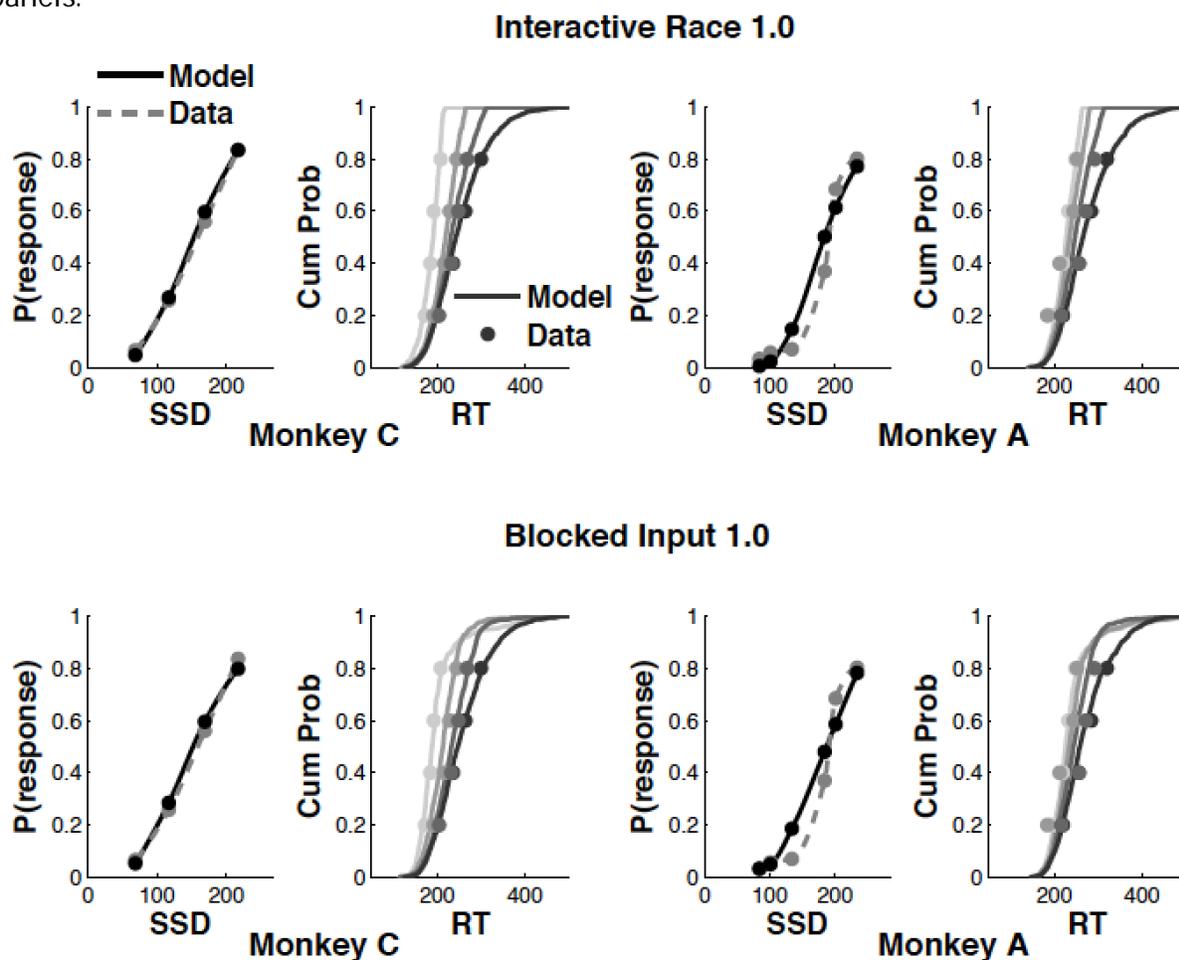


Figure 4. Predicted activation of go and stop units over time for interactive race model 1.0 and blocked input model 1.0 for monkey C and monkey A. Both models address movement initiation. For each model, the left panel shows activation of the go unit and the right panel shows activation of the stop unit. The horizontal line represents the threshold, estimated from model fits. The left-most vertical line represents stop-signal onset. The right-most vertical line represents stop signal reaction time estimated from the simulated data using the independent race model. The dashed vertical line represents mean divergence time. Cancel time is divergence time minus stop signal reaction time.

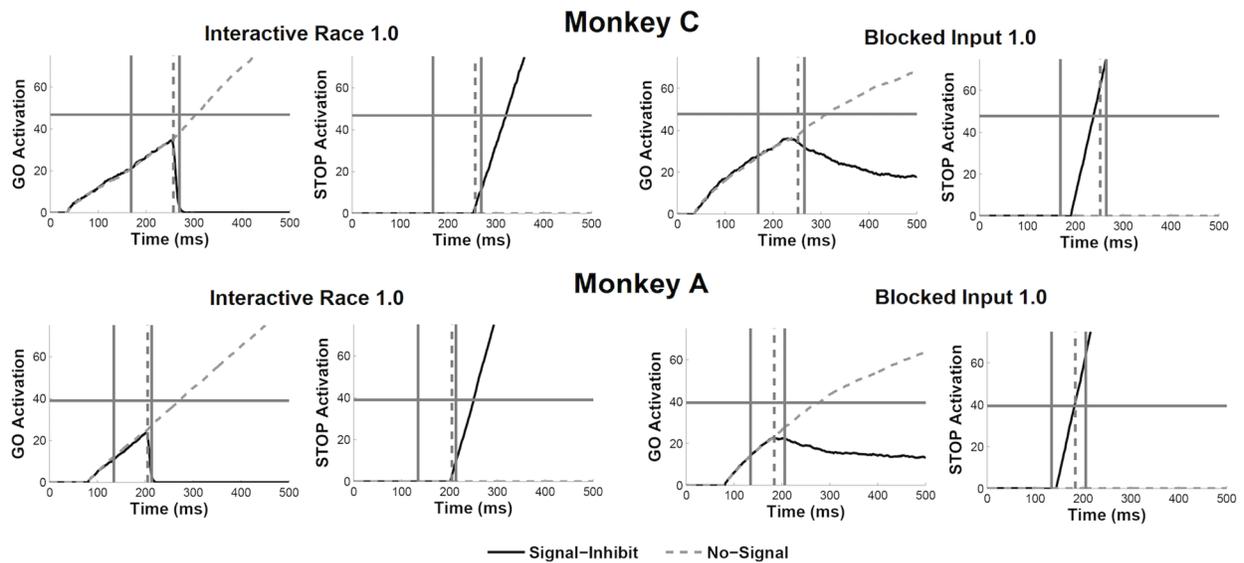
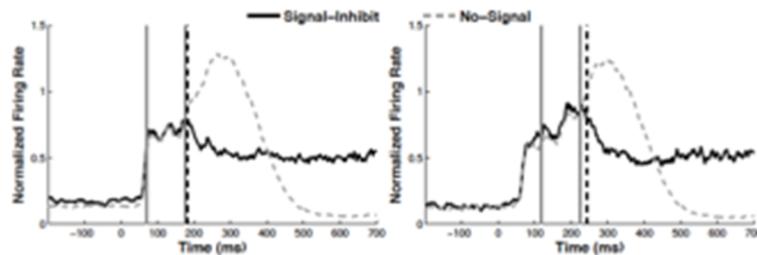


Figure 5. Mean normalized firing rate for movement-related neurons in frontal eye fields of monkeys C (top two) and A (bottom four) on signal inhibit trials (solid dark line) and latency-matched no-signal trials (dashed grey line). Each panel depicts a different stop signal delay (68 and 117 ms for monkey C; 84, 101, 134, and 184 ms for monkey A). The vertical grey line on the left represents the onset of the stop signal. The vertical grey line on the right represents stop signal reaction time estimated from the independent race model. The dashed vertical line represents the average time at which signal-inhibit and no-signal functions diverged significantly (mean divergence time). Mean cancel time is mean divergence time minus stop signal reaction time.

### Monkey C



### Monkey A

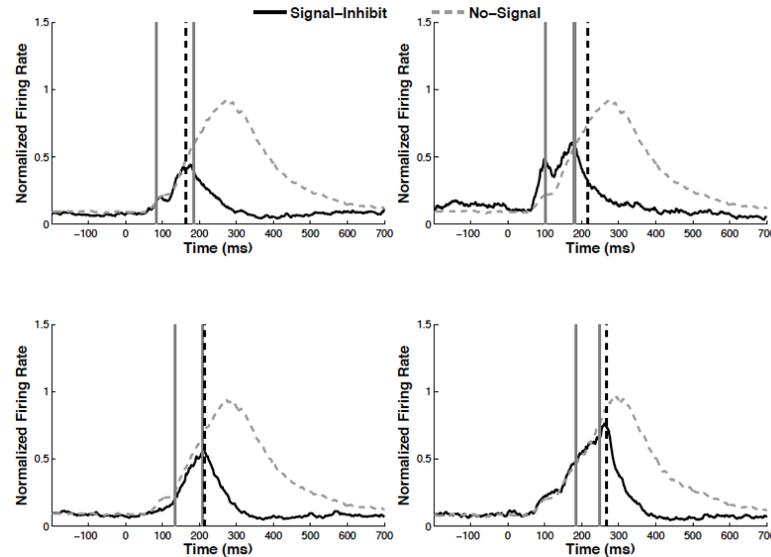


Figure 6. Mean growth and decay rates for observed and predicted activation functions for monkeys C and A. Observed growth and decay rates are estimates from neural activity that was recorded from frontal eye fields (FEF) while the behavior that was modeled were gathered. Predicted growth and decay are from interactive race model 1.0 and blocked input model 1.0. Error bars are bootstrapped 95% confidence intervals of the observed data.

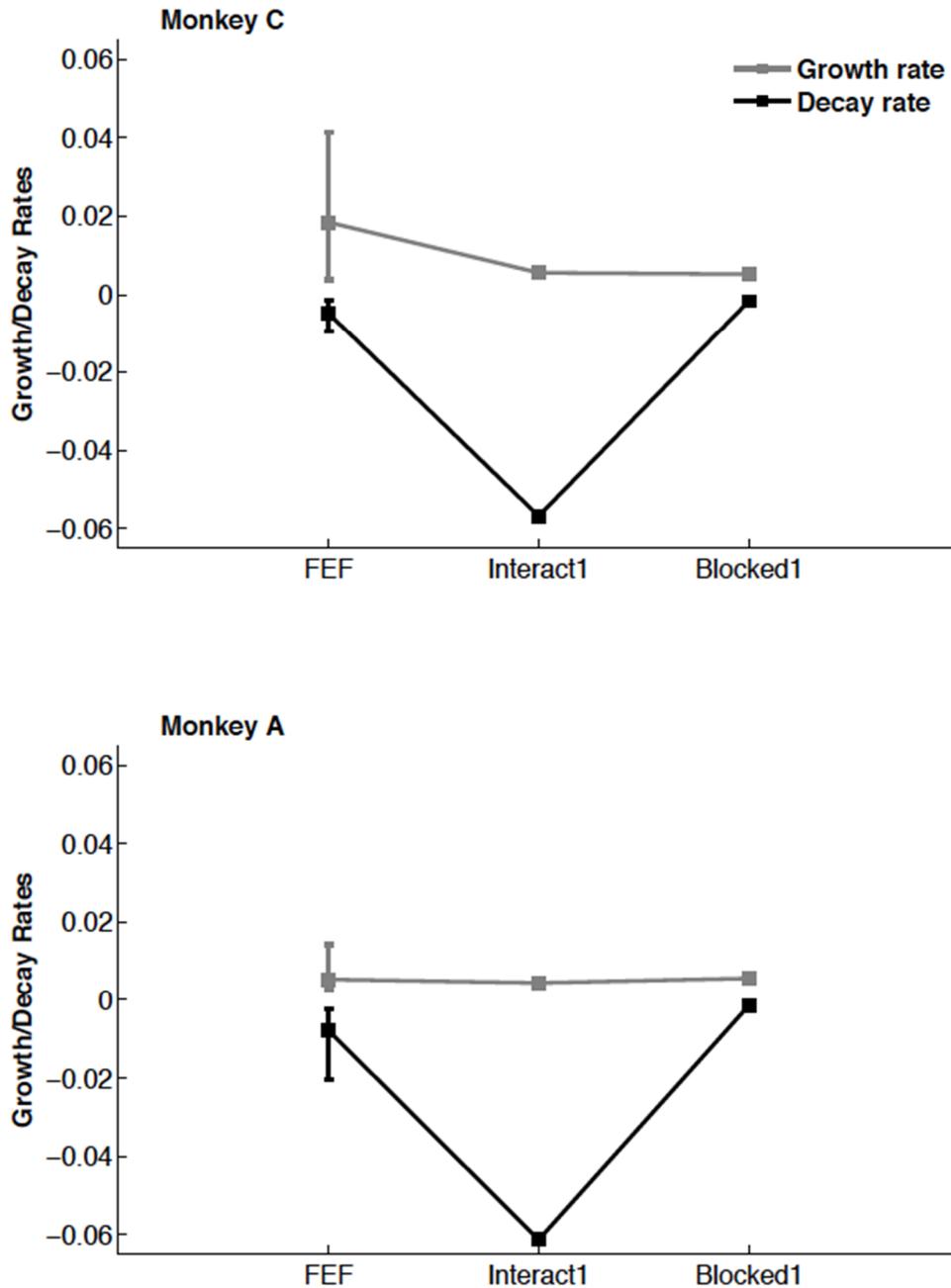


Figure 7. Mean observed and predicted cancel times for monkeys C and A. Observed cancel times are estimates from neural activity that was recorded from frontal eye fields (FEF) while the behavior that was modeled were gathered. Predicted cancel times are from interactive race model 1.0 and blocked input model 1.0.

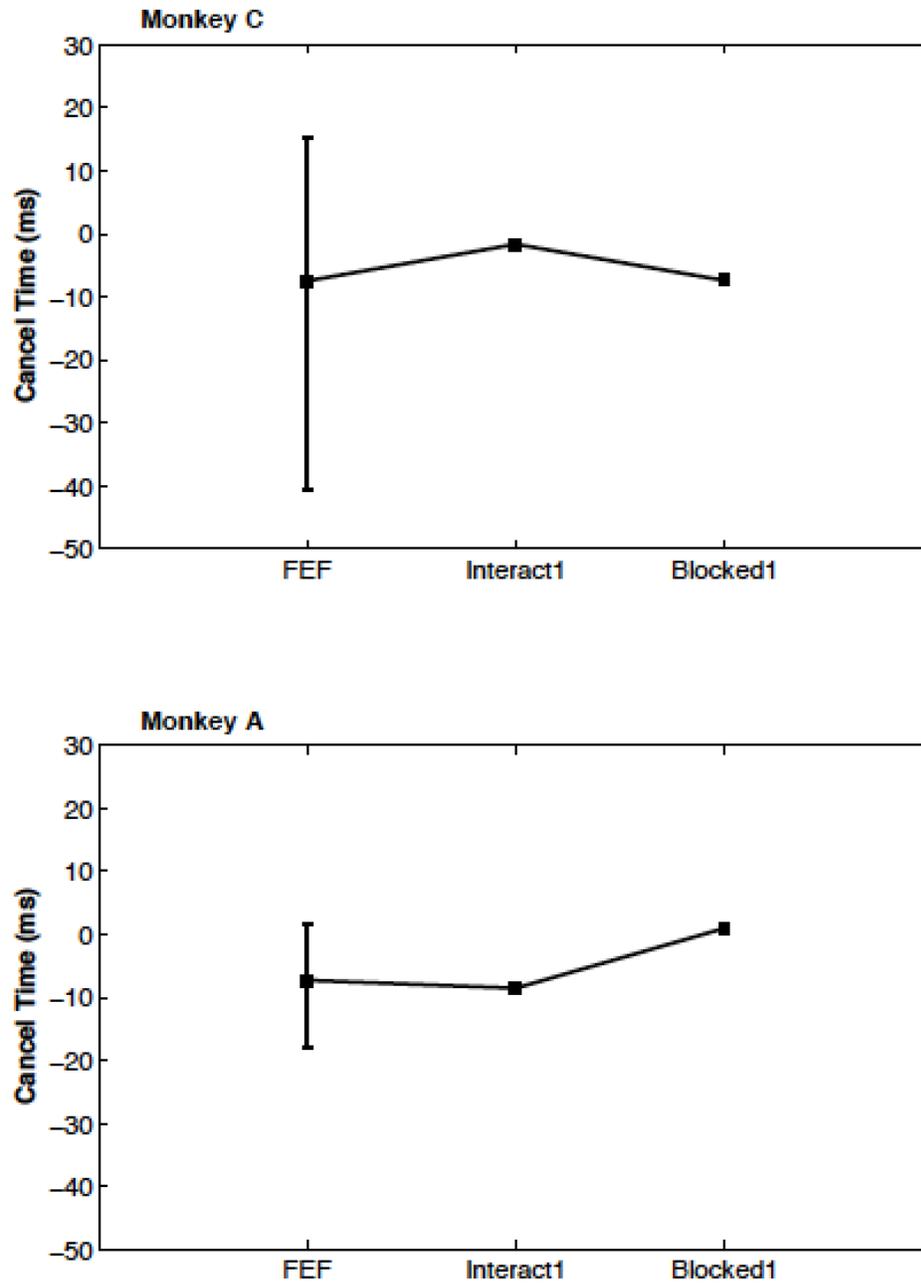
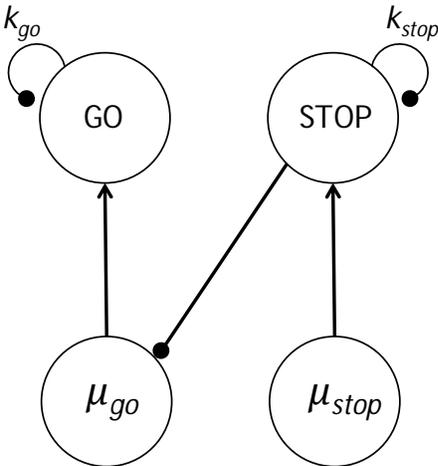


Figure 8. Blocked input model 1.0 of movement of initiation: Architecture (Panel A) and timing of events on a trial (Panel B).

## Blocked Input 1.0

### A. Architecture



### B. Timing

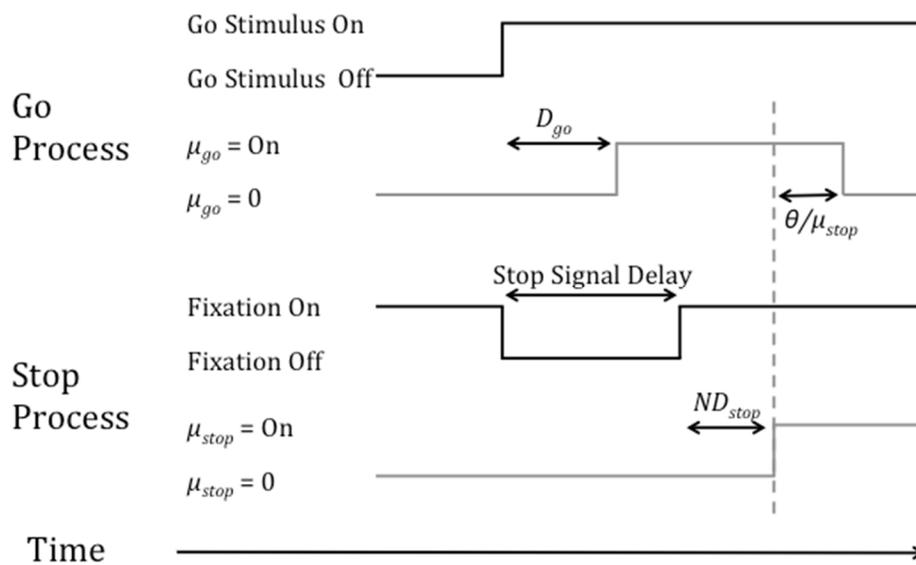
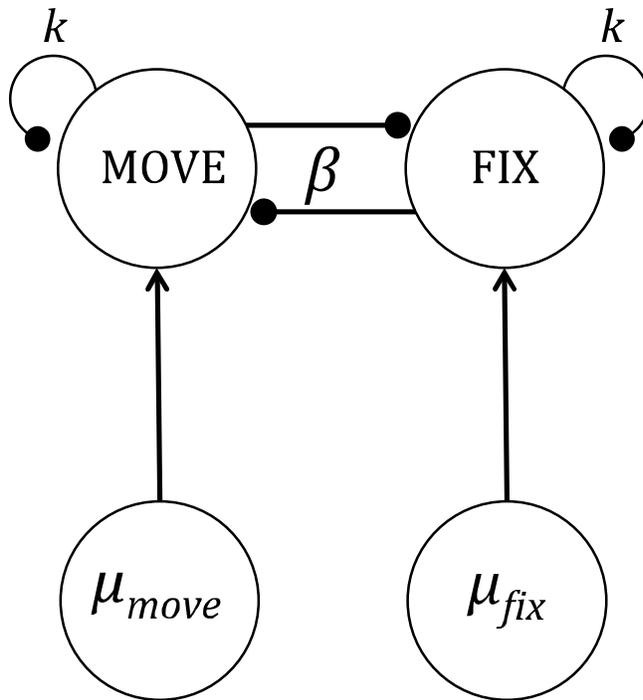
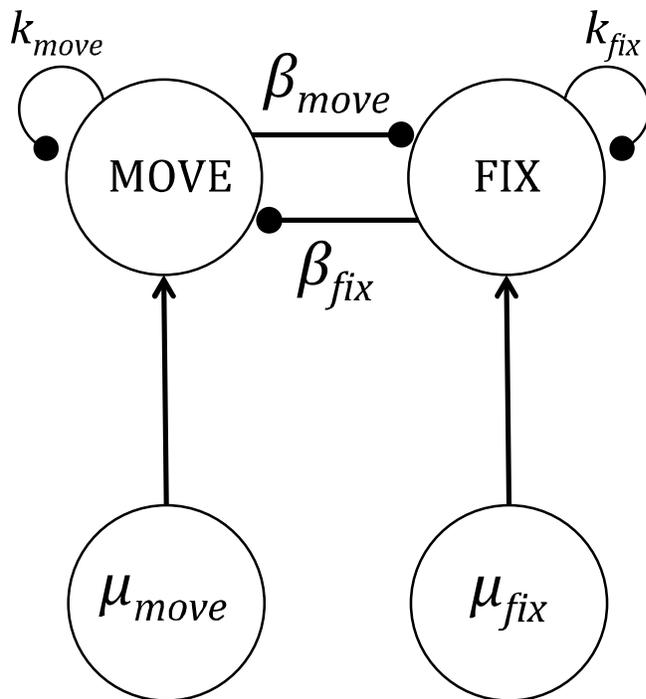


Figure 9. Architectures of four models that address fixation and movement initiation (Panel A: Baseline Model; Panel B: Interactive Race 2.0; Panel C: Blocked Input 2.0; Panel D: Boosted Fixation 1.0).

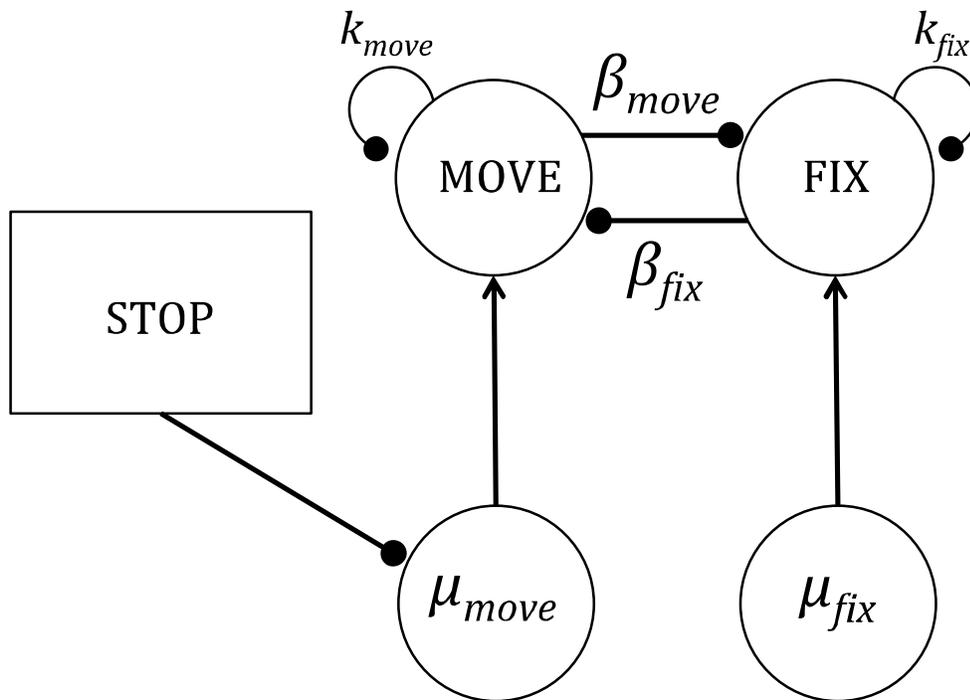
### A. Baseline (Symmetrical Inhibition) Model



## B. Interactive Race Model 2.0



## C. Blocked Input Model 2.0



## D. Boosted Fixation Model 1.0

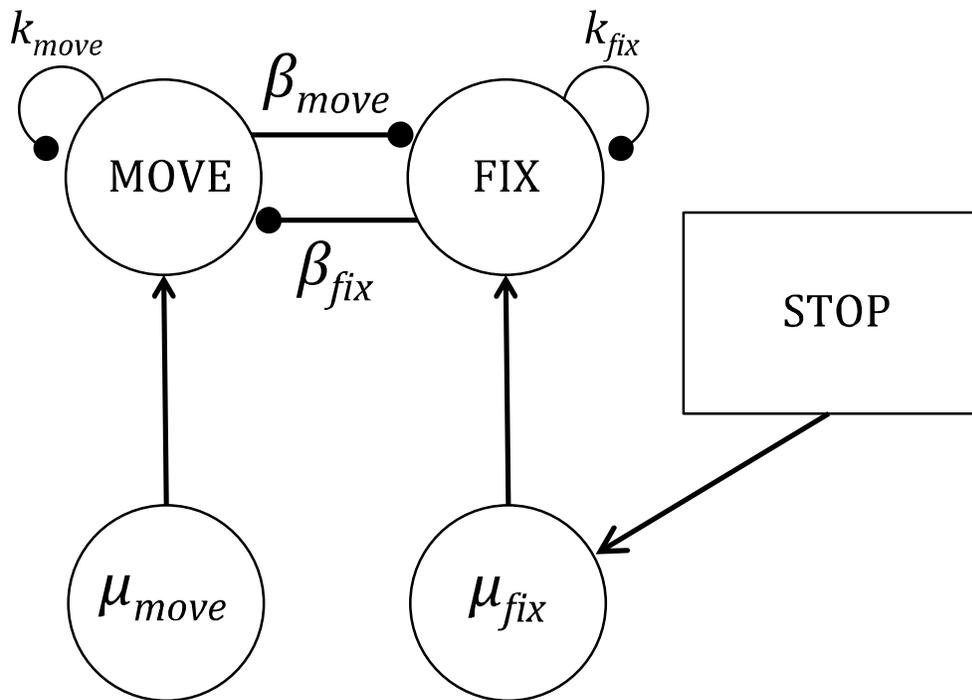
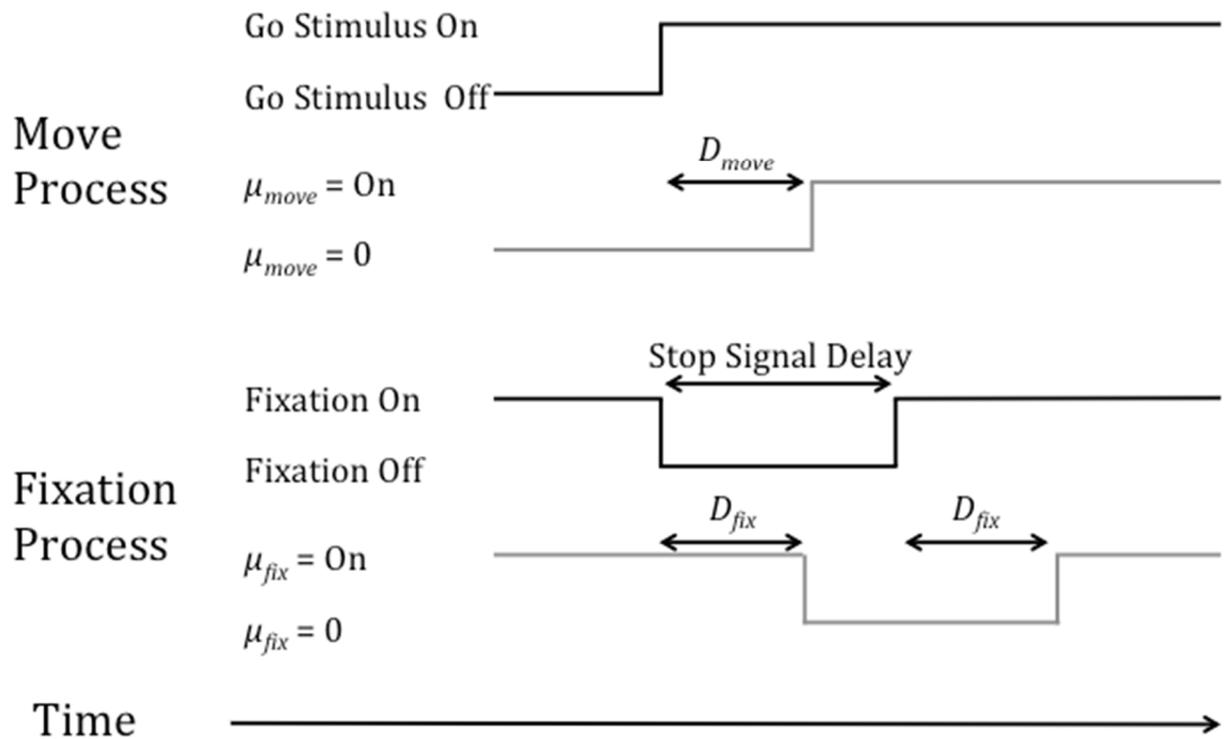
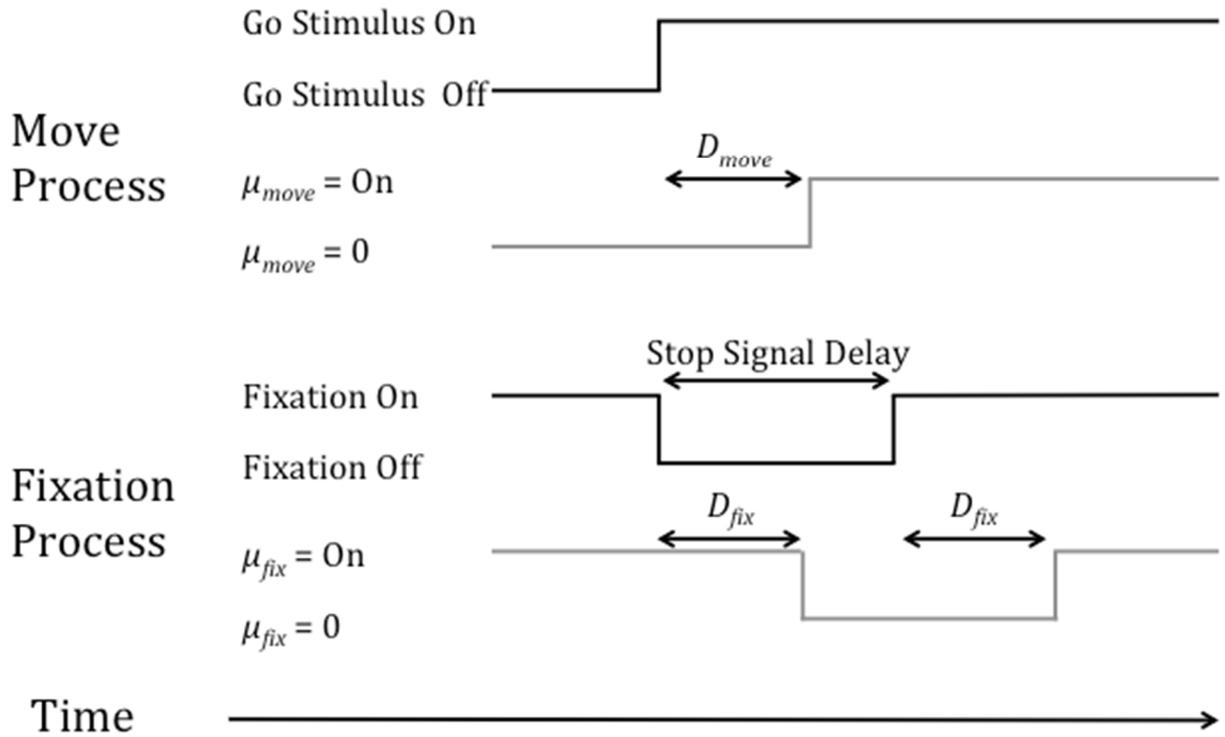


Figure 10. Timing diagrams for move and fixation processes for the baseline model (panel A), interactive race model 2.0 (panel B), blocked input model 2.0 (panel C), and boosted fixation model 1.0 (panel D). All models address fixation and movement initiation.

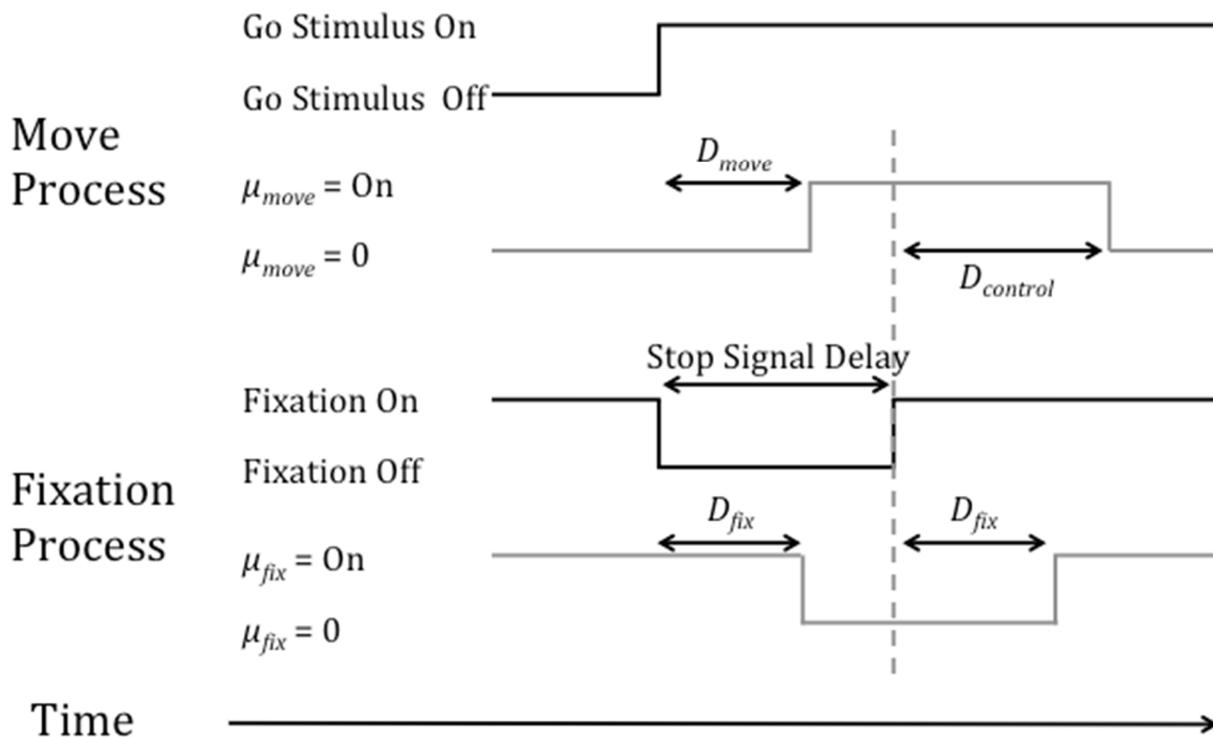
### A. Baseline Model: $\beta_{stop} = \beta_{go}$



## B. Interactive Race Model 2.0: $\beta_{stop} > \beta_{go}$



### C. Blocked Input Model 2.0



## D. Boosted Fixation Model 1.0

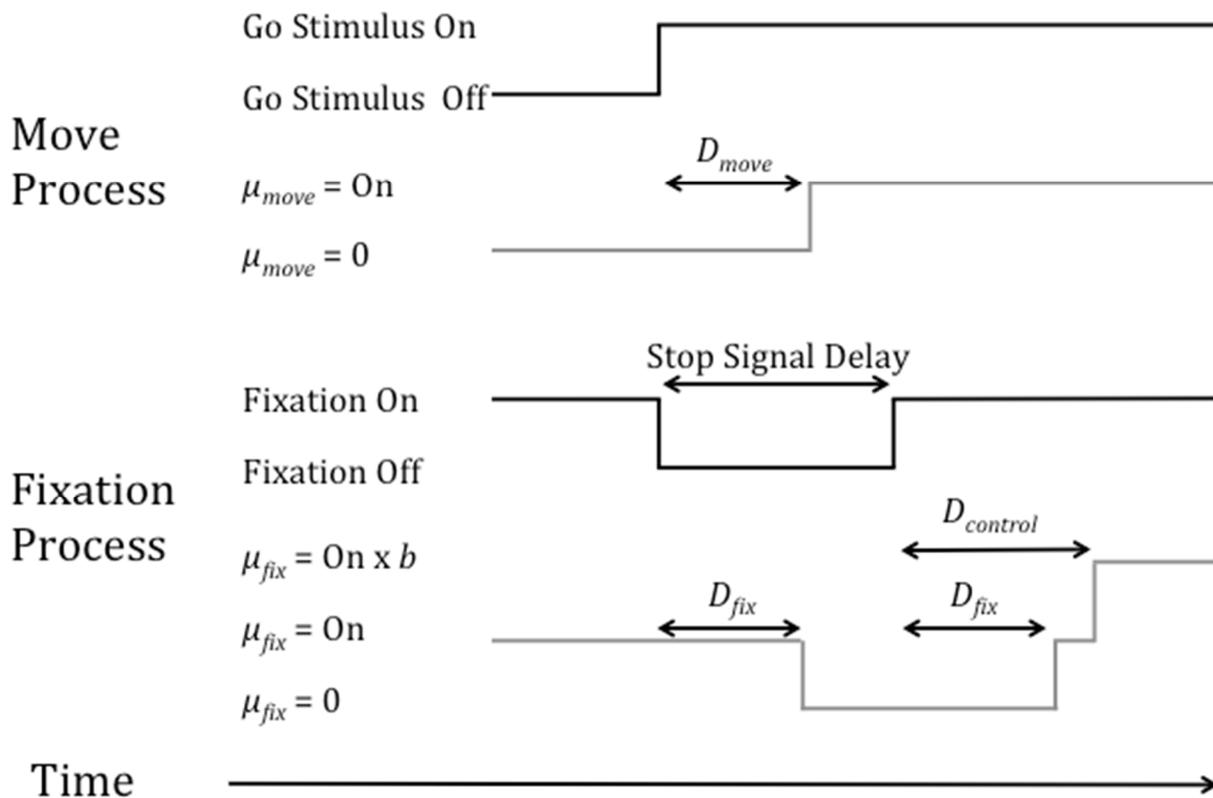


Figure 11. Observed and predicted inhibition functions and cumulative response time distributions for no-signal and signal-respond trials for monkey C. No-signal RTs are plotted in black. Signal-respond RTs are plotted in gray with lighter shades for shorter stop signal delays (SSD). Predicted values are generated from the constrained fit of the baseline model, and the unconstrained fits of interactive race model 2.0, blocked input model 2.0, and boosted fixation model 1.0. All models address fixation and movement initiation.

### Monkey C

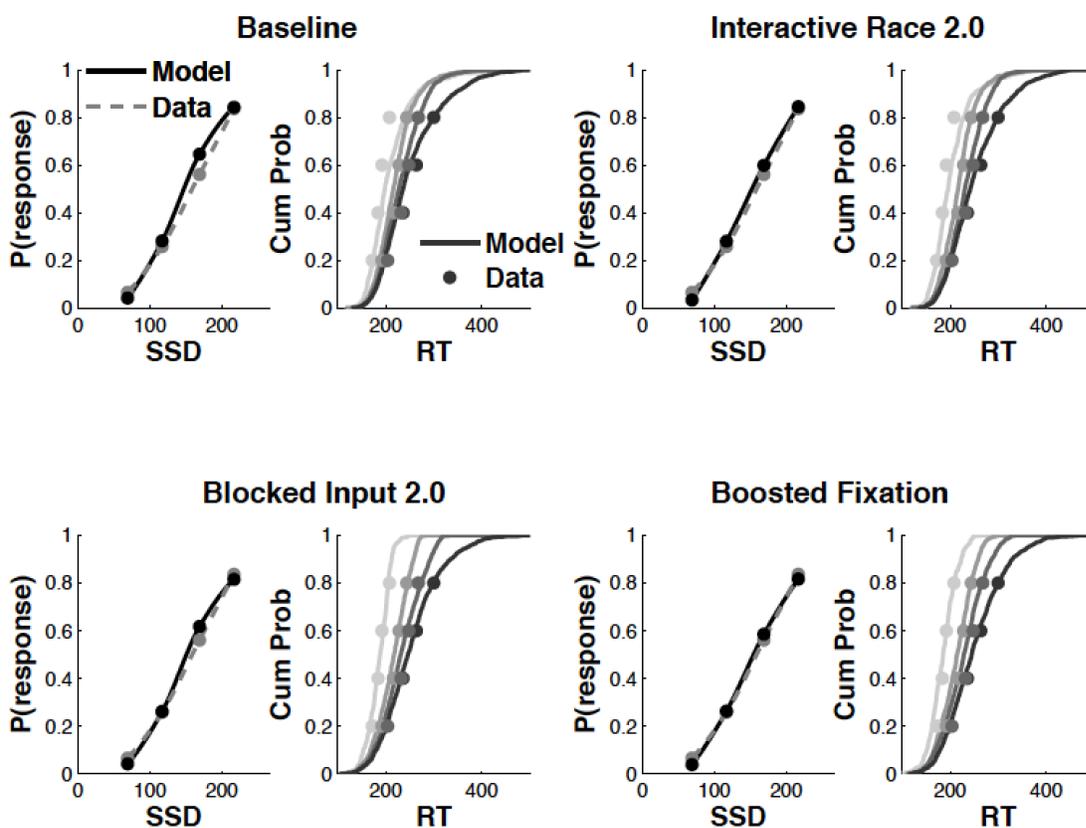


Figure 12. Observed and predicted inhibition functions and cumulative response time distributions for no-signal and signal-respond trials for monkey A. No-signal RTs are plotted in black. Signal-respond RTs are plotted in gray with lighter shades for shorter stop signal delays (SSD). Predicted values are generated from the constrained fit of the baseline model, and the unconstrained fits of interactive race model 2.0, blocked input model 2.0, and boosted fixation model 1.0. All models address fixation and movement initiation.

### Monkey A

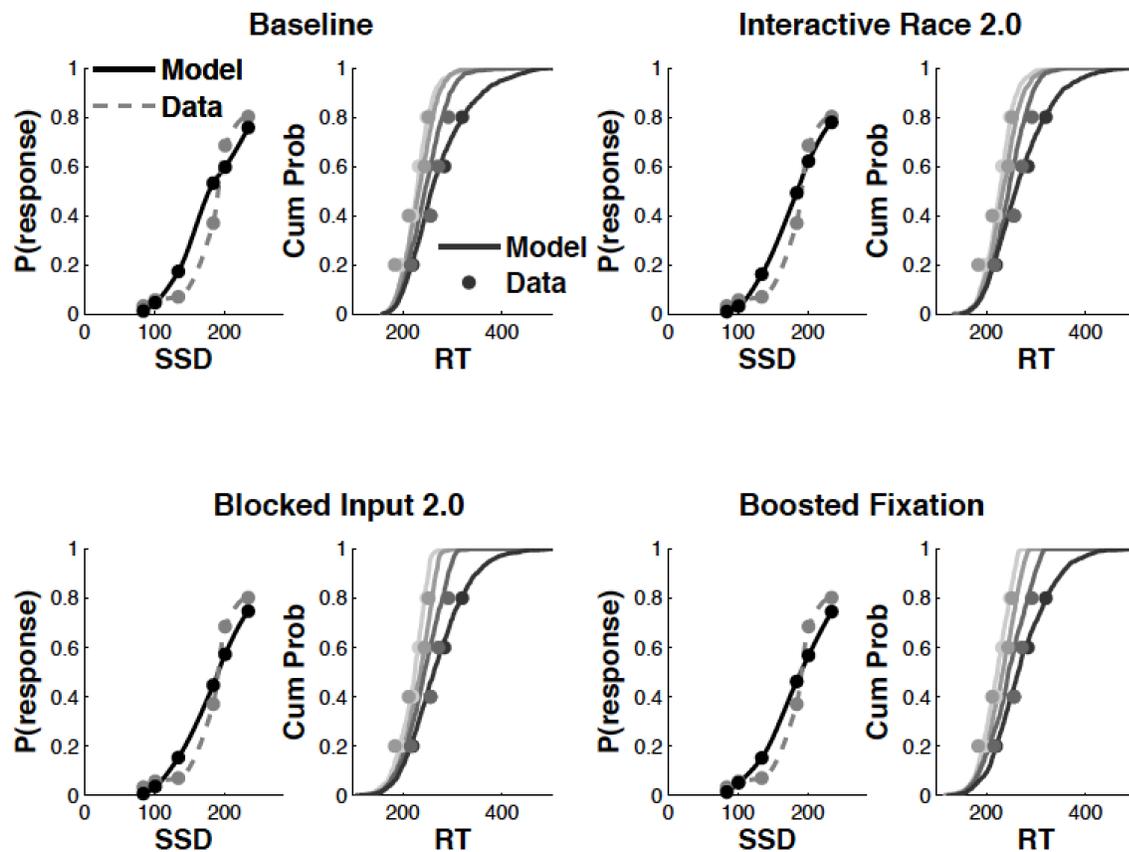


Figure 13. Predicted activation functions for move and fix units in the baseline model, interactive race model 2.0, blocked-input model 2.0, and boosted fixation model 1.0 for monkey C and monkey A. All models address fixation and movement initiation. For each model, the left panel shows activation of the go unit and the right panel shows activation of the stop unit. The horizontal line represents the threshold, estimated from model fits. The left-most vertical line represents stop-signal onset. The right-most vertical line represents stop signal reaction time estimated from the simulated data using the independent race model. The dashed vertical line represents mean divergence time. Cancel time is divergence time minus stop signal reaction time.

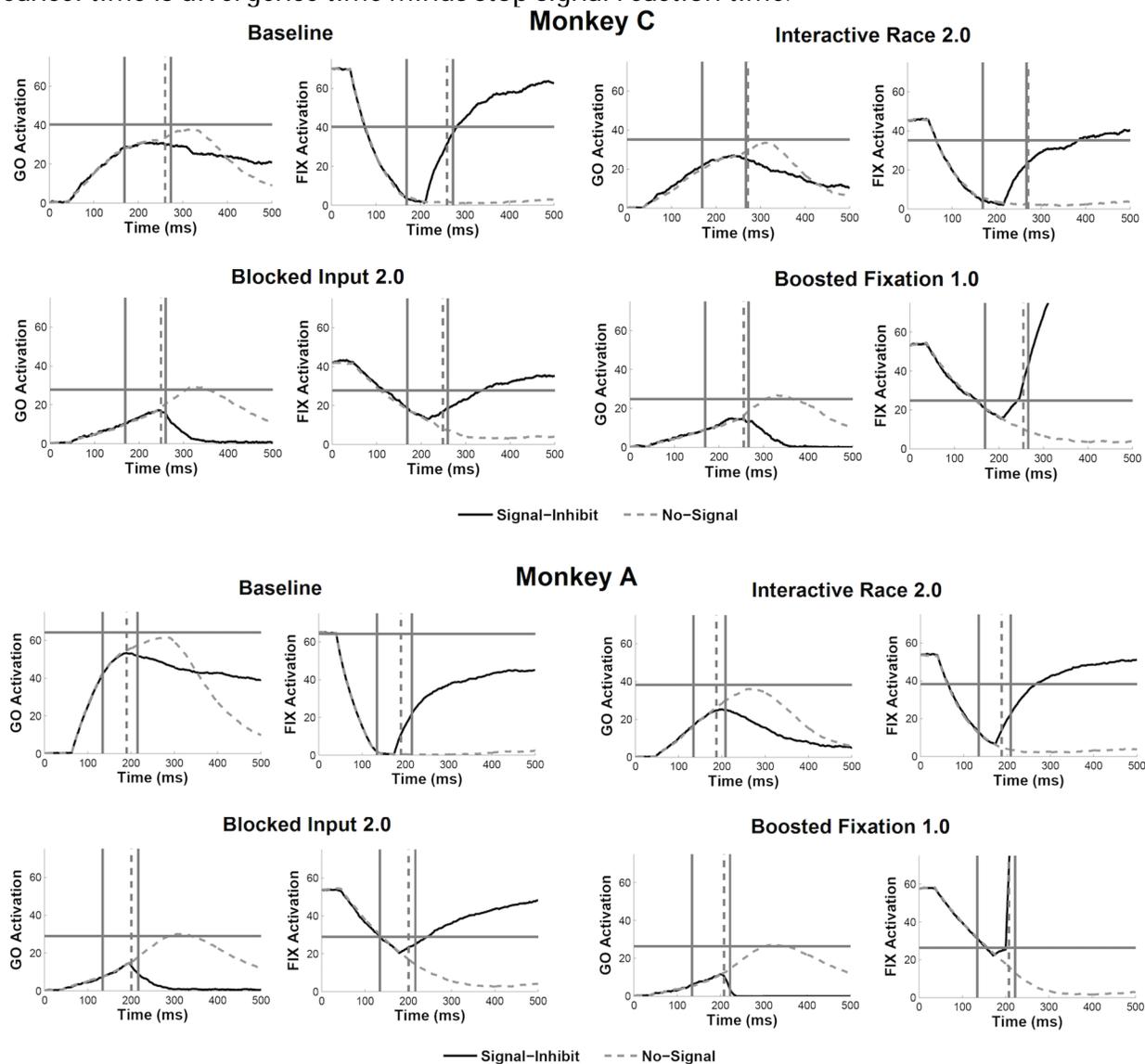


Figure 14. Mean growth and decay rates for observed and predicted activation functions for monkeys C and A. Observed growth and decay rates are estimates from neural activity that was recorded from frontal eye fields (FEF) while the behavior that was modeled were gathered. Predicted cancel times are from the baseline model, interactive race 2.0, blocked input 2.0, and boosted fixation 1.0.

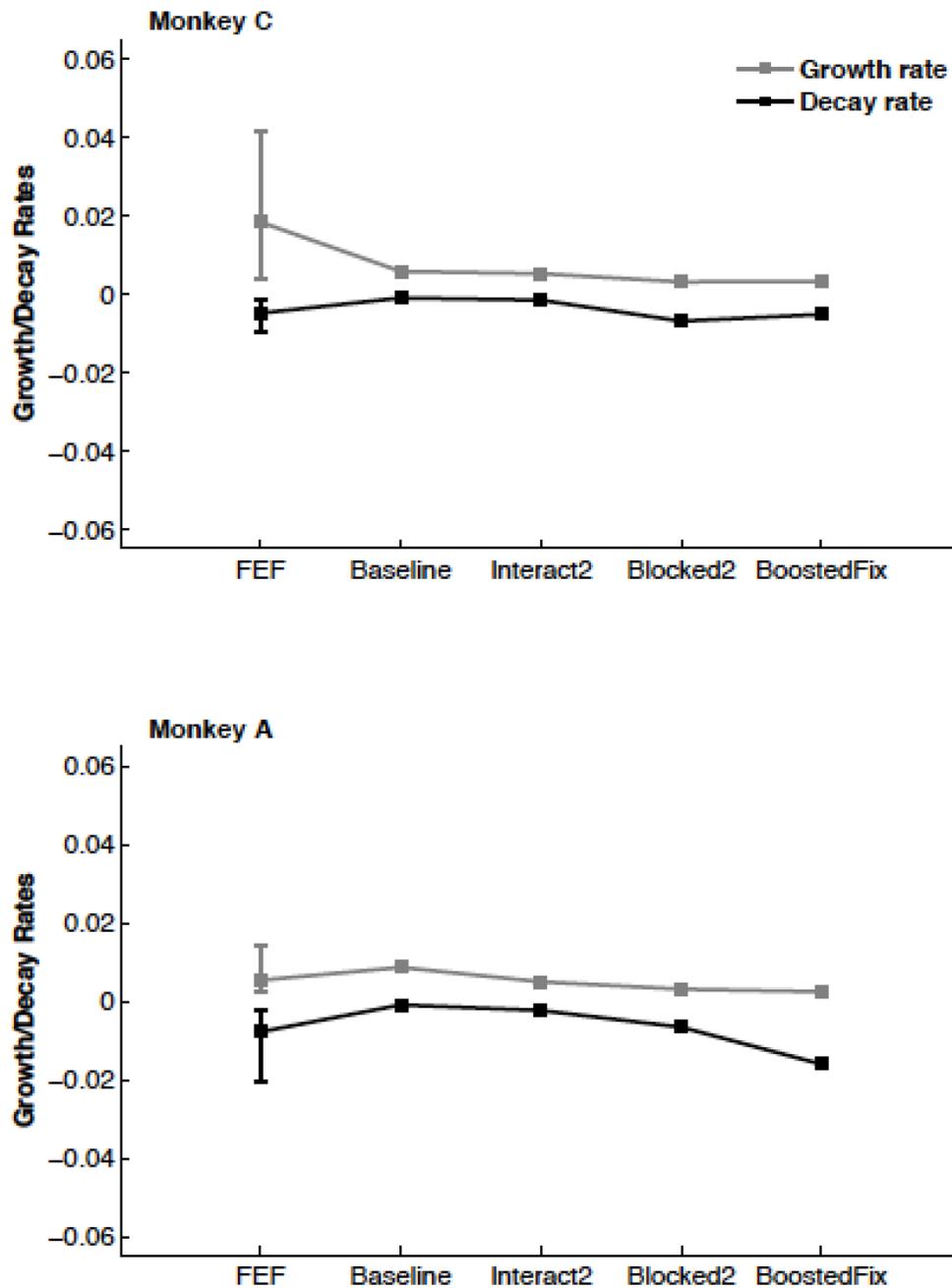


Figure 15. Mean observed and predicted cancel times for monkeys C and A. Observed cancel times are estimates from neural activity that was recorded from frontal eye fields (FEF) while the behavior that was modeled were gathered. Predicted cancel times are from the baseline model, interactive race 2.0, blocked input 2.0, and boosted fixation 1.0.

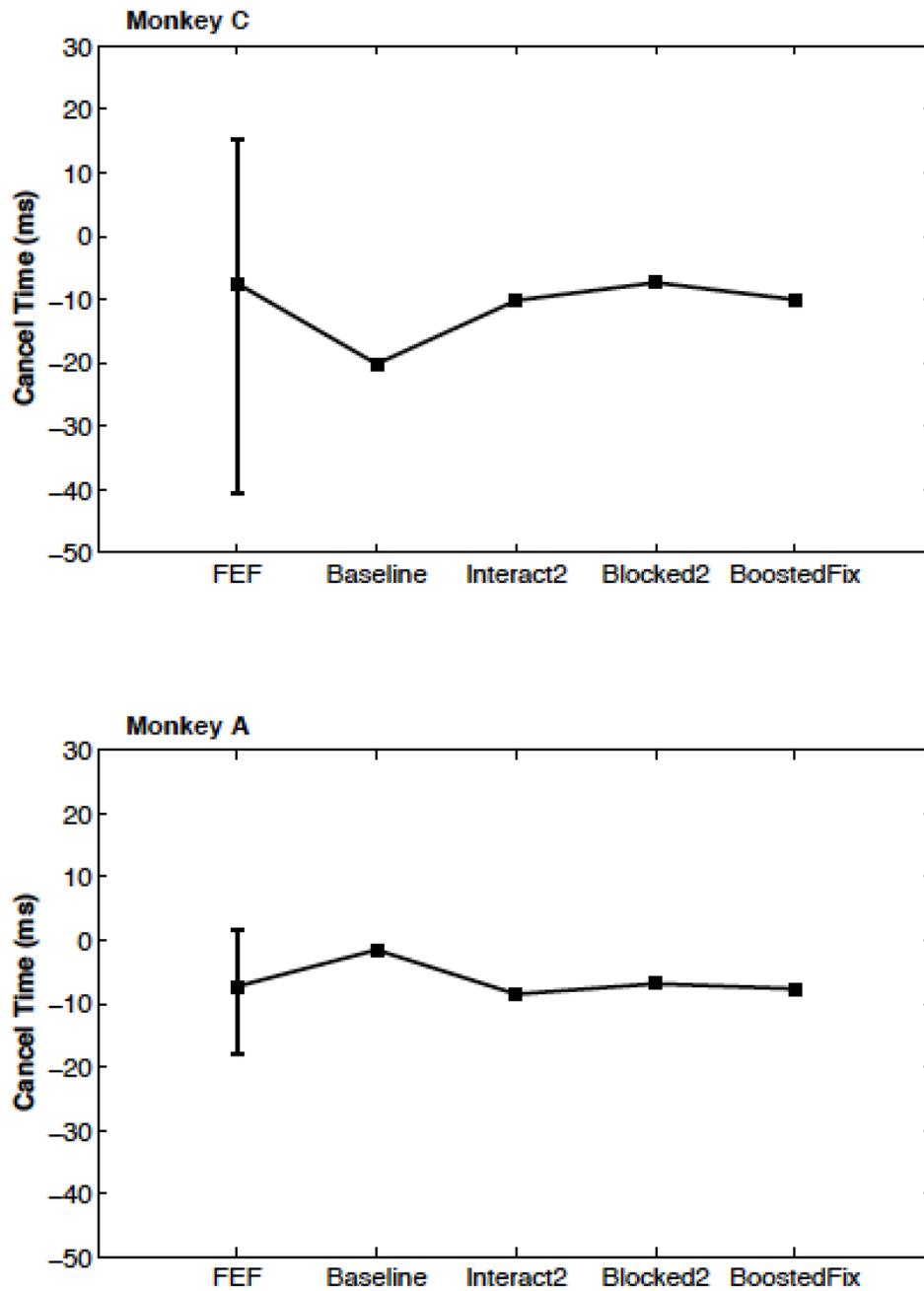


Figure C1: Mean predicted cancel times for monkey C for interactive race model 1.0 with  $\beta_{stop}$  varied and for blocked input model 1.0 with  $k_{go}$  varied.

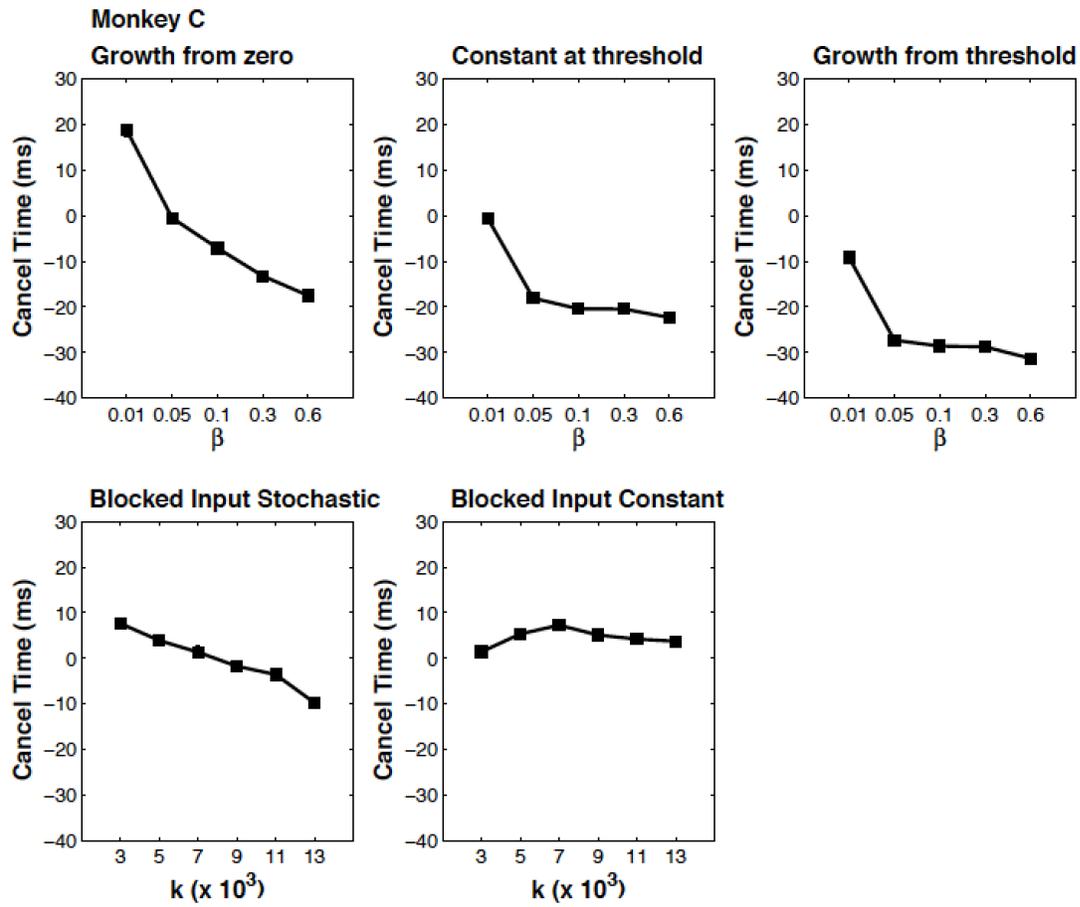


Figure C2: Mean predicted cancel times for monkey A for interactive race model 1.0 with  $\beta_{stop}$  varied and for blocked input model 1.0 with  $k_{go}$  varied.

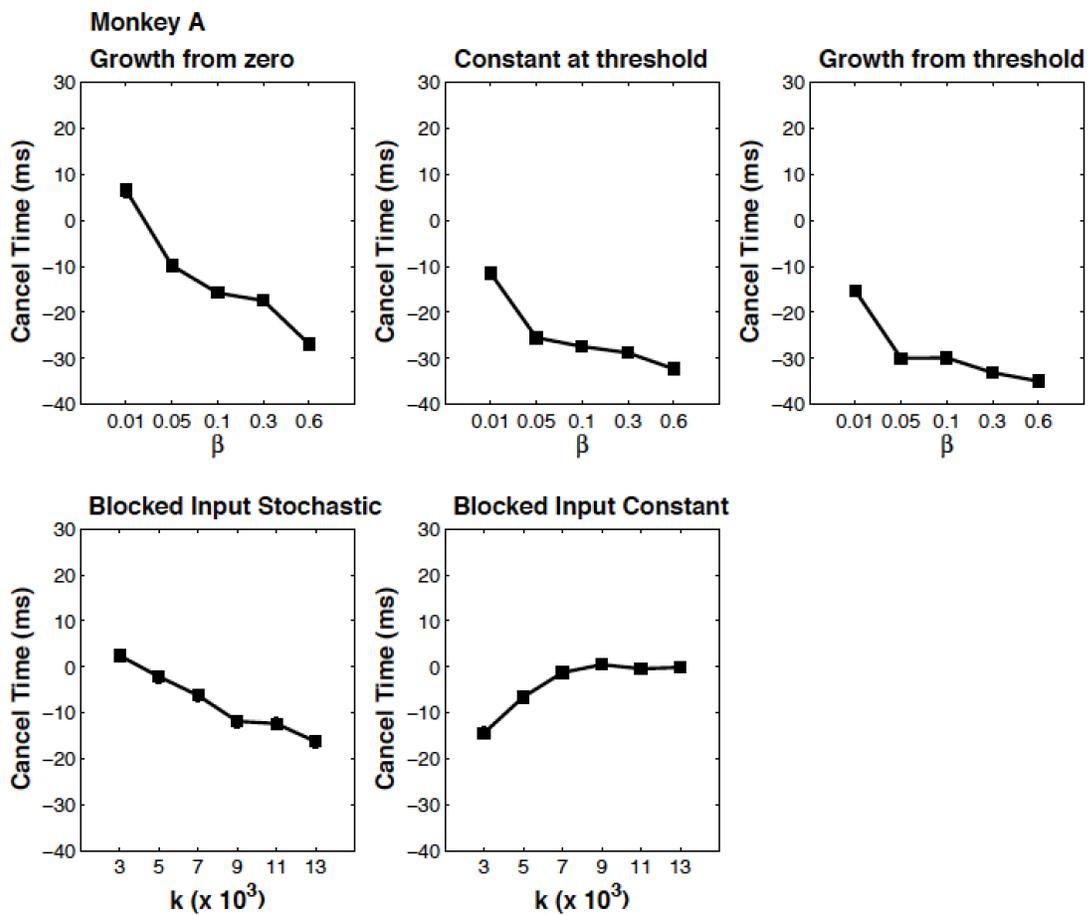


Figure C3: Mean  $\chi^2 \pm$  standard deviation for fits of interactive race model 1.0 with  $\beta_{stop}$  varying and fits of blocked input model 1.0 with  $k_{go}$  varying.

