Chapter 15
Inhibitory Control in Mind and Brain: The Mathematics and Neurophysiology of the Underlying Computation

Gordon D. Logan, Jeffrey D. Schall and Thomas J. Palmeri

Abstract We develop desiderata for a computational theory of response inhibition that links mathematical psychology with neuroscience. The theory must be explicit mathematically and computationally, and grounded in behavior and neurophysiology. The theory must provide quantitative accounts of complexities of behavior in response inhibition tasks and must predict the neural activity that underlies performance. We evaluate three current theories of response inhibition in the stop signal paradigm using these desiderata, and we find that one theory fulfills the desiderata better than the others.

15.1 Introduction

> Yawning, Goldilocks walked into the bedroom and saw three beds. "This one's too big," she said. "This one's too small. But this one's just right." She crawled under the covers, fell fast asleep, and dreamed of unimagined wonders.

We are lucky to live in an era in which the dreams we dared to dream are coming true. Mathematical psychology and neuroscience are merging, and the merger is yielding amazing insights into the mind and brain that were unimaginable 20-years-ago. Mathematical psychology has provided us with precise, explicit descriptions of mental processes that are linked tightly to behavior, making strong predictions about behavior that stand up to rigorous empirical tests. Accurate prediction of response time (*RT*) distributions for correct and error responses is now commonplace, and it is the standard by which models are judged. Neuroscience has opened the black box and shown us how the neural processes underlying behavior interact and unfold in real time. Analysis of spike trains from single neurons, local field potentials from groups of neurons, and electroencephalographic activity at the dura, skull, and scalp have revealed the time-course of information processing. Studies of anatomy, lesions, and brain imaging have shown us the networks of neurons that process information. In recent years, we have seen a proliferation of theories that merge the insights from

G. D. Logan (✉) · J. D. Schall · T. J. Palmeri
Department of Psychology, Vanderbilt University, Nashville, TN 37203, USA
e-mail: gordon.logan@vanderbilt.edu

B. U. Forstmann, E.-J. Wagenmakers (eds.), *An Introduction to Model-Based Cognitive Neuroscience,* DOI 10.1007/978-1-4939-2236-9_15

mathematical psychology and neuroscience, identifying the computational mechanisms in mathematical models with individual neurons and systems of neurons that implement the computation, and testing the identification rigorously by fitting both behavioral and neural data. In all these models, the fundamental insight that made the dream come true is the idea that mind and brain are the computers that produce behavior, and the computation is one and the same.

15.2 Imagining the Dream

We dreamed of a theory that applies that fundamental insight to response inhibition, especially in the *stop-signal* or *countermanding* task [13]. We dreamed of a theory that was formulated explicitly in mathematics or computer simulation, grounded in behavior, computation, and neurophysiology. The theory should accurately predict important behavioral phenomena with models that are connected to the extensive theory of stochastic accumulation to a threshold. The theory should specify linking propositions that connect the mathematical description to neurons, groups of neurons, or brain regions [22, 23]. The linking propositions identify the points of contact between theory and neural data, and specify the aspects of the data that are relevant to the theory. In the stop-signal task, a theory of response inhibition must provide a quantitative account of the probability of inhibiting a response and explain how it varies with the time available to stop (*stop-signal delay*, or *SSD*). The theory must provide a quantitative account of RT distributions for error and correct responses. In the stop-signal task, this means accounting for the relation between failures to inhibit (*signal-respond* or *non-cancelled* trials) and successful responses to the go task (*no-stop-signal* or *cancelled* trials), and accounting for changes in the signal-respond RT distribution with SSD.

Our dream theory provides a list of desiderata that we have used to guide our own modeling: The theory must account for behavior, neurophysiology, and computation, it must be explicit mathematically or computationally, and it must fit the data better than plausible alternatives. In this chapter, we use these desiderata to evaluate current theories of response inhibition in the stop-signal task. The theories are formulated at three different levels of analysis. The highest level addresses networks of brain regions that participate in response inhibition, specifying the interactions within and between regions. The middle level addresses firing rates in systems of neurons that participate in response inhibition, specifying excitatory and inhibitory connections. The lowest level addresses spiking neurons, specifying the connections between spike trains and the underlying biochemistry. Like Goldilocks, we will conclude that one of these levels is too big, one is too small, and one is just right. But we are getting ahead of ourselves. Let us begin by describing behavior in the stop-signal task and the independent race model that accounts for it.

Waking just enough to notice the world around us, we realize there are other dreamers and other dreams. In the other dreams, Goldilocks might prefer a bigger or smaller level of theorizing, fulfilling desiderata that emphasize large networks or biochemistry. Rolling over, we snuggle back into our own dream for the rest of this chapter.

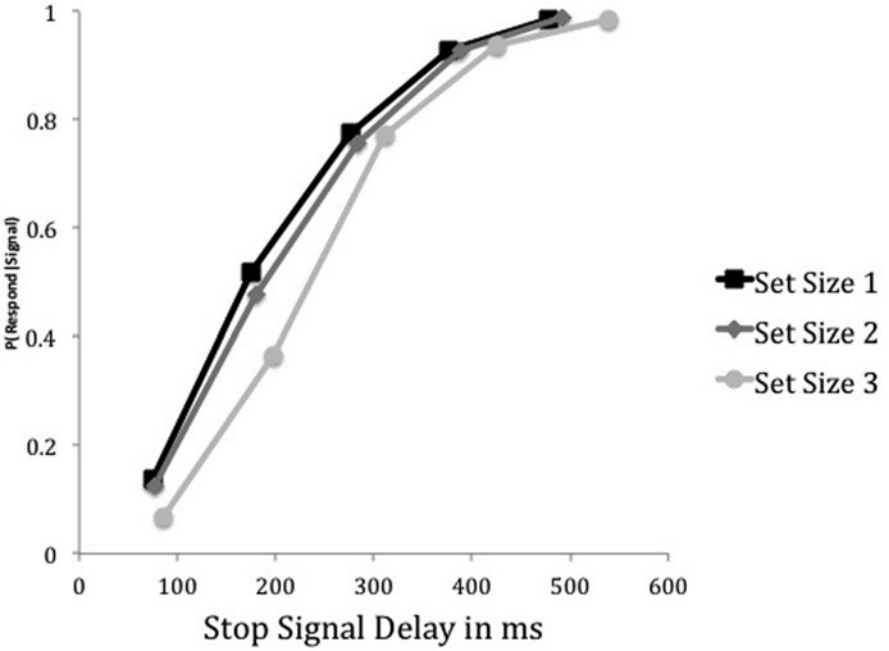

Fig. 15.1 Inhibition Function from a memory-search experiment in which the number of items in the memory set was varied. The probability of responding given a stop signal increases as stop-signal delay (*SSD*) increases and decreases as response time (RT) in the go task increases (RT1 < RT2 < RT3)

15.2.1 Response Inhibition in the Stop-Signal Paradigm

The ability to inhibit our responses voluntarily is a paradigm case of cognitive control. It shows we have "the freedom to do otherwise," which is a hallmark of free will. It reveals itself in many behavioral paradigms, but it is revealed most clearly, simply, and directly in the stop-signal paradigm (for reviews, see [12, 13, 26]). In this paradigm, subjects perform a "go" task, in which they make a speeded response to an imperative stimulus. On some trials, a "stop signal" is presented that tells subjects to inhibit their response to the go signal. Whether or not they are able to is the main datum of interest. Many studies show that the ability to inhibit responses is probabilistic, and the probability of inhibition depends primarily on SSD (see Fig. 15.1). Stop-signal delay controls the amount of time available to detect the stop signal and countermand the go response before the go response is executed; response inhibition is more likely when more time is available. Signal-respond RT is also an important datum. It is usually faster than RT on trials with no stop signal, as if it comes from the faster tail of the go RT distribution (see Fig. 15.2).

These effects have been observed in several species, including rats, monkeys, and humans, in several subject populations, including children, adolescents, young adults, and the elderly. These effects have been observed in several psychiatric

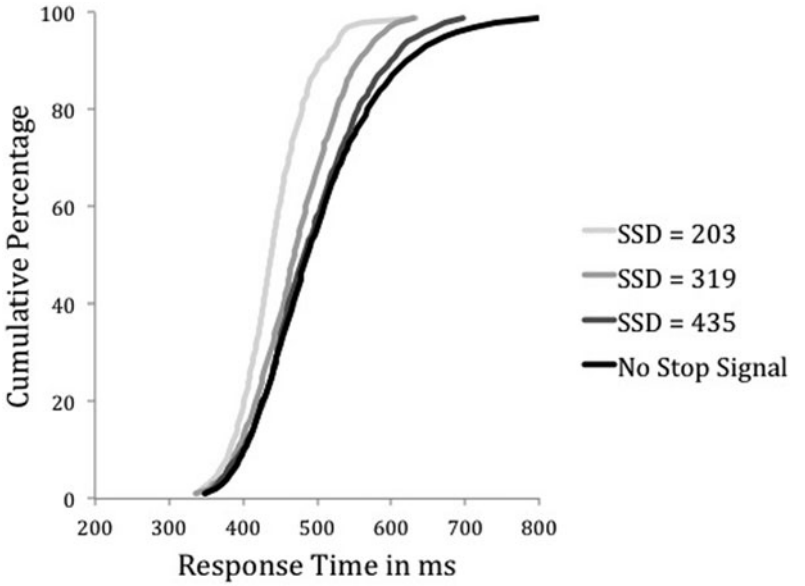

Fig. 15.2 Distributions of response time on no-stop-signal trials and on signal-respond trials with stop signal delay (SSD) equal to 231, 364, and 496 ms. Signal-respond distributions are faster than no-stop-signal distributions. They begin with a common minimum and end with a shorter maximum

disorders, including attention deficit hyperactivity disorder and schizophrenia, and in several neurological disorders, including stroke and Parkinson's disease. They have been observed in different stimulus and response modalities, in different tasks, in different experimental conditions, and with different strategies. The patterns are the same qualitatively, but they differ quantitatively, and the quantitative differences reveal important changes or deficits in cognitive control.

15.2.2 Independent Race Model

Two facts led Logan and Cowan [13] to propose the independent race model of stop signal performance: (1) The probability of response inhibition depends on the time available to detect the stop signal before the go response is executed, and (2) signal-respond RTs are faster than RTs on no-stop-signal trials. These facts suggested that response inhibition depends on the outcome of a race between a go process, initiated by the go stimulus, and a stop process, initiated by the stop signal. If the stop process finishes before the go process, the response is inhibited, producing a signal-inhibit trial. If the go process finishes before the stop process, the response

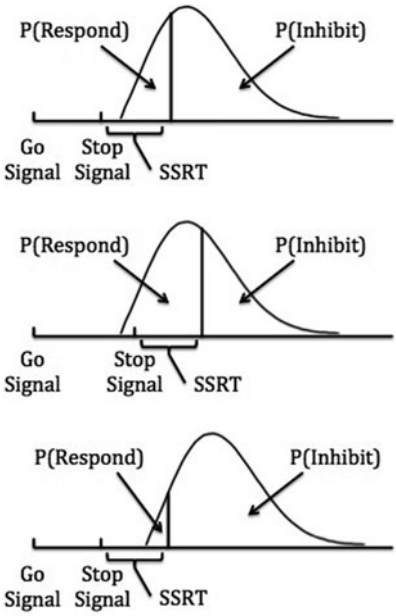

Fig. 15.3 Predictions of the independent race model, assuming *SSRT* is constant. Onset of Go Signal followed by onset of Stop Signal after a stop-signal delay. *Vertical line* across the distribution represents the finishing time of the stop process. Probability of responding is area to *left* of line; probability of inhibiting is area to *right* of line. Top panel: standard condition. *Middle* panel: Stop-signal delay increases, so probability of responding increases. Bottom panel: Go response time increases, so probability of responding decreases

is not inhibited, producing a signal-respond trial. The model assumes that the finishing times for the stop and go processes are independent random variables, and demonstrates that the fundamental results in the stop-signal paradigm follow from these assumptions (see Fig. 15.3).

The independent race model provides a measure of the latency of the stop process, called *stop-signal reaction time* (*SSRT*). This is an important contribution because the stop process is not directly observable. If the stop process finishes before the go process, there is no response whose latency can be measured. If the stop process finishes after the go process, we know SSRT must have been longer than signal-respond RT, but we do not know how much longer. The independent race model provides several converging methods for estimating SSRT from the observed data. These measures of SSRT have been important in documenting differences in the ability to inhibit responses across lifespan development, between clinical and control groups, and between neurological patients and controls. They have also been important in understanding the neurophysiology of response inhibition. Neural processes that cause response inhibition must modulate before SSRT; neural processes that are consequences of response inhibition modulate after SSRT.

Since it was formulated in 1984, the independent race model has been used in virtually every stop-signal experiment. It provides important measures of cognitive

control, like SSRT, and it provides a benchmark against which other models can be evaluated. Its prevalence results from its generality: It is formulated in terms of generic finishing time distributions for the stop and go processes. It makes no commitment to the underlying computational or neural processes that generate these finishing times. It expresses relationships that must hold for any and all distributions, regardless of the process that generates them. This is important because the independent race model provides an important check for the models we consider here that address the computations performed by the underlying neural processes: these models must predict the empirical relationships predicted by the independent race model.

The independent race model is like a dream: it captures the essence but not the details. It formulates the constraints that any model of response inhibition must follow, but it does not provide the structure that seems necessary to explain recent developments in stop-signal research. For example, many studies have shown that go RT is slower when stop trials occur more frequently, as if the go process changes to balance the competing demands of stopping and going. Many other studies have shown that go RT is slower on trials following stop signals than on trials before them, suggesting that a stop trial results in some kind of strategic adjustment to the go process. To explain how these adjustments occur, we need a more detailed model of the go process that tells us which parts can support this strategic adjustment. The independent race model provides no model of the underlying process. It can describe these effects, but it cannot explain them.

15.3 Feeding the Dream

Developments in mathematical psychology and neuroscience around the turn of the twenty-first century set the stage for the development of models that link mind and brain. Mathematical psychologists developed a variety of *stochastic accumulator models* that explained RT distributions for correct and error responses as resulting from processes that accumulate information until a threshold for responding is reached. Many studies evaluated the strengths and weaknesses of random walk, diffusion, race, and leaky competitive accumulator models, using increasingly sophisticated methods for assessing goodness of fit and increasingly stringent comparative tests of one model against another e.g. [20]. Models must fit large amounts of data with a small number of free parameters, and they must fit better than plausible alternatives when model complexity is taken into account. Researchers either compare one model architecture against another or compare different models in the same architecture to determine which parameters are necessary and sufficient to account for the data. These models and the approach they took to modeling inspired more specific models of response inhibition with greater explanatory power.

At the same time, neuroscientists were training animals to perform the stop-signal task and recording from their brains as they performed it. Hanes and Schall [7] showed that monkeys performed a saccadic version of the stop signal task much like humans. The probability they would inhibit their eye movements depended on SSD and their signal-respond RTs were faster than their no-stop-signal RTs. Hanes, Patterson and

Schall [8] recorded from frontal eye fields in monkeys performing the saccadic stop signal task, isolating neurons involved in gaze shifting and gaze holding that represent a larger circuit of such neurons that extends from cortex through basal ganglia and superior colliculus to brainstem. They found that these neurons modulated on stop-signal trials, modulating just before SSRT when the monkey stopped successfully. Paré and Hanes [15] reported similar results in superior colliculus. Meanwhile, studies of humans with lesions in frontal cortex revealed deficits in stop-signal inhibition, and functional magnetic resonance imaging (*fMRI*) on healthy young adults suggested the involvement of a circuit including frontal cortex, basal ganglia, and subthalamic nucleus [1]. These rich neural data sets demand computational explanations that are more detailed than the description the independent race model provides.

15.4 Dreaming the Dream

In recent years, many theories of response inhibition have been developed. We focus on three models that account for behavior, computation, and neurophysiology in the stop-signal task. One focuses on brain regions, one focuses on processes that generate spikes and spike trains, and one focuses on firing rates in single neurons. Like the Goldilocks in our dream, we conclude that one is too big, one is too small, and one is just right. Of course, other Goldilocks' in other dreams may reach different conclusions.

15.4.1 Single Neurons: The Interactive Race Model

Boucher et al. [5] formulated an interactive race model to address a paradox they encountered in linking models to neurons: How can a model that assumes independent stop and go processes explain behavior that is supported by interacting circuits of mutually inhibitory gaze-holding and gaze-shifting neurons? They addressed this question by instantiating the stop and go processes as mutually inhibitory leaky competitive accumulators ([25]; see Fig. 15.4). The go accumulator begins after an afferent delay, D_{go}, accumulating activation until it reaches a threshold, whereupon a response occurs. The stop accumulator begins after an afferent delay, D_{stop}, inhibiting the go response in proportion to its activation. If the stop accumulator becomes active soon enough (if SSD + D_{stop} < go RT), it prevents the go accumulator from reaching threshold and the response is inhibited. If the stop process becomes active too late (if SSD + D_{stop} > go RT), the go accumulator reaches threshold and the response is not inhibited.

Boucher et al. [5] specified the stochastic differential equations that govern the stop and go accumulators and used them to drive computer simulations. They fit the simulations to behavioral data from two monkeys, who also provided neural data from the same test sessions, manipulating the mean and standard deviation of go and stop accumulation rates and the mutual inhibition from stop to go and go to stop to

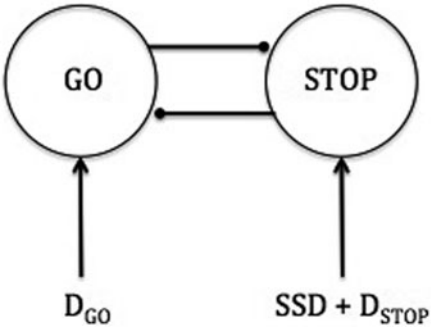

Fig. 15.4 Interactive race model. Arrows represent excitatory connections; dots represent inhibitory connections. The GO unit receives input after an afferent delay (D_{GO}) and the STOP unit receives input after stop-signal delay (*SSD*) plus an afferent delay (D_{STOP}). GO and STOP units inhibit each other. Inhibition from STOP to GO is much greater than inhibition from GO to STOP. A go response occurs if GO activation reaches threshold. The go response is inhibited if inhibition from the STOP unit prevents it from reaching threshold

optimize goodness of fit. The model fit the data well, providing accurate quantitative accounts of the inhibition function, no-stop-signal RTs, and signal-respond RTs at several SSDs (see [5], Fig. 6). Thus, the model fulfills the behavioral side of the desiderata of our dreams.

Boucher et al. then simulated the growth and modulation of activation of the go and stop accumulators, using the parameters that produced the best fits to the behavioral data, and matched the simulated patterns of activation to measured patterns of activity in gaze-holding and gaze-shifting neurons that were recorded while monkeys performed the stop signal task. To assess the match between simulated and recorded activity, Boucher et al. had to decide which aspect of the recorded activity to assess. The pattern of activation for an individual neuron has many idiosyncrasies, but all patterns show some general characteristics. In order to fit the "signal" and not the "noise," Boucher et al. focused on distributions of *cancel times*, which are the times at which neural activity modulates on trials on which subjects stop successfully, relative to SSRT. They assessed this in the simulated data in the same way they assessed it in neural data, by determining the point at which activation on successful stop trials first differed significantly from activation on latency-matched no-stop-signal trials. In the neural data, this point ranges from 50 ms before to 50 ms after SSRT, with a mean 5–10 ms before SSRT. The model predicted distributions with the same range (see [5], Fig. 7). Note these are genuine predictions. They were generated with a fixed set of parameters that provided the best fit to the behavioral data, without any further adjustment to optimize the fit to neural data. Thus, the model fulfills the neural side of the desiderata of our dreams.

The final desideratum is comparative model fitting. Boucher et al. [5] compared the interactive race model with a version of the independent race model in which the stop and go process were modeled as leaky accumulators with no competition. After

their respective afferent delays (D_{stop} and D_{go}) they accumulate activation until one of them reaches a threshold. If the stop process finishes first, the response is inhibited; if the go process finishes first, the response is executed. Boucher et al. found that the independent race model fit the behavioral data as well as the interactive race model, suggesting mimicry. Normally, parsimony would favor the simpler independent race model over the more complex interactive race model. However, Boucher et al. argued that the interactive race model accounted for the neural data, predicting modulation of go activation on stop-signal trials and predicting cancel time distributions accurately, while the independent race model did not. They argued that this favored the interactive race model. Thus, the interactive race model fulfills all of the desiderata of our dreams: it is computationally explicit, it explains the underlying processes computationally and neurally, it provides accurate quantitative accounts of behavioral and neural data, and it won in competitive tests against a plausible alternative. If Goldilocks were a mathematical psychologist, we believe she would find our model just right.

What about the paradox? The interactive race model assumes an interaction between gaze-holding and gaze-shifting units, like the interaction between gaze-holding and gaze-shifting neurons that underlies eye movements. How can it account for data that are described just as well by the independent race model? The answer lies in the values of the best-fitting parameters: In order to fit the behavioral data, D_{stop} had to be long—almost as long as SSRT—and inhibition from the stop process on the go process had to be much stronger than the inhibition from the go process on the stop process. Thus, the stop process and the go process were independent for most of their durations, and response inhibition resulted from late and potent inhibition just before a go response occurred.

15.4.2 Spikes and Spike Trains: The Spiking Neuron Model

Lo et al. [10] implemented the Boucher et al. [5] interactive race model in Lo and Wang's [9] spiking cortico-basal ganglia circuit model of RT (see Fig. 15.5). The model assumes hundreds of units representing populations of movement neurons, fixation neurons, and inhibitory interneurons, and a control unit that turns the fixation neurons on and off. Each population produces Poisson spike trains that depend on the ratio of parameters representing NMDA and AMPA inputs. The model addresses fixation activity at the beginning of a trial and the transition from fixation to movement as well as the rise in movement activation to threshold. The model produces the transition from fixation to movement, and ultimately RT, by turning off the control unit that excites fixation units, thereby releasing tonic inhibition on the movement units and allowing their activity to rise to threshold.

Lo et al. [10] fit data from one of the two monkeys Boucher et al. [5] modeled. They fixed the number of units and many of the parameters across all conditions and manipulated three parameters to maximize goodness of fit: The mean and standard deviation of a Gaussian distribution for the time at which the control unit turned off,

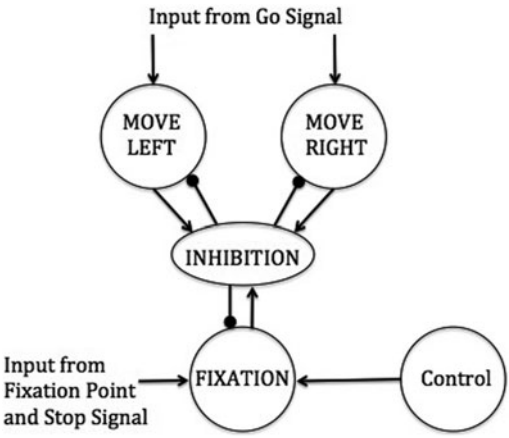

Fig. 15.5 Spiking neuron model. Arrows represent excitatory connections; dots represent inhibitory connections. The MOVE units are excited by input from the go signal. The *FIXATION* unit is excited by input from the fixation point and the stop signal and by control input. *MOVE* and *FIXATION* units activate an *INHIBITION* unit that inhibits them all. The Control unit tonically excites the *FIXATION* unit. A go response occurs when the Control unit releases excitation on the fixation unit. The go response is inhibited if the stop signal excites the *FIXATION* unit before a *MOVE* unit makes its response

and the time at which the stop signal turned the fixation units back on (analogous to D_{stop} in [5]). Their fits to RT distributions for no-stop-signal and signal-respond trials and their fits to inhibition functions were about as good as the fits Boucher et al. [5] obtained. Like Boucher et al., Lo et al. found that D_{stop} had to be relatively long to produce appropriate signal-respond RT distributions; inhibition of stop on go had to be late and potent. The Lo et al. model also predicted modulation of movement and fixation neurons and cancel time distributions qualitatively as well as Boucher et al. [5], although these predictions were not assessed quantitatively.

Lo et al. [10] modeled the effects of changes in baseline activation in fixation and movement units on the probability of successful inhibition. Successful inhibition was less likely when movement units were more active during the baseline period and more likely when fixation units were more active. They tested these predictions by reanalyzing data from the Hanes et al. [8] and Paré and Hanes [15] countermanding studies, and found lower baseline firing rates in movement neurons prior to successful inhibition.

What about our dream? The Lo et al. [10] model fulfills our behavioral desideratum, fitting the behavioral data as well as the Boucher et al. [5] model. The Lo et al. model fulfills our neural desideratum as well, describing stop and go units as spiking neurons and linking the computation to the biochemistry that generates spikes. However, the model does not fulfill our computational desideratum very well. RT depends on turning off a control unit that tonically excites fixation units, which releases inhibition on movement units and allows their activity to rise to threshold. The variability in RT depends primarily on the variability in the time at which the control signal is turned off, which is determined arbitrarily by a Gaussian distribution whose mean and standard deviation were free parameters that were adjusted to optimize

goodness of fit (113 and 95 ms, respectively). The control unit is like a homunculus outside the model that intervenes at the right time to produce the right effect. It is not grounded in the physiology, like movement and fixation units. There are no linking propositions [22, 23] that tie it to neurons or neural structures analogous to the linking propositions that tie movement and fixation units to gaze-shifting and gaze-holding neurons. We prefer models like the Boucher et al. [5] model, in which variability in RT is produced by variable growth in stochastic accumulation [18, 19] over the Lo et al. model, in which variability in RT is produced by an arbitrary control unit.

The Lo et al. [10] model partially fulfills our desideratum of comparative model fitting. Lo et al. compared their fits to Boucher et al.'s [5] fits of the interactive race model and the stochastic-rise-to-threshold version of the independent race model and found that their model fit about as well. They discovered the importance of differences in baseline activity in movement and fixation units in predicting the probability of successful inhibition, but that is not likely to be a unique prediction of their model. Differences in baseline activation could be implemented in the Boucher et al. [5] model, and would likely produce similar results. Thus, the Lo et al. model does not distinguish itself from plausible alternatives in comparative model fits, as our dream model would.

Lo et al. [10] modeled the underlying physiology at a finer grain than Boucher et al. [5], modeling spikes and spike trains rather than firing rates. However, this required many parameters (AMPA and NMDA ratios for each interaction between units) in addition to the three parameters that were varied to optimize goodness of fit. These parameters were fixed for the fitting, but they were tweaked to produce firing rates in the desired range for movement and fixation cells before they were fixed. From the perspective of mathematical psychology, where fitting large amounts of data with a small number of parameters is desirable, this is not a virtue. If Goldilocks were a mathematical psychologist, she would find the focus of this model (on spikes and spike trains) too small.

15.4.3 Brain Regions: The Frontal Cortex-Basal Ganglia Model

Wiecki and Frank [28] formulated a model of inhibitory control that extends Frank's [6] model of basal ganglia to include cortical structures. The new model describes interactions between units in frontal cortex (dorsolateral prefrontal cortex, right inferior frontal gyrus, frontal eye fields), basal ganglia (striatum, globus pallidus external segment, substantia nigra pars compacta, substantia nigra pars reticulata, subthalamic nucleus), and superior colliculus (see Fig. 15.6). It addresses the stop-signal task and an anti-saccade task in which a peripheral target is presented and subjects must inhibit their natural tendency to look directly at it and shift their gaze to a position opposite to it. The model explains stop-signal performance by assuming that the stop signal activates right inferior frontal gyrus, which activates subthalamic nucleus, which activates substantia nigra pars reticulata, which then inhibits superior colliculus. If the superior colliculus is inhibited before its activation reaches threshold, the response is inhibited, producing a signal-inhibit trial. If superior colliculus reaches threshold before it is inhibited, the response is executed, producing a signal-respond trial.

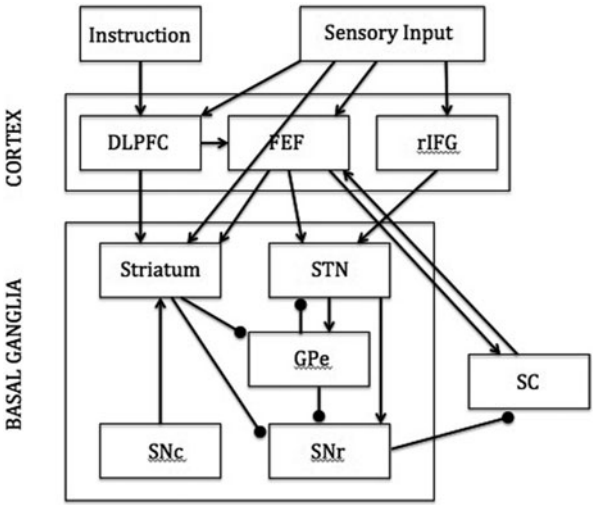

Fig. 15.6 Cortico-basal ganglia model. Arrows represent excitatory connections; dots represent inhibitory connections. *DLPFC* dorsolateral prefrontal cortex; *FEF* frontal eye fields; *rIFG* right inferior frontal gyrus; *STN* subthalamic nucleus; *GPe* globus palladus external segment; *SNc* substantia nigra pars compacta; *SNr* substantia nigra pars reticulate; *SC* superior colliculus. Response inhibition occurs when rIFG activates STN, which activates SNr, which inhibits SC

Wiecki and Frank [28] simulated performance on the stop-signal task but did not fit their model to the data. They simulated inhibition functions and RT distributions on no-stop-signal and signal-respond trials but did not compare the simulated functions quantitatively to observed data. The only observed data they reported were RT distributions taken from one of the monkeys studied by Boucher et al. [5] and Lo et al. [10], and their simulations overestimate the variability in the observed distributions (see their Fig. 12.). They reported simulated activation for units in striatum, substantia nigra pars reticulata, subthalamic nucleus, and dorsolateral prefrontal cortex, but did not compare the changes in activation with observed neural data.

What about our dream? The model fulfills our computational desideratum, explaining the mathematics and computations that occur within and between units, but it does not fulfill the other desiderata as well as we would like. The lack of quantitative fits falls short of fulfilling our behavioral desideratum. Every model of the stop-signal task predicts inhibition functions and RT distributions for no-stop-signal and signal-respond trials, so the model's predictions of the shapes of these functions are far from unique. Moreover, other models predict these functions quantitatively, and the models rise and fall on the accuracy of their quantitative fits. In fact, the independent race model [13] predicts these effects without specifying any of the underlying computations, so it is not clear that the machinery in the Wiecki and Frank [28] model is doing any of the work. We view this as a shortcoming of their model.

The Wiecki and Frank [28] model promises to fulfill our neural desideratum but also falls short. It is clear that stop-signal performance depends on the integrated

action of many brain structures, and the model includes the relevant structures. However, the linking propositions that connect model units to brain structures and interactions between model units to interactions between brain structures are not evaluated very rigorously. The model provides a framework in which these desiderata could be fulfilled, but does not go as far as we would like toward fulfilling them. Quantitative comparisons of critical features of the data (e.g., cancel times) would be steps in the right direction.

The Wiecki and Frank [28] model of the stop task does not fulfill our desideratum of competitive model testing very well. It demonstrates that the model could work, but it does not pit the model against plausible alternatives. Wiecki and Frank evaluate the effects of lesioning model structures and manipulating motivation, comparing different versions of their model, but the evaluation is qualitative, not quantitative. They also apply the model to related tasks, like the antisaccade task, again evaluating the fit qualitatively.

From the perspective of mathematical psychology, this model does not fare well. There are many parameters and essentially no data points. If Goldilocks were a mathematical psychologist, she would find to focus of this model (on brain regions and not on quantitative data) too big. If Goldilocks were a computational neuroscientist with Wiecki and Franks' perspective, she would find this theory just right.

15.5 Waking Up

It is the dawn of a brand new day. We dreamed our dream and still want more. The integration of mathematical psychology and neuroscience has only just begun. We still dream of a grand model that integrates it all, from spikes to brains, and fits a large amount of data with a small number of parameters. In our view, the frontal cortex-basal ganglia model may be too big, the spiking neuron model may be too small, and the interactive race model may be just right, but we dream of a model that integrates all three. The models have moved us significantly forward, but much remains to be done. In the remaining pages, we sketch out our future dreams and some cold, hard realities that we must face.

15.5.1 Choice

Perhaps the most pressing problem is to deal with choice, both in the go task and the stop task. Boucher et al. [5] considered only one accumulator for the go task. Lo et al. [10] and Wiecki and Frank [28] proposed two accumulators, one for each possible go response, but did not model activity in the competing accumulator. This may be appropriate for saccadic stop-signal tasks, where choice errors are exceedingly rare [7], but it is not appropriate for manual stop-signal tasks, which dominate the literature [26]. The probability and latency of choice errors need to be modeled. The alternative responses must be modeled as stochastic accumulators, and their

interaction with the stochastic accumulator for the correct response must be specified. Race models, feed-forward inhibition models, and lateral inhibition models are viable alternatives [18–20]. Choice tasks provide the opportunity to manipulate several factors that affect the go process concurrently, and these factors may influence different parameters of the go process selectively. Selective influence provides important leverage in modeling: Some parameters should stay constant across conditions while others vary, and this adds important constraints in fitting data. We are currently working on developing models that implement choice in the go task. We recently extended the independent race model to deal with choice in the go task and found some evidence for selective influence [14].

Choice is also possible in the stop process. Several investigators have studied varieties of "selective inhibition," in which some responses but not others must be stopped when a stop signal occurs [2], or all responses must be stopped when a stop signal occurs but not when another similar "ignore" signal occurs [4]. Selective stopping may pose a significant challenge for modeling. Bissett and Logan [4] found that selective stopping to one stimulus but not another often produces violations of the independence assumptions of the race model. This is important because all of the models we have discussed, from Logan and Cowan [13] to Wiecki and Frank [28], assume that the stop process and go process are independent for much of their duration. Independence makes modeling simpler. Non-independent stop and go processes are much harder to characterize. We are beginning to work on models of selective stopping.

15.5.2 Mechanisms of Response Inhibition

The models we discussed consider only one mechanism for inhibiting responses: inhibiting the growth of activation in go accumulators. Other mechanisms have been proposed in the literature and must be distinguished from this one [3, 13]. Salinas and Stanford [21] note that the main computational requirement for a mechanism of response inhibition is to halt or reverse the growth of activation in the go accumulator. They propose a generic model that halts and reverses the growth but do not commit to the underlying mechanism. In their view, it need not be inhibition. In our view, which mechanism underlies inhibition is an empirical question, which we are invested in answering.

Logan and Cowan [13]; also see [3] proposed a *blocked input* mechanism for countermanding responses. They suggest that go responses are driven by input from perceptual systems, and go responses can be countermanded by blocking the input to the motor system. The input can be blocked in several ways. One possibility is deleting the goal to act. In production system models, action depends on two conditions: a goal and an appropriate stimulus. The action can be countermanded by removing the goal, by removing the stimulus, or by removing both. Another way to countermand responses is to suppress the input from perceptual systems. In stochastic accumulator models, this involves setting the drift rate to zero (or less). A

third possibility is to break the connection between perceptual and motor systems. The mapping of go stimuli onto go responses is often arbitrary (e.g., "press the left key if an X appears") and must be maintained somewhere in the cognitive system [11]. Disabling the mapping rules would prevent the growth of activation in the motor system. In our model of visual search [18, 19], the connection between perceptual and motor activity is controlled by a gate that prevents noise from accumulating in stochastic accumulators. Responses could be countermanded by raising the gate to a much higher level.

Boucher et al. [5] evaluated a blocked input model, in which the drift rate for the go process was set to zero after the stop accumulator reached threshold. They found it did not fit the data as well as their interactive race model. We have re-evaluated the same model and several variants, and we do not replicate Boucher et al.'s findings. In our simulations, the blocked input model fits the data as well as or better than the interactive race model. We are currently working hard on this issue.

15.5.3 Model Mimicry

The models we discussed make very similar predictions for behavior and physiology. Quantitative fits to behavior—inhibition functions and RT distributions for no-stop-signal and signal-respond trials—were equivalent for the Boucher et al. [5] interactive race model, the Boucher et al. stochastic accumulator version of the independent race model, and the Lo et al. [10] spiking neuron model. Even the Wiecki and Frank [28] frontal cortex-basal ganglia model produced the same qualitative trends. Perhaps considering other data sets and more complex experimental designs can break this mimicry. All of the models were fit to a single data set from one monkey from Hanes et al. [8], (Boucher et al. also fit data from another monkey). In most applications of mathematical psychology, this would not be sufficient. However, the goal of these models is to predict behavior and neurophysiology simultaneously, and that requires fitting data sets in which behavioral and neural measures were gathered in the same session in the same subject. So far, the only data that meet this criterion are from Hanes et al. [8] and Paré and Hanes [15]. We are currently working toward gathering behavioral and neural data from monkeys performing a stop-signal task in which we manipulate choice difficulty in the go task.

Neural measures exhibit mimicry too. The Boucher et al. [5] interactive race model and the Lo et al. [10] spiking neuron model predict similar modulation of activity in movement and fixation neurons and predict similar distributions of cancel times. Our current investigations of blocked input models mimic these predictions. Moreover, the stochastic accumulator version of the independent race model that Boucher et al. investigated could predict similar modulation and cancel time distributions if the measures were defined a little differently. The modulation of go activation could be defined as the maximum value of the go accumulator on signal-respond trials. With that definition, the independent race model would modulate much like the interactive race model. It would stop before it reached threshold, and the level of activation

would be lower the shorter the SSD, like the observed data. Similarly, cancel time distributions could be generated for the independent race model by comparing the maximum activation on signal-respond trials to the activation on latency-matched no-stop-signal trials. These cancel times would fall in the range of observed cancel times (i.e., SSRT $\pm$ 50 ms).

The mimicry in the neural measures may be broken by examining different neural measures and examining activation and modulation quantitatively (e.g., [18, 19]). For example, Pouget et al. [16] compared baseline, onset of growth, growth rate, and threshold measures in neurons on trials that followed stop signal and no-stop-signal trials to determine the cause of slowing after a stop signal. They found the onset of growth changed, but none of the other measures did. Similar measures could be taken for stop-signal and no-stop-signal trials to determine the cause of stopping. The neural measures could be compared with measures taken from simulations of the models to determine which model provides the best account of the physiology. However, model simulations suggest that such measures may not always agree well with the values of the parameters that generated them, especially if there is noise in data and the model predictions, and there always is. Measured onsets do not always correspond to non-decision times, measured rates of growth do not always correspond to drift rates, and measured thresholds do not always correspond to model thresholds [17].

15.5.4 Fitting Behavior and Physiology Simultaneously

Our strategy has been to fit models to behavioral data and then use the best-fitting parameter values to generate predictions for neural measures. The virtue of this strategy is that the predictions are genuine predictions. No further adjustment of the parameters is required or allowed to generate the predictions. We find it impressive that the predictions are so close to the observed data. However, this strategy requires an arbitrary and artificial distinction between behavioral and neural data. One is used to fit the model and the other is used to test its predictions about the dynamics of the units embodying the model. We are currently searching for methods that allow us to fit behavioral and neural data simultaneously, giving each equal weight in assessing goodness of fit (e.g., [24]). Those methods promise a true integration of mathematical psychology and neuroscience—a dream worth waking up to.

Exercises

1. In what sense does the stop signal paradigm measure response inhibition?
2. The independent race model addresses finishing time distributions without specifying the processes that generate the finishing time distributions. How is this an advantage and how is this a disadvantage.

3. The interactive race model does not describe neural activity at the beginning of the trial when the eyes are fixated. During this period, fixation cells are active and their firing rate is stable. After the target appears, activity in fixation-related neurons drops and activity in movement-related neurons increases. Do you think that including this fixation activity at the beginning of a trial in the modeling will change the models' predictions? How?
4. The spiking neuron model assumes a control process that removes inhibition on the go process to generate a response. What problems do you see with this assumption?
5. The cortico-basal ganglia model has not been tested with rigorous fits to data. Do you think it would fit well if such fits were attempted?

Further Reading

Logan and Cowan [13] is a seminal paper in stop-signal modeling. Everyone who works with the stop signal task should be familiar with this model and the approach.

Logan [12] is a "user friendly" introduction to the stop signal task that may be more accessible to novice readers than Logan and Cowan [13].

Boucher et al. [5] is the first model to bring computational modeling and neurophysiology together in the stop-signal paradigm and is worth reading for its place in history.

Verbruggen and Logan [26] provide a useful but brief review of recent research on the stop signal task. Verbruggen and Logan [27] provide a review of recent modeling work on the stop-signal paradigm.

Anything by Kurt Vonnegut Jr. or Robertson Davies.

Acknowledgement This work was supported by NIH grant R01EY021833

References

1. Aron AR, Duston S, Eagle DM, Logan GD, Stinear CM, Stuphorn V (2007) Converging evidence for a fronto-basal-ganglia system for inhibitory control of action and cognition. J Neurosci 27:11860–11864
2. Aron AR, Verbruggen F (2008) Dissociating a selective from a global mechanism for stopping. Psychol Sci 19:1146–1153
3. Band GP, van Boxtel GJ (1999) Inhibitory motor control in stop paradigms: revew and reinterpretation of neural mechanisms. Acta Psychol 101:179–211
4. Bissett PG, Logan GD (2013) Selective stopping? Maybe not. J Exp Psychol Gen (in press)
5. Boucher L, Palmeri TJ, Logan GD, Schall JD (2007) Inhibitory control in mind and brain: an interactive race model of countermanding saccades. Psychol Rev 114:376–397
6. Frank MJ (2006) Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. Neural Netw 19:1120–1136
7. Hanes DP, Schall JD (1995) Countermanding saccades in macaque. Vis Neurosci 12:929–937

8. Hanes DP, Patterson WF, Schall JD (1998) Role of frontal eye field in countermanding saccades: visual, movement and fixation activity. J Neurophysiol 79:817–834
9. Lo CC. Wang XJ (2006) Cortico-basal ganglia circuit mechanism for a decision threshold in reaction time tasks. Nature Neurosci 9:956–963
10. Lo CC, Boucher L, Paré M, Schall JD, Wang XJ (2009) Proactive inhibitory control and attractor dynamics in countermanding action: a spiking neural circuit model. J Neurosci. 29:9059–9071
11. Logan GD (1979) On the use of a concurrent memory load to measure attention and automaticity. J Exp Psychol Hum Percept Perform 5:189–207
12. Logan GD (1994) On the ability to inhibit thought and action: a users' guide to the stop signal paradigm. In: Dagenbach D, Carr TH (eds) Inhibitory processes in attention, memory, and language. Academic, San Diego, pp 189–239
13. Logan GD. Cowan WB (1984) On the ability to inhibit thought and action: a theory of an act of control. Psychol Rev 91:295–327
14. Logan GD, Van Zandt T, Verbruggen F, Wagenmakers EJ (2014) On the ability to inhibit thought and action: general and special theories of an act of control. Psychol Rev 121:66–95
15. Paré M, Hanes DP (2003) Controlled movement processing: superior colliculus activity associated with countermanded saccades. J Neurosci 23:6480–6489
16. Pouget P, Logan GD, Palmeri TJ, Boucher L, Paré M, Schall JD (2011) Neural basis of adaptive response time adjustment during saccade countermanding. J Neurosci 31:12604–12612
17. Purcell BA (2013) Neural mechanisms of perceptual decision making. Doctoral Dissertation, Vanderbilt University
18. Purcell BA, Heitz RP, Cohen JY, Schall JD, Logan GD, Palmeri TJ (2010) Neurally constrained modeling of perceptual decision making. Psychol Rev 117:1113–1143
19. Purcell BA, Schall JD, Logan GD, Palmeri TJ (2012) From salience to saccades: multiple-alternative gated stochastic accumulator model of visual search. J Neurosci 32:3433–3446
20. Ratcliff R, Smith PL (2004) A comparison of sequential sampling models for two-choice reaction time. Psychol Rev 111:333–367
21. Salinas E, Stanford TS (2013) The countermanding task revisited: fast stimulus detection is a key determinant of psychophysical performance. J Neurosci 33:5668tb–5685
22. Schall JD (2004) On building a bridge between brain and behavior. Annu Rev Psychol 55:23–50
23. Teller DY (1984) Linking propositions. Vis Res. 24:1233–1246
24. Turner BM, Forstmann BU, Wagenmakers EJ, Brown SD, Sederberg PB, Steyvers M (2013) A Bayesian framework for simultaneously modeling neural and behavioral data. Neuroimage 72:193–206
25. Usher M, McClelland JL (2001) The time course of perceptual choice: the leaky, competing accumulator model. Psychol Rev 108:550–592
26. Verbruggen F, Logan GD (2008) Response inhibition in the stop-signal paradigm. Trends Cogn Sci 12:418–424
27. Verbruggen F, Logan GD (2009) Models of response inhibition in the stop-signal and stop-change paradigms. Neurosci Biobehav Rev 33:647–661
28. Wiecki TV, Frank MJ (2013) A computational model of inhibitory control in frontal cortex and basal ganglia. Psychol Rev 120:329–355