WHAT COMPUTATIONS SUPPORT RESPONSE INHIBITION? THE ROLES OF CONTEXT-MONITORING, SELECTIVE STOPPING, AND STRATEGIC CONTROL

by

CHRISTOPHER H CHATHAM

B.A., University of Pennsylvania, 2002

M.A., University of Colorado, Boulder, 2007

A dissertation submitted to the

Faculty of the Graduate School of the

University of Colorado in partial fulfillment

of the requirement for the degree of

Doctor of Philosophy

Department of Psychology and Neuroscience

2011

This dissertation entitled:
What computations support response inhibition? The roles of context-monitoring, selective stopping, and strategic control
written by Christopher H. Chatham
has been approved for the Department of Psychology and Neuroscience

_____
Yuko Munakata

_____
Marie T. Banich

_____
Tim Curran

_____
Randy C. O'Reilly

_____
Francois Meyer

Date: _____

The final copy of this thesis has been examined by the signatories, and we
Find that both the content and the form meet acceptable presentation standards
Of scholarly work in the above mentioned discipline.

IRB protocol #s 0507.2, 0307.2, 10-0330

Chatham, Christopher H. (PhD Candidate, Cognitive Neuroscience)

What computations support response inhibition? The roles of context-monitoring, selective stopping, and strategic control

Dissertation directed by Professor Yuko Munakata

Response inhibition is thought to be central to the function of the frontal lobes and intimately related to all higher cognitive functions. The effortful and controlled component to response inhibition is often assumed to be the act of motoric stopping per se, such that motoric stopping can be strategically directed either at all actions (thus taking a global form) or only a subset (thus taking a more selective form). Here we challenge these core assumptions, first by offering a more viable alternative to dominant accounts of global stopping, and second by raising both empirical and computational dilemmas for dominant accounts of selective stopping. Using behavioral, neuroimaging, electroencephalographic and computational approaches, we demonstrate that the global stopping of responses does not require control mechanisms beyond those involved in detecting contexts in which old behaviors are inappropriate – in other words, the ability to monitor context in the service of goals. A computational model is used to demonstrate that these context-monitoring processes may underlie the behavioral, neuroimaging, and pharmacological phenomena of global stopping in the absence of any controlled stopping mechanisms. This work also challenges claims that more strategically selective forms of stopping are enabled by the controlled use of a slower but more response-specific stopping mechanism. We find a developmental double dissociation between the speed of stopping and its specificity, indicating that they cannot unambiguously be taken to reflect the use of a single neural mechanism. A computational model raises further challenges for extant theories of

selective stopping, raising the possibility that the control mechanisms supporting this ability are not actually directed at specific responses. Together these results challenge dominant accounts of both global and selective inhibitory control, and clarify emerging debates on the frontal substrates of response inhibition, while raising novel questions about the substrates and processes that support selective stopping. More broadly, this work provides a coherent account of prefrontal cortex function across inhibitory domains, which may in turn enable both a more precise characterization of frontal disinhibitory syndromes as well as enable more targeted diagnosis and treatment of pathologies associated with response inhibition deficits.

For Mom and Dad

ACKNOWLEDGEMENTS

This dissertation is the combined effort of innumerable people. My advisor, Yuko Munakata, has not only trained me in science but also provided the initial theoretical insights that are the foundation of the current work. The processes supporting response inhibition, as I have learned, are exceedingly complex. Her ineffable poise is a continuing personal inspiration for me, and her lucidity of thought will always be thoroughly humbling.

Other faculty members have also been vital. Marie not only trained me in functional magnetic resonance imaging, but was also central in developing the much needed clarification to the theoretical ideas described here. Tim trained in me in event-related potentials, but his unwavering focus on the atomic core of science – what, and only what, the data say – transcends all techniques. Randy's computational models have also had a major influence on my work, though his indefatigable optimism and creativity is the perfect model. I am honored to have had the opportunity to work with these magnificent people, an opportunity which was made possible only by Yuko's decision to admit me as a student. I'm still not entirely sure how that happened, but I am deeply thankful for being given this chance.

My friends have also played a crucial role in this dissertation. Many helped me to hone both my hypotheses and methods, but all have provided me with invaluable emotional support when I needed it most. I owe a great debt to Seth Herd, Angela Brant, Luka Ruzic, Colin Innes, Jon Nagel, and Allen Milletics in particular. I intensely admire you all.

Finally, my parents – to whom this work is dedicated – have both mercifully buoyed me up (arguably even when I didn't deserve it) and helpfully prodded me along. Their unconditional love and unending encouragement are the unshakeable foundations on which all of this is built.

CONTENTS

CHAPTER 1: CONTEXT MONITORING AND RIGHT VENTROLATERAL PREFRONTAL CORTEX 1
Abstract 1
Introduction 2
Process Impurity in Measures of Response Inhibition 3
Our Proposal: The Context Monitoring Hypothesis 6
Context-Monitoring In Models of Response Inhibition 7
The Importance of Behavioral-Relevance 9
The Importance of Contextual-Frequency 12
Evidence from Psychopathological Disorders and Neural Insult 17
Evidence from Psychopharmacological Manipulations 19
Red Herrings 23
Genuine Outstanding Challenges 27
Novel Predictions 30
Conclusions 31
Footnotes 32
CHAPTER 2: COGNITIVE CONTROL REFLECTS CONTEXT MONITORING, NOT MOTORIC STOPPING, IN RESPONSE INHIBITION 34
Abstract: 34
Introduction 35
Materials and Methods 40
Participants. 40
Behavioral Task. 40
Statistical Analysis of fMRI 41
Statistical Analysis of ERPs. 43
Statistical Analysis of Pupillometry. 44
Statistical Analysis of Behavior – Double Go Task. 44
Statistical Analysis of Behavior – Stop Task. 46
Results 47
Univariate fMRI Results 47
Multivariate Pattern Analysis 49
Event-related potentials 51
Pupillometry 52

Model-based decomposition of behavior and correlations with brain activity 53
Discussion.....59
Relation to Recent Work. 60
Broader Implications 64
CHAPTER 3: A NEUROCOMPUTATIONAL MODEL OF MONITORING AND STOPPING IN RESPONSE INHIBITION 67
Abstract.....67
Introduction 67
Methods 74
Implementation. 74
Training & Testing. 76
Results 78
Discussion.....89
Conclusion 96
CHAPTER 4: STRATEGIC STOPPING AND FOREKNOWLEGDE IN RESPONSE INHIBITION 97
Abstract.....97
Introduction 97
Methods 102
Participants. 102
Choice RT. 102
Global Stop. 103
Double RT. 103
Uncued Selective Stop. 104
Cued Selective Stop. 104
Trimming and Preprocessing. 104
Results 107
Discussion.....111
Conclusions 116
CHAPTER 5: SELECTIVE STOPPING AND COMPUTATIONAL CONUNDRUMS 117
Abstract.....117
Introduction 117
The problem of interference in selective stopping 119
Methods 122

Implementation. 122
Training & Testing. 124
Results 125
Discussion 135
Conclusions 139
CHAPTER 6: SUMMARY AND CONCLUSIONS 141
Implications of Context Monitoring: Broader Domains 145
Monitoring of the Environment 146
Monitoring of Language Processes 149
Monitoring of Memory 150
Implications of Context-monitoring: Frontal Lobe Damage and Psychopathology 152
Implications of Context-Monitoring: Targeted Interventions 154
Final Remarks 156
REFERENCES 157
APPENDIX I. SUPPORTING INFORMATION FOR CHAPTER 2 175

FIGURES

Figure

1. The importance of processes that support stop signal detection……4

2. Design of a context-monitoring task……..37

3. Hybrid fMRI results from the context monitoring task…….48

4. Multivariate pattern analyses……..50

5. Event-related potentials …….51

6. Patterns of mental effort ……53

7. Response slowing and behavioral model of context-monitoring……..55

8. Extensions to PBWM used to simulate context-monitoring…….72

9. Inputs to PBWM, and its outputs……78

10. Simulated benchmark phenomena from the Stop task……..80

11. Post-signal and post-error slowing…….83

12. Effects of stop signal frequency on the model……..84

13. Demonstration of the crucial nature of prefrontal layers……..85

14. Potentiation of inhibitory control with simulated atomoxetine……..87

15. Activation dynamics of output and subthalamic layers……88

16. Weight analyses as evidence for a domain-general prefrontal function …89

17. The developmental study on selective stopping……106

18. Accuracy on Stop Signal trials across tasks and ages……108

19. Reaction times in the Selective Stop tasks…….109

20. Typical developmental effects on global SSRT……..110

21. A counterintuitive effect of foreknowledge……111

22. Architecture of a selective stopping model……121

23. Inputs and outputs of this selective stopping model……123

24. Selective stop model: accuracy and reaction time phenomena…….125

25. Nonindependence of stopping and going in the selective stop model……..126

26. The effects of foreknowledge in the model……127

27. Reaction time distributions of the model and adult subjects…….129

28. Reaction time distributions of the model early in training……..131

29. Alternative architectures of the model……133

30. Reaction time distributions of humans in selective stopping……..135

CHAPTER 1: CONTEXT MONITORING AND RIGHT VENTROLATERAL PREFRONTAL CORTEX

Abstract

Response inhibition is widely considered to be a critical executive function deployed in the service of cognitive control, and is thought to crucially rely on the integrity of the right ventrolateral prefrontal cortex (rVLPFC). However, the paradigms thought to assess response inhibition typically involve both the actual stopping of responses and the requirement to detect and interpret behaviorally significant stimuli. Thus, experimental tasks which measure response inhibition may also assess those *context-monitoring* processes which enable the detection of stimuli that may demand changes in behavior given contextual factors. Here we provide an integrative review of the literatures on response inhibition and rVLPFC by contrasting two competing accounts. Dominant "motoric stopping" accounts posit that responses are inhibited via stopping-specific processes localized to the rVLPFC, while context-monitoring accounts posit that responses are inhibited via more domain-general processes in the rVLPFC. We evaluate these competing accounts in the context of a broad range of relevant electrophysiological, hemodynamic, and neuropsychological data. On the basis of our critical review, we conclude that context-monitoring theories are viable accounts of phenomena widely assumed to reflect stopping in the response inhibition literature, and in some cases are more consistent with available data. However, neither account can explain all existing phenomena, potentially pointing the way to a more integrative and comprehensive account of rVLPFC function.

Introduction

The ability to stop certain unwanted thoughts or actions is a widely recognized cognitive capacity, with numerous subtypes and with theoretical importance to a number of domains (Aron et al., 2007). Dysfunction in such inhibitory abilities has been invoked to explain a number of neuropsychological disorders, including attention deficit disorder, borderline personality disorder, obsessive-compulsive disorder, and post-traumatic stress disorder (Barkley, 1997; Chambers, Garavan, & Bellgrove, 2008; Falconer et al., 2008; Groman, James, & Jentsch, 2008; Hertel, 2007; Rentrop et al., 2008). Similarly, different inhibitory subprocesses have been used to characterize the functions of disparate regions of the prefrontal cortex (Aron, Robbins, & Poldrack, 2004; Diamond, 1990; Roberts & Wallis, 2000) and to be the principal set of capacities enabled by maturation of the prefrontal lobes (Dempster, 1992). It is hard to exaggerate the influence that these constructs have had on cognitive theorizing.

A wealth of evidence has been interpreted to reflect an overarching role for response inhibition in daily life. In healthy adults, factor analyses of individual differences indicate the variance shared by tasks requiring response inhibition fully overlaps with that shared among all tasks requiring the control of thought and behavior (Friedman et al., 2008). In agreement with this view, the neural circuitry thought to support response inhibition is intricately linked to the loops between prefrontal cortex, thalamus and basal ganglia that are involved in higher-level cognition (Aron, 2007). While some theorists have questioned the existence of resource-demanding and cognitively-controlled inhibition at the cognitive and conceptual levels (Munakata, 2001), the inhibition of physical responses has sometimes been specifically exempted from these criticisms (MacLeod et al., 2003). This exemption, in turn, has grounded

continued theorizing that the mechanisms of response inhibition may somehow apply more widely (Aron, 2007).

Process Impurity in Measures of Response Inhibition

Response inhibition is a difficult construct to measure – in part, because successful response inhibition will often lead to no measurable response at all. Such difficulties can be circumvented by the use of the Stop Signal task (Logan & Cowan, 1984) in which subjects must simply respond to an imperative stimulus unless it is followed by a "stop signal." The stop signal indicates that subjects must *not* execute any response on that trial. By using subject performance to adapt the delay between the imperative stimulus and the stop signal that occasionally follows it, one can experimentally determine the delay at which any given subject is just as likely to successfully inhibit a response as to erroneously commit a response (i.e., 50% accuracy). According to the canonical "race model" of this task (Logan & Cowan, 1984), 50% accuracy reflects the case where a "stop process," initiated by the onset of the stop signal, is finishing on average just as quickly as the "go process" initiated by the onset of the preceding imperative stimulus. The duration of this otherwise-unobservable "stop" process can thus be estimated utilizing straightforward subtractive logic. The resulting measure is termed the "Stop Signal Reaction Time" (SSRT).

The race model of the stop signal task has been incredibly useful for deriving this quantitative estimate of the efficiency of response inhibition. However, it is important to note that SSRT includes both the efficiency of motoric stopping and the efficiency and reliability of triggering that motoric stopping process in the first place (Figure 1A and B illustrate this phenomenon). For example, one might be exceptionally good at cancelling ongoing or planned responses, but exceptionally poor at noticing or acting upon the brief or rare stop signals. In that

case, SSRT might be interpreted to reflect poor response inhibition, but for the wrong reason – the impairment lies in noticing and processing the stop signal, not in the cancellation of a response *per se*. Thus, good inhibitory performance requires not only the ability to stop the planned or ongoing response, but also the ability to detect and interpret the infrequent but task-relevant stop signal itself. These capacities are inherently intermingled in the use of SSRT.

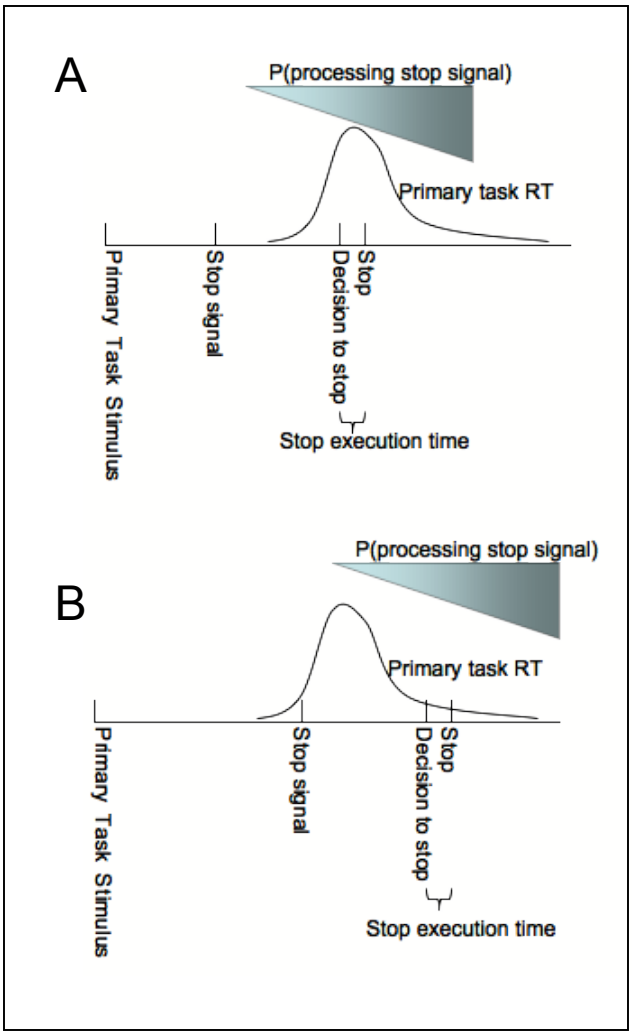


Figure 1. Processing and detecting the stop signal can explain the variability in stopping success with signal delay as illustrated above. A. When stop signals are provided early, the detection of the stop signal is more likely to complete before the response is committed, enabling successful inhibition. B. Conversely, the primary task response is more likely to have occurred by the time later stop signals are fully processed, yelding less successful inhibition.

While this point might at first glance seem to be a kind of "storm in a teacup," it could in fact have substantial theoretical and applied implications. For example, groups that demonstrate prolonged SSRT relative to healthy adult controls are often interpreted to have poor response inhibition abilities – when their deficits might have nothing more to do with response inhibition than they have to do with response commission. Such these difficulties might instead lay in effectively detecting or interpreting the stop signals. Likewise, areas of the brain whose function is correlated with SSRT (most notably, the right ventrolateral prefrontal cortex; rVLPFC) have often been interpreted to instantiate the act of motoric stopping itself, when their function might possibly have rather little to do with motor stopping and much more to do with effective detection and interpretation of the infrequent, but behaviorally-relevant stop signals. Dissociating between motoric stopping and these more attentional processes is thus not only important for the basic mission of cognitive neuroscience, but could be highly informative for the diagnosis and treatment of psychopathologies that are currently characterized by deficits in response inhibition.

In this article, we review computational, neuroimaging, and neuropsychological evidence which suggests that the widely used Stop task and the closely related Go/NoGo task may primarily reflect what we term *context monitoring* processes – processes which enable the detection of stimuli that might indicate the need for a change in behavior given the current context, regardless of whether those stimuli demand motoric stopping *per se* (Mars, Piekema, Coles, Hulstijn, & Toni, 2007; Ray, Huang, Constable, & Sinha, 2006; Stuss & Alexander, 2007; Sharp et al., 2010; Hampshire et al., 2010; Dodds et al, 2011; Cai & Leung et al, 2011; Verbruggen et al., 2010; Chatham et al., submitted). We describe how this hypothesis can explain the contributions of rVLPFC to a variety of tasks thought to measure response inhibition.

Finally, we summarize the conceptual contributions and novel empirical measures proposed by this view, along with the novel theoretical and empirical insights it offers.

Our Proposal: The Context Monitoring Hypothesis

We will use the term *context monitoring* to refer to the cognitive capacity to detect, attend to, and interpret *behaviorally-significant stimuli*, particularly those that are *contextually-infrequent* (Ranganath & Rainer, 2003) or otherwise unexpected (e.g., Horstmann, 2006). We propose that such context-monitoring is the critical executive function tapped in putative measures of response inhibition, and that it rather than motoric stopping is more consistent with psychopathological deficits in response inhibition tasks as well as the changes in inhibitory control that are effected by psychopharmacological manipulations. We review five types of evidence in support of this context-monitoring hypothesis:

1. Formal modeling work that indicates the central importance of context-monitoring in the way SSRT is measured, and in the way that response inhibition phenomena may emerge from underlying neuronal computations;
2. Emerging empirical work which indicates the importance of *behavioral-relevance* to the processes subserved by rVLPFC, irrespective of motoric stopping demands
3. Empirical work which indicates the importance of *contextual-frequency* to the processes subserved by rVLPFC, irrespective of motoric stopping demands
4. Evidence from psychopathological disorders and neural insult which demonstrates that apparent changes in response inhibition are uniformly paralleled by the electrophysiological changes predicted by our hypothesis;

5. Evidence from pharmacological manipulations which demonstrate that apparent changes in response inhibition are uniformly paralleled by changes in context-monitoring.

We conclude with a discussion of the evidence that is sometimes thought to argue against this hypothesis, with special attention to the subset which does in fact pose a challenge. Because this evidence also goes uniformly unexplained by extant motoric stopping accounts, we conclude with a careful evaluation of these unexplained findings as a direction for future work on rVLPFC and the cognitive neuroscience of response inhibition.

Context-Monitoring In Models of Response Inhibition

There are numerous computational reasons to believe that context-monitoring, and not motoric stopping *per se,* could be central to the measurement of response inhibition in the Stop task. For example, extensive monte-carlo simulations of the race model (Band et al., 2003) included the possibility that subjects may categorically fail to trigger a stop process on a small proportion of signal trials (with a probability of less than or equal to 25%), as might occur through failures to detect the Stop Signal. Even these conservative estimates of the likelihood of "triggering failures" nonlinearly inflated estimates of SSRT.

Failures to detect the Stop signal might be understood as a kind of goal neglect, in which subjects fail to maintain vigilance for the stop signal. Goal neglect, and its conceptual converse, vigilance or goal maintenance, are central to current theories of the structure of executive functions (e.g. Friedman et al, 2009) and explain numerous phenomena that might otherwise be ascribed to various forms of inhibition (De Jong, Berendsen, & Cools, 1999; Kane & Engle, 2003; Munakata, 2001; J. R. Reynolds, Braver, Brown, & Van der Stigchel, 2006; Towse, Lewis, & Knowles, 2007), even in the Stop task (e.g.., the increased activation of "default mode"

regions prior to failed stop trials, among other effects; Li, Yan, Bergquist, & Sinha, 2007; Logan, 1981; Logan & Burkell, 1986; Verbruggen & Logan 2009; Bissett & Logan, 2010). However, If the goal to detect the stop signal is instantiated not in an all-or-none, but rather in a graded fashion (Munakata, 2001), then it may be inappropriate to view such goal neglect as a completely categorical phenomenon.

Instead, one might expect that detection of the signal itself would involve a goal-driven process that unfolds gradually across time, such that detection of the signal must first cross some threshold before "stopping" *per se* can occur. Ironically, stopping – not triggering – may be the more discrete and temporally delimited process, such that a relatively automatic and rapid "circuit breaker" mechanism is engaged whenever contextual changes indicate a need for behavioral control. The idea that motoric stopping *per se* is a rapid, global, and automatic operation is supported by the substantial response interference observed when only one of two responses must be stopped, and even when subjects are forewarned which of those responses will need to be stopped (Aron & Verbruggen, 2009; Coxon, Stinear, & Byblow, 2007). Thus it is possible that motoric stopping is not engaged in a very selective or controlled fashion, but rather is automatically engaged even when such subjects know that such global inhibition will be inappropriate[1].

Further support for the importance of stop signal monitoring and detection comes from a two-unit neural network implementation of the race model (Boucher, Palmeri, Logan, & Schall, 2007). This model indicates that the "stimulus encoding" of the stop signal can take twice as long as the other components of stopping ("interruption itself" and "release from inhibition"), and that the duration of these latter components (22ms) approaches the latency of synaptic integration time. Thus the actual stopping of go processes once stop signal processing has

completed “can be considered effectively instantaneous” (Boucher et al., 2007, p391) It is questionable whether such duration-limited processes related to stopping *per se* (independent of stimulus detection and encoding) could drive the large individual differences in SSRT, which are regularly around 200ms in magnitude. Instead, differences in the rate of stimulus detection and encoding processes may drive these individual differences.

A Bayesian diffusion model implementation of the classical race model similarly points to an important role for processes like context monitoring (Shenoy & Yu, 2011). In this model, a monitoring process continually assesses the accumulation of sensory information about the stimuli present on any given trial with respect to prior beliefs about the prevalence of Go and Stop signals, to estimate a posterior probability that the current trial is a Stop trial. This monitoring process is complemented by a second (and Bayes optimal) process which selects among possible actions: either it chooses to respond, or it chooses to wait longer in case a Stop signal may appear. The only model parameters which are explicitly free to vary among individuals pertain to the accumulation of noisy sensory information. As such, this model explicitly posits that variability in monitoring processes would underlie most of the variance in a response inhibition task like the Stop task.

The Importance of Behavioral-Relevance

Recent fMRI and TMS studies support the aforementioned computational arguments that context-monitoring processes may be more central to the Stop task than motoric stopping *per se*, particularly with respect to the issue of whether the infrequent stimuli are behaviorally-relevant. For example, the rVLPFC actually more strongly recruited in response to rare behaviorally-relevant stimuli that indicate subjects must commit, rather than stop, a response (Hampshire et al., 2010). Similarly, the rVLPFC is more strongly recruited during tasks in which a rare

behaviorally-relevant stimulus indicates that subjects must commit an *additional* response rather than stop a response (Dodds et al., 2011; Chatham et al., under revision). Temporary deactivation of the rVLPFC via TMS yields a deficit in the ability to quickly commit these additional responses; the magnitude of this effect is equivalent to the deficit in SSRT that is yielded by the same procedure (Verbruggen et al., 2010). These studies clearly indicate that recruitment of the rVLPFC is not specific to the motoric stopping demands of the Stop task, and thus that rVLPFC function should not be unequivocally interpreted as reflecting the motoric stopping demands required during response inhibition tasks.

Significantly different results have been achieved from studies that assess the recruitment of the rVLPFC during tasks that present rare but behaviorally-*irrelevant* stimuli (i.e., stimuli that indicate subjects must merely "continue" with a planned or prepotent response, rather than to stop or to commit an additional response). In these studies, the rVLPFC is less strongly recruited by infrequent and task-irrelevant stimuli than by infrequent stimuli that are task-relevant by virtue of indicating a need to stop (Chikazoe et al., 2009; Chikazoe, Konishi, Asari, Jimura, & Miyashita, 2007; Ramautar et al., 2006; Rubia et al., 2001; Barch, et al, 1997; Sharp et al., 2010; Dodds et al., 2011). Two of these studies have found slight differences in the precise localization of these responses, such that a more dorsal region of the rVLPFC (the inferior frontal junction, or rIFJ) was more responsive to these "continue" stimuli than to stimuli that demanded motoric stopping; motoric stopping demands were instead more associated with the *pars opercularis* (Brodmann Area 44) subregion of rVLPFC (Chikazoe et al., 2009; Cai & Leung, 2011). However, work with transcranial magnetic stimulation (TMS) indicates neither of these possible subdivisions of rVLPFC is functionally specific to motoric stopping (Verbruggen et al., 2010); If there is indeed such a functional distinction, this work suggests that it would instead lie

between detecting infrequent stimuli (as subserved alone by a region dorsal to the rVLPFC, the right inferior frontal junction; rIFJ) and further processing of such stimuli that are also relevant for behavior (as subserved by rVLPFC proper, in particular the *pars opercularis* and the adjoining anterior insula; Verbruggen et al., 2010)

Collectively, this evidence would appear to refute all claims that the rVLPFC is functionally specific to demands on motoric stopping. Upon closer inspection, however, these results are not so clear-cut: subjects may in fact engage a motoric stopping process even when this is not strictly required by a given task. Indeed, according to the context-monitoring hypothesis we advanced above, such motoric stopping is actually an automatic consequence of the outcome of context-monitoring processes.

When the presence of motoric stopping has been examined within tasks with rare and behaviorally-relevant stimuli that do not demand stopping, it has been uniformly detected. Specifically, the presentation of "continue" stimuli leads to interference on these to-be-committed responses in both the Stop task (Sharp et al., 2011; Cai & Leung, 2011) and the closely related NoGo task (Chikazoe, 2009). This slowing is from one perspective unsurprising: subjects must first evaluate this infrequent stimulus, to determine whether it is a "continue" stimulus or a stimulus that demands motoric stopping. Subjects may engage motoric stopping while such evaluation takes place.

Other studies suggest that this motoric stopping is not limited to cases in which stimulus evaluation of this kind must take place. For example, slowing of responses is also observed following the presentation of rare behaviorally-relevant stimuli that require not only a planned or prepotent response must be committed but also that an additional response must be provided (Verbruggen et al., 2010; also Chapter 2). This occurs even when there is no possibility that this

stimulus will demand the subject to stop their planned motor action, and even when subjects are explicitly informed to minimize any possible delay in their actions (Chatham et al under revision).The presence of this slowing suggests that subjects may in fact engage motoric stopping processes, even when an abstract task analysis would suggest that motoric stopping is unnecessary and undesirable.

On the other hand, this slowing does not appear to be strongly related to the functional recruitment of the rVLPFC. A between-subjects contrast of those who undergo substantial slowing and those who do not reveals a correlation with activity in the pre supplementary motor area (pre-SMA; Sharp et al., 2010), but not with rVLPFC (Sharp et al 2010; also Chatham et al., under revision). Within-subjects analysis of the precise duration of this motoric slowing also fails to reveal any correlation with activation in the rVLPFC, but shows a trend towards a correlation with activity in the subthalamic nucleus (Chatham, Claus, Kim, Curran, Banich & Munakata, 2011). Thus motoric stopping does not seem to be particularly related to the function of the rVLPFC. To the extent motoric stopping occurs automatically, it may also not play a very central role in the cognitive demands of response inhibition tasks.

The Importance of Contextual-Frequency

The processing of contextually-infrequent stimuli has long been studied in the oddball paradigm. Oddball paradigms all share a common feature: They expose subjects to a sequence of stimuli of varying probability, including a standard stimulus (occurring with 60% to near 100% probability), and one or more deviant stimuli (each occurring with lower probability). Oddball paradigms differ in the discrepancy between standards and deviants (in terms of probability, perceptual deviance, and other factors like modality) as well as in terms of which (if any) stimuli require attention or responses. Despite these considerable variations, oddball tasks

consistently elicit one or more event-related potentials (ERPs) from a collection of ERPs termed the "N2-P3 complex."

The N2-P3 complex consists of at least three dissociable components: a negative peak over anterior electrodes between 200 and 350 ms after stimulus onset (the N2), a subsequent low-amplitude positive peak over frontocentral electrodes occurring within roughly the same time frame (the novelty P3, or P3a), and a large positive peak maximal at centroparietal electrodes after 300ms (the P3b). The N2 is itself sometimes further dissociated into as many as three separate components even within a single modality (Folstein & Van Petten, 2008), and dipole source modeling indicates that medial frontal regions are responsible for its generation. We will not focus here on the N2 complex, because it is recognized to reflect processes other than motoric stopping, irrespective of the task used to elicit it (Donkers & Van Boxtel, 2004; Nieuwenhuis et al., 2003; Azizian et al., 2006)

In contrast, interpretations of the P3 component have been task-dependent. P3 components recorded during tasks requiring response inhibition (the so-called "Stop P3" and "No Go P3") have been interpreted to reflect motoric stopping processes *per se*. However, there are extensive similarities between these P3 components and those recorded in oddball paradigms without concomitant demands on motoric stopping. P3 components as elicited by such oddball paradigms have identified a source in the rVLPFC (as well as the adjoining anterior insula; Bubic, von Cramon, Jacobsen, Schroger, & Schubotz, 2008; Doeller et al., 2003; Kiehl, Laurens, Duty, Forster, & Liddle, 2001; Linden, 2005; Linden et al., 1999; Mulert et al., 2004; Opitz, Rinne, Mecklinger, von Cramon, & Schroger, 2002; A. Stevens, 2000), the same region interpreted to be specialized for the stopping demands of response inhibition. However, numerous theorists from other domains have linked this area to the exogenously-triggered

orienting of attention (Corbetta, Patel, & Shulman, 2008; Corbetta & Shulman, 2002; Ranganath & Rainer, 2003) and some to monitoring in particular (Stuss & Alexander, 2007).

The individual components that compose the P3 complex have been cleanly dissociated with the use of spatiotemporal principal components analysis (Spencer, Dien, & Donchin, 2001). This technique yields spatiotemporally uncorrelated "virtual" ERPs, at least two of which have a predominantly frontal and right-lateralized topography. These virtual ERPs are elicited a) in response to rare stimuli, b) when the perceptual features of those rare stimuli are highly deviant from the standard, and c) when those stimuli are task-relevant (relative to conditions where they serve as distractors). In other words, they precisely fit the characteristics of the hypothetical context-monitoring capacity introduced above, especially with respect to issues of contextual frequency (a and b, above).

Manipulations of stop signal frequency typically show increased rVLPFC activity, or increased contributions of ventral prefrontal areas to the P3, with increasingly rare stop or NoGo trials (Ramautar, Kok, & Ridderinkhof, 2004; Ramautar, Slagter, Kok, & Ridderinkhof, 2006),[2] consistent with a role for the rVLPFC in detecting rare stimuli. One might assume that such effects reflect increasing difficulty with stopping an increasingly prepotent motoric response; however, such intuitions are apparently not borne out by empirical data, which demonstrates no change in SSRT as a function of stop signal frequency (Logan & Cowan, 1984). As such, the increased rVLPFC response to rare Stop signals is apparently unexplained by accounts which posit the centrality of motoric stopping and rVLPFC to response inhibition processes.

Even when stimulus probability is matched, oddball effects are enhanced to stimuli that are more perceptually deviant relative to those that are less deviant (Sawaki & Katayama, 2008). Thus, an oddball effect in the rVLPFC might be evoked to a NoGo stimulus (e.g., "A") if it is

less similar to the fixation stimulus used on every trial (e.g., +) than an equiprobable Go stimulus (e.g., X). This raises the possibility that perceptual confounds could underlie the greater rVLPFC activity observed to NoGo stimuli in several studies that do not counterbalance the identity of equiprobable stimuli (Kaladjian et al., 2007; Liddle, Kiehl, & Smith, 2001; Mazzola-Pomietto, Kaladjian, Azorin, Anton, & Jeanningros, 2009) – evidence previously taken to support a specific role for the rVLPFC in the motoric stopping demands of response inhibition tasks. Moreover, P3 components are exquisitely sensitive even to minute differences in stimulus characteristics: P3 components are elicited by tones differing as little as .8% in frequency, even when the average discrimination threshold is only .73% (Tervaniemi, Just, Koelsch, Widmann, Schroeger, 2004). Similarly, some estimates of area discrimination thresholds are as high as 20% (Morgan, 2005), and yet even 24% differences in area yield a P3 component (Sawaki & Katayama, 2008). Given this extreme sensitivity to perceptual characteristics, it is very important to counterbalance stimuli to control for potential perceptual confounds.

Independent of stimulus probability and perceptual deviance, certain events may also be contextually-infrequent by virtue of their response demands. Some studies have controlled for this factor by requiring motoric stopping on fully 50% of trials, under the assumption that any differences between stopping and responding could not therefore reflect differences in response frequency. However, several of these studies "pre-train" subjects on tasks involving the same stimuli but without the demand for motoric stopping – meaning that subsequent motoric stopping demands are infrequent in the larger context of the experiment (Kok, Ramautar, De Ruiter, Band, & Ridderinkhof, 2004; Ramautar et al., 2004). This contextual-infrequency may be sufficient for engaging rVLPFC: Activity in this region is not observed in equiprobable response inhibition

paradigms when such pretraining is omitted (as described by Konishi, Nakajima, Uchida, Sekihara, & Miyashita, 1998).

Contextual frequency also has an influence at a shorter time-scale; for example, some tasks require that subjects inhibit their responses to certain stimuli only if they follow another particular stimulus (Mostofsky et al., 2003). A recent metanalysis indicates that rVLPFC activity can be predicted across Go/NoGo studies on the basis of this factor alone (Simmonds, Pekar, & Mostofsky, 2008).

Finally, events may also be contextually-infrequent if they differ markedly in difficulty or priority from other events, disregarding any modulations of frontal activity due to task difficulty itself, which may be particularly large in the rVLPFC (Barch et al., 1997). Task difficulty is a near-universal problem for studies of response inhibition: error rates are typically much higher on trials that require inhibition than those that do not. Intermixing response inhibition and infrequent response commission trials in the same task may thus induce a strategic prioritization of inhibition (Morein-Zamir, Chua, Franks, Nagelkerke, & Kingstone, 2007), further complicating the interpretation of studies which intermix trials that differ in inhibitory demands.

These confounds are not unique to Stop signal and Go/NoGo tasks, but apply more widely to a number of paradigms thought to measure response inhibition. For example, rVLPFC involvement in the antisaccade task – in which subjects must putatively inhibit the tendency to look towards stimuli with rapid visual onset – is limited to those paradigms with infrequent antisaccade trials (Chikazoe et al., 2007; Ettinger et al., 2008; Hodgson et al., 2007). Likewise, rVLPFC involvement in the Stroop task is not unique to incongruent trials, but is also observed during other infrequent Stroop trial types (Melcher & Gruber, 2006;similarly, Milham, Banich & Barad, 2003 also find a small cluster of rVLPFC activation). In the Posner cueing paradigm,

invalid spatial cues – towards which subjects must putatively inhibit their spatial attention in order to detect the target in another location – also activate the rVLPFC, but this appears to be due to the fact that invalidly-cued trials are unexpected, as suggested by the similarity of the rVLPFC response to invalidly cued trials vs. unexpected targets (Forstmann et al., 2008; Vossel, Weidner, Thiel, & Fink, 2008).

In summary, the contextual-frequency of stimuli seems to be a determining factor in the functional recruitment of the rVLPFC, irrespective of motoric stopping demands. One dimension of contextual-frequency is stimulus probability: Stop signals as well as infrequent stimuli that do not require motoric stopping reveal increasing contributions of anterior prefrontal areas to the P3 complex or increasing recruitment of the rVLPFC. Another dimension of contextual-frequency is perceptual deviance: paradigms with equiprobable Go/NoGo stimuli tend to show increased rVLPFC responses or P3 components when stimulus identity is not counterbalanced, such that differences in perceptual deviance could be a confounding source of variance. Finally, stimuli can be contextually-infrequent for more endogenous reasons, for example because certain trial types are intrinsically more difficult, of a higher strategic priority than others, or are otherwise unexpected from the local temporal context. All of these factors seem important in determining rVLPFC activity and the associated prefrontal subcomponents of the P3, confirming the importance of contextual-frequency in the processes subserved by the rVLPFC.

Evidence from Psychopathological Disorders and Neural Insult

Disinhibition is thought to be a "core deficit" in a number of psychopathological disorders. However, many of these same disorders are also associated with deficits in monitoring (as assessed by responses to contextually-infrequent and behaviorally-relevant

stimuli that do not require inhibition). For example, delayed or reduced P3 components in oddball tasks have been demonstrated for attention deficit disorder (Barry, Clarke, & Johnstone, 2003), borderline personality disorder (Drake, Phillips, & Pakalnis, 1991), and post-traumatic stress disorder (Karl, Malta, & Maercker, 2006). All of these disorders are associated with less efficient stopping in terms of SSRT (Casada & Roache, 2005; Nigg, Silk, Stavro, & Miller, 2005), but the coincident abnormalities in the processing of infrequent stimuli indicate that such inhibitory deficits might be more parsimoniously interpreted as failures of a more general ability (as did Alderson, Rapport, Sarver & Kofler, 2008), such as context-monitoring.

Conversely, the context-monitoring hypothesis and motoric stopping accounts are both indefinite in identifying a deficit in obsessive-compulsive disorder (OCD). Some studies of those with OCD demonstrate superior monitoring (as reflected in earlier P3 components in oddball tasks Morault, Bourgeois, Laville, Bensch, & Paty, 1997), as well as more efficient stopping in the Stop Signal task (Krikorian, Zimmerman, & Fleck, 2004), while other studies demonstrate a deficit in monitoring (as reflected in a delayed or reduced P3 components; Sanz, Molina, Martin-Loeches, Calcedo, & Rubia, 2001) and impaired stopping in the Stop Signal task (Menzies, Achard, Chamberlain, Fineberg, Chen, Campo, Sahakian, Robbins & Bullmore, 2007). Given this inconsistency in inhibitory deficits among those with OCD, neither monitoring nor inhibition can be unambiguously considered a core deficit in this disorder.

Patients with focal brain damage also often show symptoms that are more consistent with deficits in monitoring than inhibition, so much so that inhibition has been omitted from recent revisions to taxonomies of executive function based on neuropsychological data (Stuss & Alexander, 2007). For example, in the Wisconsin Card Sorting Test, patients with damage to the right lateral PFC spontaneously revert to sorting cards by outdated sorting criteria (so-called

"loss of set" errors) even when explicitly informed of which criterion to use at the beginning of the task and of the fact that the criterion changes every ten trials. Because these "loss of set" errors all followed at least three correct trials under a single sorting criterion, and were tabulated separately from those errors resulting from sorting by the immediately preceding criterion, these patients have not simply failed to inhibit previous sorting criteria; instead they may be losing vigilance and responding randomly, or perhaps even preemptively changing their sorting criteria due to deficits in their ability to monitor for the *need* to change criteria.

The explicit task-setting from which these patients failed to benefit strongly reduces the "loss of set" errors observed among patients with damage to left lateral PFC, consistent with production/monitoring accounts of hemispheric asymmetry (Cabeza, Locantore, & Anderson, 2003). Monitoring deficits are also thought to be revealed by the inability of right lateral PFC patients to benefit from larger inter-stimulus intervals, their greater variability in performance even at the beginning of a block, and their lack of post-error adjustments to behavior (Stuss & Alexander, 2007).

Evidence from Psychopharmacological Manipulations

A wealth of pharmacological evidence is also consistent with the context-monitoring hypothesis: Stopping and context-monitoring efficiency (as indicated by SSRT and the P3 latency/amplitude recorded in oddball tasks without concomitant inhibitory demands, respectively) are affected by the same drugs in the same way. (Because Eagle, Bari & Robbins, 2008 have recently documented the effects of various neurotransmitters on SSRT, we will redescribe that evidence here only in sufficient detail to highlight the parallel findings from the P3 component).

Enhancement of adrenergic mechanisms tend to improve both context-monitoring and SSRT. For example, atomoxetine (a norepinephrine [NE] reuptake inhibitor) improves SSRT among many of those with ADHD (Chamberlain et al., 2006), and although it is unknown whether the P3 changes as a result of this treatment, response to atomoxetine treatment can be predicted based on the pre-treatment P3 profile of the patients (Sangal & Sangal, 2006). Atomoxetine is also associated with increased recruitment of the rVLPFC (Chamberlain et al., 2009). Methylphenidate (an NE and dopamine [DA] reuptake inhibitor) remediates stopping efficiency in ADHD (Aron, 2003) and also partially normalizes P3 latency (Winsberg, Javitt, Silipo, & Doneshka, 1993). Similarly, modafinil (a stimulant with widespread effects on NE, DA, and other neurotransmitter systems) improves SSRT (Turner, Clark, Dawson, Robbins & Sahakian, 2004) but can also remediate delayed P3s in humans (Sangal, Sangal, & Belisle, 1999).

In contrast, serotonergic mechanisms appear to minimally influence both monitoring and response inhibition. For example, the antiserotonergic drug methysergide does not appear to affect the latency or amplitude of the P3 (Meador, Loring, Davis, Sethi, Patel, Adams, Hammon, 1989), nor does acute depletion of the dietary precursor to serotonin, tryptophan (Ahvenien, Kähkönen, Pennanen, Liesivuori, Ilmoniemi, Jääskeläinen 2002). Similarly, serotonergic mechanisms do not appear to affect response inhibition: SSRT is unaffected by serotonin uptake blockage (Chamberlain, Mueller, Blackwell, Clark, Robbins & Sahakian, 2006) or by dietary challenge (Clark, Roiser, Cools, Rubinzstein, Sahakian, Robbins, 2005). In general, this evidence supports the idea that serotonin manipulations affect neither context-monitoring (as reflected in the P3) nor SSRT.

Surprisingly, given the theorized involvement of dopaminergic basal ganglia mechanisms in response inhibition (Aron & Poldrack, 2005), relatively few studies have investigated the possible effects of dopamine on response inhibition. Of those that do, SSRT is unaffected by both nonselective dopaminergic agonists and antagonists (for a review, see Eagle, Bari & Robbins, 2008). Context-monitoring mechanisms (again, as reflected in the P3) are similarly unaffected by dopamine, in terms of D2 dopamine receptor polymorphisms (Lin, Yu, Chen, Tsai, & Hong, 2001), D2 antagonists (Antal, Kéria, Bodis-Wollner & 1997), or L-dopa (Oranje, De Wied, Herman, Westenberg, Kemner, Verbaten, Kahn, 2006).

Only one study has found a dopaminergic affect on SSRT (Eagle et al., 2011), only did so by directly injecting D1 and D2 receptor antagonists into the dorsomedial striatum. Even so, this study failed to obtain any kind of monotonic dose/response curve. The results of this study are further difficult to interpret because Stop trial accuracy was uniformly above 85% - the upper limit at which SSRT can be reliably calculated (Band et al., 2003). One cautious conclusion is that any subcortical dopaminergic mechanisms uniquely recruited by response inhibition play little role in governing behavioral performance, relative to the much larger role played by context-monitoring and as reflected in the P3.

In contrast, prefrontal dopamine seems particularly important for both monitoring and stop signal performance. For example, the D4 receptor, expressed preferentially in prefrontal (relative to subcortical) regions (Mulcrone & Kerwin, 1997), is associated with novelty seeking (Schinka, Letsch & Crawford, 2002), with increased SSRT (Congdon, Lesch & Canli, 2008) and with poorer monitoring (as reflected in a lower amplitude P3; Vogel, Laucht, Furtado, Becker & Schmidt, 2006, although this is also not always consistent across studies; Demiralp, Herrmann, Erdal, Ergenoglu, Keskin, Ergen, Beydagi, 2007). Variants in the gene that codes for catechol-

O-methyl transferase (COMT; which also has a particularly strong influence on prefrontal dopamine) are similarly associated with consistent changes both in SSRT (Kramer et al., 2007) and the P3 in oddball paradigms without concomitant inhibitory demands (Gallinat et al., 2003). In sum, this evidence suggests that the effect of D4 receptors and COMT on both the P3 and SSRT may be due to the common influence of monitoring for contextually-infrequent and task-relevant stimuli, and provides further support for our claim of a prefrontal basis for monitoring.

Cholinergic mechanisms also affect context-monitoring and performance on the stop signal task in similar ways. Context-monitoring is impaired by the administration of the cholinergic antagonists scopolamine and benztropine mesylate (as reflected in an increased P3 latency; Meador, Potter, Davis, Sethi, Patel, & Adams, 1985), and improved by the acute cholinergic agonist nicotine (perhaps by affecting the earlier components of the P3 complex in particular; Polich & Criado, 2006, Edwards, Wesnes, Warburton, & Gale, 1985). Although there are no reports of the effects of anticholinergics on response inhibition, SSRT is improved with nicotine (Potter & Newhouse, 2008). In summary, pharmacological cholinergic manipulations show similar effects on context-monitoring and SSRT, suggesting again strong overlap in their neural mechanisms.

More commonly-used drugs (such as caffeine and alcohol) have less consistent effects, perhaps due to the widespread influences these drugs have on the different neurotransmitter systems, and the difficulties inherent in administering drugs that subjects are familiar with (i.e., the concomitant unblinding of subjects, as well any potential withdrawal due to abstinence). Nonetheless, previous reviews have identified several effects of acute exposure (Polich & Kok, 1995): Caffeine can increase P3 amplitude (with smaller effects on latency), and ethanol decreases P3 amplitude and increases its latency. Both of these drugs are also associated with

the concomitant changes in SSRT predicted by the context-monitoring hypothesis (de Wit, Crean, & Richards, 2000; Potter & Newhouse, 2008; Tieges, Snel, Kok, & Richard Ridderinkhof, 2009; but see Bekker, Bocker, Van Hunsel, van den Berg, & Kenemans, 2005[3]). Clearly many of these drugs affect more than just SSRT and context-monitoring abilities, but based on this pharmacological evidence, it seems that many of the same mechanisms that underlie monitoring (as reflected in the P3 component) may also drive performance in Stop Signal task.

Red Herrings

As reviewed above, the context-monitoring hypothesis offers a plausible and parsimonious account of a number of findings from a variety of domains and methods, including electrophysiology and neuroimaging, neuropsychology, psychopharmacology, and clinical psychopathology. It is at this point useful to dispense with several "red herrings" that may have contributed the absence of similar theoretical accounts in the literature.

First, context-monitoring accounts have been criticized as inconsistent with a long history of theorizing which dictates that bottom-up attentional capture (as might arise from contextually-infrequent stimuli) should be most reliant on posterior brain regions, whereas prefrontal regions like the rVLPFC should be involved in more purely "top-down" attentional processes. Relatedly, monitoring hypotheses seem too strongly stimulus-related to viably explain prefrontal function, which is thought to be an evolutionary outgrowth of the more response-related posterior frontal cortex. However, the context-monitoring hypothesis is not alone in contradicting these long-held views. The What-How model of prefrontal function (O'Reilly, 2010) also posits that ventral areas of prefrontal cortex may be particularly involved in stimulus-related processing; similarly, the ventral-dorsal attentional networks theory (Corbetta, Patel &

Shulman, 2008) posits that a ventral prefrontal network subserves the goal-contingent capture of attention. In this way, the context-monitoring hypothesis is not merely inconsistent with some historical perspectives; it is in precisely these same ways *consistent* with emerging and more recent views.

Specific concerns have also been raised about the association of context-monitoring with rVLPFC function. For example, it has been argued that monitoring accounts cannot explain the increased rVLPFC recruitment to successfully stopped trials, relative to unsuccessfully stopped trials. However, such evidence is also clearly consistent with the more pronounced use of context-monitoring processes (and thus heightened rVLPFC recruitment) on successfully-stopped trials.

Similarly, motoric stopping is induced by direct electrical macrostimulation of rVLPFC in both humans and non-human primates (Luders et al., 1988; Sasaki et al. 1989), effects which might seem straightforwardly validate motoric stopping accounts and falsify context-monitoring accounts. However, we note that stimulation of the medial temporal lobe (MTL) leads to difficulties in memory retrieval (e.g., Halgren et al 1985), but the MTL is not thus assumed to inhibit memory retrieval processes. Applying analogous logic to the rVLPFC, we suggest (as did Luders et al, 1988) that these rVLPFC macrostimulation effects may reflect some more general *loss* of function, rather than the *in*duction of a normal "stopping" function implemented by rVLPFC.

A variety of more equivocal evidence has been represented as more consistent with motoric stopping accounts of rVLPFC function than with context-monitoring accounts (e.g., Tabibnia et al, 2011). For example, it has been argued that the rapid inhibitory effects of rVLPFC on primary motor cortex, as determined through paired-pulse TMS techniques, supports

a motoric stopping role for rVLPFC (Buch et al., 2010). However, such rapid inhibitory effects are also observed with paired-pulse TMS to regions outside of the rVLPFC – suggesting that these effects may be a relatively more general feature of cortex, and not related to some functional specialization of the rVLPFC (Reis et al., 2009). Similarly, increases in the power of beta-band frequency oscillations in the ongoing electroencephalogram (EEG) have been linked to both the rVLPFC and to successful stopping (Swann et al., 2009), but increased rVLPFC beta-band coherence is also observed during simple preparation of motor responses (Fischer, Langner, Diers, Brocke & Birbaumer, 2010).

Context-monitoring accounts have also been criticized for relying on the P3 complex to demonstrate the similarities of tasks involving the detection of behaviorally-relevant but infrequent stimuli across various populations and pharmacological manipulations. Specifically, it is argued that the P3 tends to peak too late for it (the peak) to play a causal role in response inhibition. However, if a component *peaks* relatively late that does not mean that the component, or the process it reflects, is not also operative much earlier. Indeed, the Stop P3 has often been taken to reflect motoric stopping processes *per se* despite this prolonged peak latency.

Concerns have also been raised regarding the ability of the context-monitoring hypothesis to explain inhibitory control deficits in ADHD. In particular, context-monitoring has been criticized for predicting that children with ADHD should show difficulty not only in stopping responses but in committing alternative responses as well, whereas only the former is observed in Marriott et al. (2005). In actuality, this study demonstrated not only an elevated SSRT among children with ADHD but also more variable and slower execution of alternative responses – a finding that is conceptually consistent with their elevated reaction times to behaviorally-relevant but infrequent stimuli (Schachar & Logan, 1993), if both phenomena reflect deficits in context-

monitoring. Notably, motoric stopping accounts have more recently been the subject of these same criticisms with respect to ADHD given the inefficacy of therapies targeted at these mechanisms (Rappoport et al., 2001; Alderson et al., 2008).

Other criticisms of context-monitoring concern the fact that stop-signal performance is not particularly resource-demanding, insofar as it has failed to show the "psychological refractory period" (PRP) effects (e.g., Logan & Burkell, 1986). It is difficult to draw conclusions from this null effect (although see Pashler, 1994 for one attempt). More informative are both preceding and subsequent reports which demonstrate that response inhibition is not only subject to refractoriness (Horstmann, 2003; Welford, 1957), but also induces a PRP effect (Horstmann, 2003). These results can be straightforwardly explained by the hypothesis that context-monitoring is both cognitively-controlled and demanding of central executive resources.

More significantly, the context-monitoring hypothesis has been criticized as being overly broad. There is however some support for a role of the rVLPFC specifically in context-monitoring, as opposed to that of a more general role in executive control. First, consider a paradigm in which subjects must focus on a central stimulus while inhibiting the interference arising from an array of incongruent stimuli presented to the left and right of the central stimulus (a flanker task). At the onset of these stimuli, gamma-band oscillations increase in power, with dipole sources localized to the rVLPFC (Fan et al., 2007). This change in power is abolished by orienting subjects to the *vertical* position of the upcoming stimuli – regardless of whether the flanking stimuli were congruent or incongruent (the latter clearly demanding on executive control) to the central stimulus (Fan et al., 2007). There are also reasons to believe that the rVLPFC is not critical for the monitoring of rules in working memory (e.g. Rushworth et al., 1997, who incidentally also concluded that “the inclusion of a response suppression element is

not necessary for a task to be impaired" by VLPFC damage, consistent with our claims). These studies would seem to put limits on the generality of context-monitoring and rVLPFC-associated processes; indicating that they are not simply recruited in every case where executive control is required.

Genuine Outstanding Challenges

On the other hand, a handful of results are genuinely difficult to reconcile with the current framework, although they also tend to be unnatural bedfellows with motoric stopping accounts. In this light, these studies suggest that the complete account of rVLPFC function is yet undiscovered, particularly with respect to the functional hemispheric asymmetry of the rVLPFC, its functional relationship to goal maintenance, and its mapping between humans and higher primates.

In terms of functional asymmetry, the rVLPFC may show a somewhat greater sensitivity in the processing of nonverbal relative to verbal stimuli, even when these stimuli are task-irrelevant and not contextually infrequent. For example, in a flanker paradigm with equiprobable neutral and incongruent flankers (Morimoto et al., 2008), the rVLPFC was more activated in response to incongruent color than incongruent word flankers, whereas a homologous region of the left VLPFC showed the opposite pattern. This finding was interpreted to reflect right-lateralized inhibition of irrelevant color information, and left-lateralized inhibition of irrelevant word information (contrary to previous reports that the rVLPFC is not modality specific; Hazeltine, Bunge, Scanlon, & Gabrieli, 2003). However, this conclusion is cast into doubt by a correlation with performance: This effect was pronounced among those with lower accuracy, contrary to suggestions that those with superior response inhibition tend to more strongly activate the rVLPFC (Aron & Poldrack, 2006).

Such results are also difficult to reconcile with a context-monitoring account of the rVLPFC. A plausible alternative account is that greater activation of right frontal regions reflects increased processing of the color flankers, and greater activation of left frontal regions reflects increased processing of the word flankers, consistent with verbal/nonverbal accounts of functional hemispheric asymmetry. This account is further compatible with theories that VLPFC regions work in concert with more dorsal regions of the PFC in implementing the top-down selection of task-relevant information, such that greater activity in the hemisphere with more sensitivity to the flanker modality is actually an index of interference. By this expanded account, task-relevance and contextual-frequency are sufficient, but not necessary characteristics for determining the recruitment of the rVLPFC.

An additional example of the potential role of the rVLPFC in other functions thought to rely on the prefrontal cortex at large, such as goal maintenance, comes from a cued Simon task with equiprobable congruent and incongruent stimuli (Forstmann et al., 2008); in this experiment, a minority of trials were invalidly cued as to the congruency or incongruency of the location of an upcoming stimulus and the location of the required response. Given that congruent and incongruent trials were matched for probability, the context-monitoring hypothesis would predict no greater involvement of rVLPFC as a function of congruency; accordingly, no differential rVLPFC activity was observed in the invalid or the valid conditions. This result runs contrary to motoric stopping accounts of the Simon task, in which such stopping is important on incongruent trials, and perhaps particularly important in the invalid condition.

Nonetheless, greater involvement of the rVLPFC was observed on invalidly-cued incongruent relative to neutral trials among those with greater inhibitory skill, a finding that would not have been predicted by the context-monitoring hypothesis. This measure of inhibitory

skill was a composite "difference of differences" score, raising the possibility that the correlation actually reflects an advantage in goal maintenance for those with putatively better inhibition (consistent with evidence from the Stroop task; Kane & Engle, 2003). Again, these results show that top-down goal maintenance may be another function of the rVLPFC, and another important ability governing response inhibition, in addition to the important role played by the detection of contextually-infrequent and behaviorally-relevant stimuli.

Finally, our review has focused primarily on the human literature, in which the neuronal processes recruited by context-monitoring tasks (and response inhibition tasks) are typically right-lateralized. Given that non-human primates do not show such consistent lateralization of function, it is possible that both monitoring and motoric stopping accounts are not well-suited to explain non-human primate data. For example, the majority of cells in the monkey rVLPFC differentiate between Go and NoGo stimuli only when they differ in color (Sakagami et al., 2001), and not motion (64 out of 73 cells; contrary to reports from the human literature that rVLPFC responses are not modality-specific; Hazeltine, Bunge, Scanlon, & Gabrieli, 2003).

A majority of these color-sensitive cells responded preferentially to *go* stimuli (41 out of 64 cells), and, most critically, these rVLPFC cells never showed consistent firing patterns in response-locked analyses (contrary to theories that the rVLPFC is involved in the execution of response inhibition, as opposed to context-monitoring, which should be more closely related to the stimulus onsets in such tasks). Clearly, these results are incompatible with perspectives emphasizing the importance of motoric stopping demands.

But two aspects of these results are also difficult to interpret from a monitoring framework – especially since Go and NoGo stimuli were equiprobable in this design. First, the majority of color-sensitive cells in the rVLPFC differentiated between Go and NoGo only by

suppressing their activity to the color NoGo stimulus, responding similarly to color Go stimuli and both Go and NoGo when defined by motion. Conversely, a minority of rVLPFC cells increased their activity selectively to the color NoGo stimulus, not showing much response to any other stimulus type whether defined by color or motion. Thus, in this particular study, NoGo stimuli do appear to have a special status in the rVLPFC. While NoGo trials were arguably more difficult (instead of requiring *no* response, they actually required a *more delayed* response than Go stimuli), the larger issue is that our context-monitoring hypothesis is primarily based on human data, and may not straightforwardly apply to animal models, even those based on higher primates. Clearly, integration of human and animal work is an important area for future research and theorizing – not just in the domains of monitoring and response inhibition, but in executive function as a whole.

Novel Predictions

While not offering a complete account of rVLPFC function, context-monitoring hypothesis nonetheless makes specific and falsifiable predictions about how factors like task difficulty, stimulus deviance (whether in terms of stimulus probability or perceptual characteristics) and task-relevance should influence activity in the rVLPFC, independent of motoric stopping demands. To the extent that such predictions can be falsified, they may point the way towards a more comprehensive account of the rVLPFC.

Foremost among these predictions is that rVLPFC damage should alter frontal P3 components, and produce "attentional neglect" for infrequent and behaviorally-relevant stimuli, in a way that should mediate any concomitant deficits observed from these patients in tests of response inhibition. Second, the context-monitoring hypothesis predicts that relatively domain-general mechanisms should support performance on the Stop task and related response inhibition

tasks. In particular, a neural network implementation of context-monitoring may naturally emerge when an otherwise general architecture for executive control is trained on the Stop task (e.g., the PBWM framework; O'Reilly & Frank, 2006). Such architectures should be capable of giving rise to canonical phenomena from the Stop task if the context-monitoring hypothesis is correct.

Third, the context-monitoring hypothesis predicts that the variations in SSRT and response interference observed in "selective stop" paradigms (where only one of multiple responses must be stopped) are more likely to reflect differences in the recruitment of these rather domain-general mechanisms, rather than reflecting differences in the neural circuitry that is specific to motoric stopping demands. Existing accounts of such paradigms posit that selective stop paradigms recruit the use of a slower but more specific pathway for motoric stopping – the subcortical "indirect" pathway of the striatum as opposed to the hyperdirect pathway of the subthalamic nucleus – and that this indirect pathway is preferentially used when subjects have foreknowledge about which of the multiple responses may need to be stopped. By contrast, the context-monitoring hypothesis would predict that this foreknowledge does not have its effects by changing the precise subcortical route used in support of motoric stopping, which is instead hypothesized to be automatic and uniformly global. Instead, foreknowledge may have its effects (for example) by reducing the contextual-infrequency or behavioral-relevance of the stop signals, thereby yielding reduced rVLPFC activation, and concomitantly a less abrupt onset of automatic motoric stopping mechanisms.

Conclusions

Response inhibition paradigms typically require not only motoric stopping but also the ability to monitor for and detect behaviorally-relevant and contextually infrequent stimuli. This

more general attentional demand may underlie the recruitment of the rVLPFC in a variety of tasks, including those that do not require motoric stopping. This *context-monitoring hypothesis* illuminates the tight relationship between inhibitory deficits and abnormalities in oddball paradigms in populations with brain damage, with psychopathological disorders, and those undergoing pharmacological treatment. It also offers a coherent and comprehensive account of rVLPFC function in the context of response inhibition tasks. The exact computations performed by the rVLPFC in the service of monitoring remain to be specified, and some findings cannot be explained by any single existing theoretical account. Nonetheless, the context-monitoring hypothesis offers an important step towards understanding the processes that constitute the ability to inhibit, towards characterizing individual differences in that ability, and towards understanding the role of the rVLPFC in a variety of domains.

Footnotes

[1]While the use of a global braking mechanism appears to be automatic, the amount of braking can be modulated by the degree to which stop signals are unexpected: stopping interference is reduced, but not abolished, when subjects have informative foreknowledge about the occurrence of a stop signal (Aron & Verbruggen, 2009).

[2]One recent study of a paradigm without inhibitory demands reported a null interaction of rVLPFC activity with target frequency (Hampshire, Thompson, Duncan & Owen, 2009), but the frequency manipulation was small (40% vs 50% targets).

[3]Bekker et al report a null effect in the case of nicotine; on the other hand, Bekker et al used an unusually high proportion of stop trials (40%), which may attenuate the influence of monitoring or inhibition.

CHAPTER 2: COGNITIVE CONTROL REFLECTS CONTEXT MONITORING, NOT MOTORIC STOPPING, IN RESPONSE INHIBITION

Adapted from Chatham, C.H., Claus, E.D., Kim, A., Curran, T., Banich, M.T., Munakata Y (in press) Cognitive Control Reflects Context-Monitoring, Not Motoric Stopping, In Response Inhibition. *PLoS One.*

Abstract:

The inhibition of unwanted behaviors is considered an effortful and controlled ability. However, inhibition also requires the detection of contexts indicating that old behaviors may be inappropriate – in other words, inhibition requires the ability to monitor context in the service of goals, which we refer to as context-monitoring. Using behavioral, neuroimaging, electrophysiological and computational approaches, we tested whether motoric stopping per se is the cognitively-controlled process supporting response inhibition, or whether context-monitoring may fill this role. Our results demonstrate that inhibition does not require control mechanisms beyond those involved in context-monitoring, and that such control mechanisms are the same regardless of stopping demands. These results challenge dominant accounts of inhibitory control, which posit that motoric stopping is the cognitively-controlled process of response inhibition, and clarify emerging debates on the frontal substrates of response inhibition by replacing the role of a controlled mechanism for motoric stopping with context-monitoring.

Introduction

The inhibition of responses is critical for enabling controlled behavior: bad habits, unfamiliar situations, and dangerous environments often require that default behaviors be stopped and more context-appropriate actions performed (Aron, 2007). Response inhibition has been localized to brain regions implicated in behavioral control, such as the right ventrolateral prefrontal cortex (rVLPFC), and response inhibition has been statistically equated with the behavioral and genetic variance common across multiple tests of cognitive and behavioral control (Aron, Robbins & Poldrack, 2004; Friedman et al., 2008). Moreover, the inhibition of responses has been exempted from skepticisms about the existence of other forms of inhibition (MacLeod, Dodd, Sheard, Wilson & Bibi, 2003), supporting theorizing that similar mechanisms enable the inhibition of thoughts and emotions. Thus, modern theorizing is largely consistent with a hypothesis proposed 130 years ago: that "the centers of inhibition being thus the essential factor of attention, constitute the organic basis of all the higher intellectual faculties" (Ferrier, 1876).

However, response inhibition not only requires that inappropriate behavior be stopped – it also requires the detection of behaviorally-relevant signals. For example, one goal may be to cross a street; this requires actually crossing the street, and stopping these motor actions if oncoming traffic is approaching, but to do so the environment must be monitored so that motoric stopping can be performed as appropriate. In other words, the environmental context must be monitored to support behavior that may be contingent on that context. Both motoric stopping and context-monitoring are also intermingled in the most precise laboratory assessment of response inhibition, in which subjects must cancel a prepotent or planned response after the

presentation of a signal to stop (Chikazoe et al., 2009; Hampshire et al., 2010; Sharp et al., 2010; Cai & Leung, 2011; Dodds et al., 2011). Estimates of the time that subjects require to stop an action, or "Stop Signal Reaction Time" (SSRT; Logan & Cowan, 1984) are thought to include the time spent detecting and encoding the signal to stop, but do not explicitly distinguish between this process and the motoric process involved in actually stopping the action. Thus, the apparent centrality of a controlled stopping process in this task could in principle reflect the centrality of controlled context-monitoring processes.

Here, we experimentally determine whether context-monitoring or motoric stopping constitutes the cognitively-controlled process recruited during response inhibition by examining tasks with identical context-monitoring demands, one of which requires stopping and one of which does not. In both tasks, 75% of trials ("No Signal" trials) require a 2-alternative forced choice (2AFC; Fig. 2A); in the remaining 25% of trials ("Signal" trials), the 2AFC is followed by a behaviorally-relevant stimulus (the "signal") after a variable delay (Fig. 2B). In the Stop Task, Signal trials require the stopping of motor responses on that trial. In contrast, in the "Double Go Task," Signal trials require subjects to repeat their response for that trial as quickly as possible (see methods and Supporting Information). Thus, both tasks require monitoring for the context that signals what actions should be executed, but only the Stop Task explicitly requires actions to be stopped.

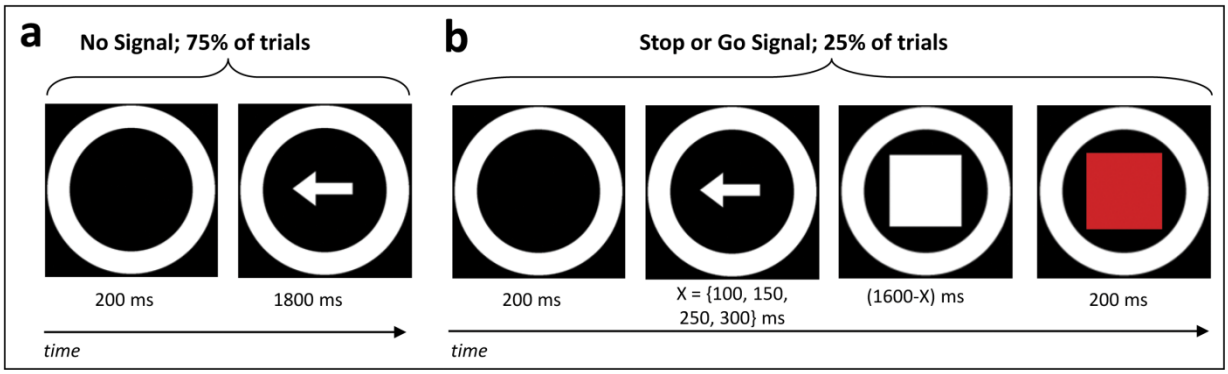


Figure 2. Task Design. Identical stimuli and trial structure were used across tasks in three separate experiments. In both the Stop and the Double Go tasks, most trials are "No Signal" trials where only a 2AFC decision is required (A). However, the tasks differ on "Signal" trials (B) where an additional stimulus, a white box, is presented with a variable inter-stimulus interval following the onset of the 2AFC stimulus. On Double Go_{Signal} trials, this additional stimulus indicates that the appropriate 2AFC button press be repeated. On $Stop_{Signal}$ trials, this stimulus indicates that the 2AFC button press must be stopped. Thus, although only the Stop Task requires motoric stopping, both tasks share demands on context-monitoring.

The cognitive control required for response inhibition is thought to rely on the prefrontal cortex, to be most crucial at the moment when motoric stopping is required, to be associated with substantial mental effort, to be recruited in a goal-directed fashion, and to support consistent individual differences. We therefore assessed each of these characteristics of context-monitoring and stopping, to determine which of these components may reflect the cognitively-controlled process recruited during response inhibition.

We used functional magnetic resonance imaging (fMRI) to assess the recruitment of the prefrontal cortex in our tasks. Numerous previous fMRI studies have demonstrated transient activation within the right ventrolateral prefrontal cortex (rVLPFC) and the adjoining anterior insula during trials that require motoric stopping (e.g., Chikazoe et al., 2009; Hampshire et al., 2010; Dodds et al., 2011; Sharp et al., 2011). Collectively, this and related evidence has been interpreted to indicate that the rVLPFC is a dedicated substrate for inhibition, and that this

function may also be deployed proactively to support behaviors like "responding with restraint" (Aron, 2010; Jahfari et al., 2010). Alternatively, it is possible that these hemodynamic patterns reflect the context monitoring demands of the Stop task, which could also be deployed proactively as well as transiently at the moment a goal-relevant feature of the environment (e.g., a Stop Signal) is encountered. Recent work has begun to examine this alternative possibility using fMRI but has unfortunately yielded inconsistent results: either more (Hampshire et al., 2010; Dodds et al., 2010), less (Chikazoe et al., 2010; Cai & Leung, 2011) or roughly equivalent (Sharp et al., 2010) rVLPFC activity is observed during the Stop task, in either overlapping (Sharp et al., 2010; Dodds et al., 2011; Hampshire et al., 2010) or distinct (Chikazoe, et al., 2009; Cai & Leung, 2011) subregions of the rVLPFC. In addition, none of these studies have examined whether the sustained component to the rVLPFC hemodynamic response could reflect a tonic and proactive process of context-monitoring (in which case sustained activity should also be observed in a context-monitoring task) rather than a process of responding with restraint. Finally, all of these studies have examined only the univariate patterns in hemodynamics, and have not assessed whether rVLPFC demonstrates multivariate commonality across tasks involving context-monitoring (as would be predicted by context-monitoring accounts), or whether any such commonality is relatively decreased on trials requiring motoric stopping (as would be predicted by stopping accounts). Below, we measured each of these aspects of the recruitment of the rVLPFC during the Stop and Double Go tasks to test these differing predictions of the context-monitoring and motoric stopping accounts.

We also assessed whether the event-related potentials commonly associated with response inhibition tasks, and often presumed to reflect motoric stopping processes, might instead reflect context monitoring processes. The most characteristic ERP from response

inhibition tasks is the "Stop P3" or "No/Go P3," a frontocentral positivity elicited following the onset of stimuli which demand motoric stopping (e.g., Smith, Johnstone & Barry, 2008). We tested whether this "Stop P3" would be more strongly expressed during the Stop task than the Double Go task (as motoric stopping accounts would predict), and whether the correlation of all ERPs across these tasks would be reduced following the onset of the Signals (as would also be predicted by motoric stopping accounts). In contrast, accounts positing the centrality of context-monitoring to the Stop task would predict roughly equivalent frontocentral ERPs across these tasks, despite their differing demands on motoric stopping.

Finally, we assessed the task-evoked pupillometric response (TEPR), a well-validated measure of mental effort (Kahneman & Beatty, 1966; Beatty & Lucero-Wagner, 2000), to determine whether the more effortful component to the Stop task reflects motoric stopping (in which case pupil diameter should be increased on $Stop_{Signal}$ trials) or whether it might reflect the act of monitoring context for goal-relevant signals (in which case, pupil diameter may show a more complex pattern, such as a modulation of pupil diameter by the relevance of a monitored signal to the planned response). Previous work examining pupil diameter in the Stop task has utilized it mainly as a control measure of arousal in TMS studies (Chambers et al., 2006; Verbruggen et al., 2010).

To foreshadow our results, our results uniformly suggest that, during response inhibition, cognitive control is primarily engaged for the purpose of monitoring the environmental context in the service of goals, rather than for motoric stopping *per se*.

Materials and Methods

Participants.

For experiment 1, 86 subjects (mean age 19.11 years; SD = 1.17 years; 32 males) were recruited using the University of Colorado undergraduate research pool and successfully completed the Double Go and Stop Tasks. 2 subjects failed eyetracker calibration and were excluded from pupillometric analyses. For experiment 2, 45 subjects (mean age 19.86 years; SD = 2.21; 23 males) were recruited using the University of Colorado undergraduate research pool and successfully completed the Double Go and Stop Tasks. 7 of these subjects were excluded from ERP analysis for artifacts caused by excessive blinking (>60% of trials). For experiment 3, 19 subjects (mean age 23.3: SD = 4.4; 10 males) were recruited from the local community and successfully completed the Double Go and Stop Tasks. One subject was excluded from fMRI analyses due to motion artifact.

Behavioral Task.

All subjects in all experiments completed the Double Go Task prior to completing the Stop Task. This fixed task order was adopted for reasons described in Supporting Materials – in particular, the use of a fixed task order is ideal for the investigation of individual differences (e.g., Friedman et al., 2008), which was a central goal of the study reported here. Nonetheless, appropriate precautions were taken to prevent the contamination of experimental effects with cognitive phenomena that might arise from the fixed task order (e.g., the use of within-task baselines are used for all pupillometry, ERPs, and fMRI analyses, so as to control for the relatively general effects of phenomena like fatigue).

In all respects the Double Go and Stop Task were identical within any given experiment (e.g., the precise interstimulus and intersignal intervals, the presence of "null" trials, etc), with the following exception: subjects are naturally aware of when they fail to successfully stop a response, but seem unaware of their relative speed on trials with the infrequent stimulus. To avoid any possible mismatch across the two tasks owing to this difference, we provided explicit feedback on all signal trials. Specifically, in the Double Go Task, the signal turned red if subjects were slower than their average running RT (experiments 2 & 3); in experiment 1 this was presented as sham feedback. (Double Go task trials with categorically incorrect responses – such as a failure to respond twice on Signal trials, or anything but a single correct response on No Signal trials – were extremely rare and excluded from all analysis). Similarly, in the Stop Task, the signal turned red if subjects failed to successfully stop their response on that trial (in all experiments). Additional cross-experiment differences in our tasks suggest the generality of our results across minor variations in experimental procedure (see Supporting Figure S1 & Supporting Table S1).

Statistical Analysis of fMRI

Data were acquired with a 3T GE Signal whole-body MRI scanner at the University of Colorado Health Sciences Center, using T2-weighted echo-planar imaging (EPI) (TR= 2000 ms, TE= 32 ms, flip angle= 70º). Additional acquisition details are available in Supporting Methods. Image pre-processing and analyses were conducted with FSL (FMRIB's Software Library). The first six volumes of each run were discarded to allow the MR signal to reach steady state, the remaining images in each participant's time series were motion corrected using MCFLIRT, and non-brain voxels were removed using a brain extraction algorithm (BET). The data series was

spatially smoothed with a 3D Gaussian kernel (FWHM = 5 mm), intensity normalized for all volumes, and high-pass filtered (s=50 sec).

After statistical analysis of each time series (details of the regression model are available in Supporting Methods), statistical maps were normalized into the MNI-152 stereotaxic space using FLIRT (FMRIB's Linear Image Registration Tool). Parameter estimates (PE) were transformed into a common stereotaxic space using the above-mentioned three-step registration prior to the group analyses with FLAME (FMRIB's Local Analysis of Mixed Effects). Z-statistic images were thresholded using clusters with $z > 2.58$ as well as a whole-brain corrected cluster significance threshold of $p < .05$ using the theory of Gaussian Random Fields.
ROIs for Brodmann areas were anatomically defined using the Talairach labeled atlas, and mean percent signal change was extracted using FSL's featquery tool. The subthalamic nucleus was anatomically defined using a $10mm^3$ region centered on the MNI coordinates previously used in the Stop Task to interrogate BOLD in the STN (10, -15, -5) (Aron & Poldrack, 2006). The TPJ was anatomically defined using a $30mm^3$ region centered on the MNI coordinates (-54, -52, 30) previously observed in a target detection task (Corbetta, Kincade, Ollinger, McAvoy & Shulman, 2000).

Pattern classification analyses were conducted on the beta-weights resulting from the above fMRI analysis pipeline, with four minor exceptions. First, the BOLD data were not spatially smoothed; second, the PEs were not statistically thresholded; third; the PEs were z-transformed across all voxels within a given ROI for each subject, to ensure that the classifiers were forced to operate on the basis of distributed patterns of activation instead of overall magnitudes. Finally, voxels with z-values falling outside of +/- 4.5 were windsorized. Classifiers were implemented as neural networks in Emergent (Aisa, Mingus & O'Reilly, 2008);

separate networks were then trained, using Hebbian and Contrastive Hebbian learning, for each ROI (and therefore differed in terms of the number of input units), and for identifying which individuals generated the data vs. what trial type the data was estimated from (and therefore differed in terms of the number of output units) but all other aspects of the network architecture were the same. See Supporting Methods for full details on classifier implementation.

Statistical Analysis of ERPs.

During the Double Go and Stop Tasks scalp voltages were recorded with a 128-channel geodesic sensor net (Tucker, 1993). Amplified analog voltages (0.1- to 100.0-Hz bandpass) were digitized at 250 Hz. Individual sensors were adjusted until impedances were less than 50 k. The EEG was digitally low-pass filtered at 40 Hz. Trials were discarded from analyses if they contained incorrect responses, eye movements (eye channel amplitudes over 70 V), or more than 20% of channels were bad (average amplitude over 100 V or transit amplitude over 50 V). Individual bad channels were replaced on a trial-by-trial basis with a spherical spline algorithm. EEG was measured with respect to a vertex reference (Cz), but an average-reference transformation was used to minimize the effects of reference-site activity and accurately estimate the scalp topography of the measured electrical fields. The average reference was corrected for the polar average reference effect (Junghoefer, Elber, Tucker & Braun, 1999). ERPs were obtained by stimulus-locked averaging of the EEG recorded in each condition. ERPs were baseline-corrected with respect to a 200-ms prestimulus recording interval. These baselines were calculated separately for each task, thereby controlling for nonspecific effects like fatigue.

Where montages are used, the occipital montage was centered on Oz (including Oz, O1, O2, and the contiguous set of electrodes 76, 70, 74 and 82) and the frontal montage was centered on Fz

(including Fz and the contiguous set of electrodes 4, 5, 10, 12, 16 18 and 19). For scalp-wide voltage correlations we calculated Pearson's R across tasks at every time point as the variance shared between the subjects x electrode matrix across tasks. Thus, this correlation reflects changes in voltage that covary across tasks in the same subjects at the same electrode sites. For montage-based voltage correlations we calculated Pearson correlations separately for the frontal and occipital montages both before and after signal onset.

Statistical Analysis of Pupillometry.

Pupil diameter was recorded continuously during the Double Go and Stop Tasks via a Tobii X50 infrared eyetracker calibrated to each subject. Sampling at 50 Hz was synchronized to fixation onset, and pupil diameter was calculated as the average diameter of successfully-tracked eyes for each sample. Baseline measurements of pupil diameter were calculated as the average diameter during the 200ms preceding the onset of each signal (or the corresponding time period for no signal trials); this value was subtracted from the averaged samples recorded following the onset of the signal (or the average signal onset for no signal trials). Baseline periods were calculated independently for the Stop and Double Go tasks, providing a within-task baseline to control for nonspecific cognitive effects like fatigue. These normalized, averaged pupil diameter samples were then smoothed using a box-car filter with width of 60ms.

Statistical Analysis of Behavior – Double Go Task.

In the Double Go Task, all RTs falling below 150ms or above 750ms were excluded from analysis, as well as those on No Signal trials falling outside of 3.5 standard deviations of the iteratively-calculated mean for each subject. RTs were only analyzed on correct trials (i.e., trials

in which two responses of the correct type were provided on Signal trials, and where one and only one response of the correct type was provided on No Signal trials).

Individual differences were extracted from the Double Go task using a mixture model-based adaptation of the classic race model of the Stop task (see also Supporting Methods). Specifically, to classify individual trials as slowed or unslowed, we first decomposed the distribution of equipercentile residuals into two underlying distributions: a Gaussian distribution with a mean of zero (corresponding to unslowed first RTs), and a Gamma distribution (corresponding to the slowed first RTs). The two free parameters to the Gamma and the one free parameter to the Gaussian were fit in a fixed-effects analysis using maximum likelihood estimation via with the Nelder-Meade simplex algorithm (Nelder & Mead, 1965; Ratcliff & Starns, 2009). The maximum likelihood fit is illustrated as overlaid lines on the residual histogram (Fig. S3A), which was relatively stable across multiple optimizations with different starting parameters and yielded a better overall fit (see Supporting Table S1) than a single Gaussian in terms of the Bayesian Information Criterion (BIC), calculated as:

$$BIC = -2 \cdot \sum_{n=1}^{N} \ln\left(\sum_{d=1}^{D} \Phi_d L_d(RT_n) \right) + D_p \cdot \ln(N) \qquad (1)$$

Where N is the total number of observations, D is the total number of distributions fit, D_p is the total number of free parameters used in fitting those distributions, Φ_d is the weight of the d^{th} distribution, and $L_d(RT_n)$ is the likelihood of the n^{th} RT given the best fit parameters for the d^{th} distribution (μ and σ for Gaussian and k and Θ for Gamma).

We next categorized individual trials as slowed or unslowed using the likelihood of observing each RT under either of the two fitted distributions. RTs were categorized as slowed if there was even weak evidence in favor of the RT belonging to that distribution (as quantified

by a difference in BIC of ≥ 2.35); otherwise RTs were categorized as unslowed. Other standards of evidence lead to similar results as those presented here, but do not as cleanly separate the slowed and unslowed trials (c.f. Fig S3B).

To calculate TOSD, we subtracted the signal delay from the n^{th} percentile of no signal trial RTs, where *n* corresponds to the proportion of RTs classified as unslowed at that signal delay. This approach is conceptually identical to that used to calculate SSRT in the race model, in which the signal delay is subtracted from the n^{th} percentile of No Signal RTs, where *n* corresponds to the proportion of unsuccessful stop trials at that signal delay. TOSD was calculated for each subject as the median of these estimates across all signal delays. This estimate was unreasonably high for subjects for whom no RTs had been classified as slowed (n=34 out of 150), so in those cases we used the minimum estimate of TOSD across all signal delays.

We then calculated the duration of slowing as the average difference between RTs classified as slowed and RTs of corresponding percent rank in the no signal RT distribution; subjects for whom no RTs had been classified as slowed were excluded from all analyses involving duration of slowing. The resulting estimates of TOSD and duration of slowing can be found in Table S2.

Statistical Analysis of Behavior – Stop Task.

In keeping with the recommendations based on Monte Carlo simulations (Band et al., 2003), we estimated SSRT as the n^{th} percentile of the No Signal RT distribution minus the signal delay, where *n* is the proportion of errors observed at each signal delay. This estimate was averaged across the signal delays yielding 15% to 85% accuracy for each subject to generate the

recommended dependent measure for Stop Tasks with fixed interstimulus intervals ($SSRT_{AV}$). Data from the Stop Task confirmed assumptions of the race model: RTs were faster on Signal trials than on No Signal trials ($t(145)=11.31$, $p<.0005$) and accuracy was a monotonically decreasing function of interstimulus interval (100 vs 150: $t(145)=13.52$, $p<.0005$; 150 vs 250: $t(145)=17.20$, $p<.0005$; 250 vs 300: $t(145)=7.14$, $p<.0005$).

Results

Univariate fMRI Results

First, we found that context-monitoring rather than stopping explained the transient prefrontal contribution to response inhibition. Accounts which posit that motoric stopping is the controlled process during response inhibition tasks predict rVLPFC activation only in the Stop task, but event-related fMRI revealed that the Stop and Double Go tasks activated completely overlapping regions of prefrontal cortex (Fig.3A), consistent with the tasks' shared context-monitoring demands. Specific regions of interest (ROIs) in the rVLPFC and interconnected subthalamic nucleus (STN) that have been proposed to be specific to the motoric stopping demands were uniformly more strongly recruited on Signal trials in the Double Go Task (Fig 3B&C; STN: $t(17)=5.49$, $p<.0001$; BA44: $t(17)=5.08$, $p<.0001$; BA45: $t(17)=2.83$, $p=.012$; BA47: $t(17)=2.5$, $p=.023$), challenging any characterization of these areas as specialized for motoric stopping. A significantly different pattern was observed in areas thought to have a more general attentional role (e.g., the temporo-parietal junction; TPJ (Chikazoe et al., 2007; Konishi et al., 1998); $F(1,17)=31.57$, $p<.0001$), such that both tasks recruited this area equivalently. This equal recruitment of the TPJ across tasks indicates that decreased recruitment of the rVLPFC in

the Stop task cannot be explained by globally-decreased activation during that task (e.g., as might result from fatigue; see also discussion in Supporting Information). Moreover, the increased recruitment of rVLPFC during the Double Go task is consistent with several recent findings, which also demonstrate that tasks involving both context-monitoring and response commission are associated with increased rVLPFC activity relative to tasks involving both context-monitoring and a demand to stop motor actions (Hampshire et al., 2010; Sharp et al 2010; Dodds et al., 2011; but see Cai & Leung, 2011 and discussion, below).

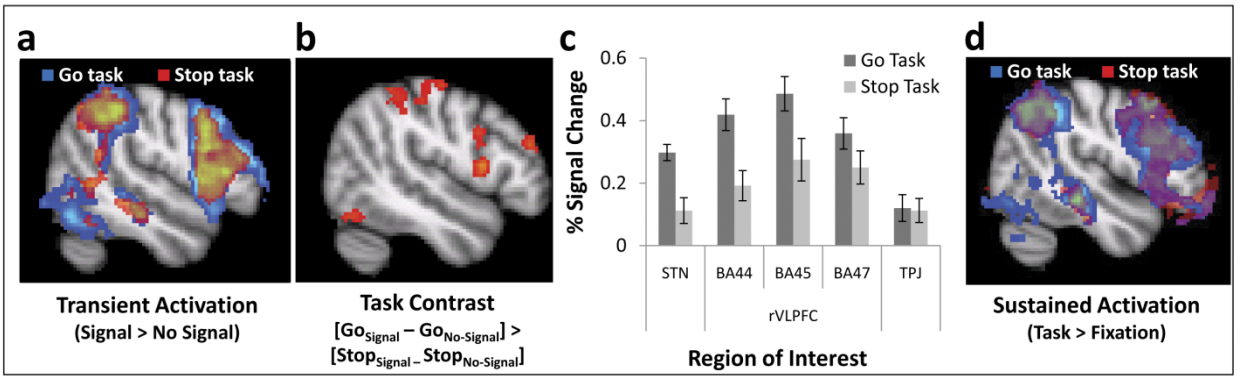


Figure 3. Hybrid fMRI analyses revealed overlapping neural activity in the Stop and (Double) Go Tasks (A), with significantly more rVLPFC activity in the Go Task (B). ROI analyses for the contrast of Signal vs. No-Signal trials (C) revealed increased activity in the Go Task throughout a putatively stopping-specific network; this pattern did not generalize to regions with more general attentional functions (e.g., TPJ). Sustained rVLPFC activity was also observed across all trials within each task (D).

Our hybrid fMRI design also allowed us to assess the extent to which neural regions were recruited in a sustained fashion across all trials within the Stop and Double Go tasks. Such sustained activity is potentially a hallmark of proactive context-monitoring processes. Indeed, this analysis revealed sustained hemodynamics in the rVLPFC during both tasks at the timescale of seconds-to-minutes (Fig. 3D), consistent with their shared sustained context-monitoring demands. In contrast, accounts positing that motoric stopping is the cognitively-controlled process during response inhibition predict no sustained rVLPFC activity in the Double Go task,

since only response commission is required by that task, and "responding with restraint" is unnecessary.

Multivariate Pattern Analysis

We next leveraged multi-voxel pattern analysis to determine whether the same information was encoded by rVLPFC regardless of whether motoric stopping is required by a given task. First, we trained classifiers to identify hemodynamic patterns that reliably predicted subject-specific patterns of rVLPFC activation in the Double Go task over 10 independent runs of the classifier (see methods in Supporting Information). Classifiers readily generalized their training on the Double Go task to distinguish individuals in the Stop task, indicating that the rVLPFC is recruited in an individual-specific but consistent way across tasks. These patterns were significantly more consistent across tasks on Signal trials in the rVLPFC – precisely when and where context-monitoring processes are most crucial, but also when motoric stopping demands differ most across these tasks (Fig.4A; BA44: $t(9)=13.5$, $p<.0001$; BA45: $t(9)=11.39$, $p<.0001$; BA47: $t(9)=12.35$, $p<.001$). Critically, the increased cross-task similarity of Signal trials relative to No Signal trials was not observed in an area known to encode responses – primary motor cortex – and this pattern was significantly different from that observed in rVLPFC ($F(1,9)=85.12$, $p<.0001$). Although these results do not conclusively demonstrate that the cognitive processes engaged by both tasks are the same, they do demonstrate that the multivariate representations in the rVLPFC fail to show differential sensitivity to the explicit stopping demands imposed by Signal trials within the Stop task (in contrast to the multivariate patterns within primary motor cortex). While this pattern contradicts the idea that

representations in rVLPFC are specialized for the motoric stopping that is required on Signal trials in the Stop task (but not in the Double Go task), it is wholly consistent with the idea that similar context-monitoring processes are elicited by Signal trials within both tasks.

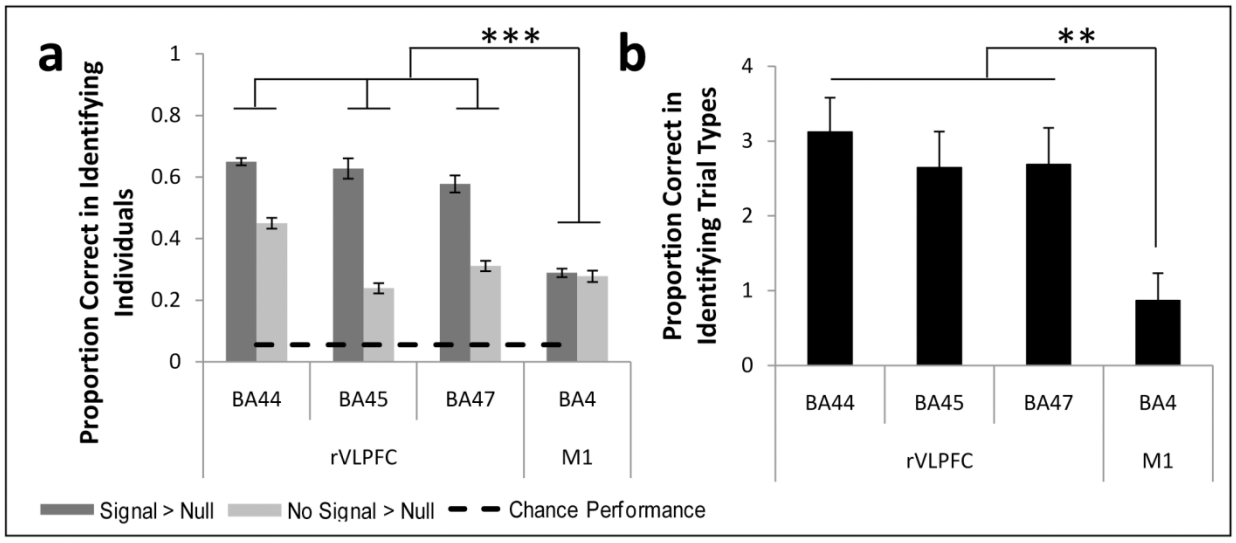


Figure 4. (A) rVLPFC was recruited in subject-specific but consistent ways regardless of stopping demands: individual differences in (Double) Go task hemodynamic activity also differentiated subjects in the Stop task. (B). rVLPFC showed trial-type-specific recruitment that was consistent across tasks, contradicting stopping-specific accounts of rVLPFC function. ** $p<.0001$ *** $p<.005$.

In a second multi-voxel pattern analysis, subject-specific classifiers were trained to decode the multivariate patterns that differentiate Double Go_{Signal} and Double $Go_{No\text{-}Signal}$ trials. Classifiers generalized this training on the Double Go task to correctly identify $Stop_{Signal}$ trials with 92-97% accuracy in the rVLPFC, significantly higher than the 59% accuracy achieved in primary motor cortex ($F(1,17)=9.413$, $p=.007$). To control for the possible effects of classifier bias on this result, we utilized signal detection theory. Classifiers readily discriminated between $Stop_{Signal}$ and $Stop_{No\text{-}Signal}$ trials in the rVLPFC but not primary motor cortex in terms of d' (Fig.4B; $F(1,17)=13.14$, $p<.005$), indicating that rVLPFC similarly encodes the different

processes invoked by Signal and No-Signal trials, despite their different demands on motoric stopping, unlike the sensitivity shown by primary motor cortex to the demands on motoric stopping invoked by $Stop_{Signal}$ trials.

Event-related potentials

The ERPs evoked by our tasks also reflected context-monitoring demands rather than stopping demands. In particular, motoric stopping accounts predict that a prefrontal ERP called the "Stop P3" reflects stopping-specific processes (Smith et al., 2008) and should therefore be enhanced in the Stop task. However, the so-called Stop P3 was enhanced in the Double Go task, in direct contradiction to the stopping account (Fig.5A; $t(35)=2.92$, $p<.03$), but consistent with our observations of increased transient hemodynamics in the Double Go task relative to the Stop task (see Univariate fMRI results, above).

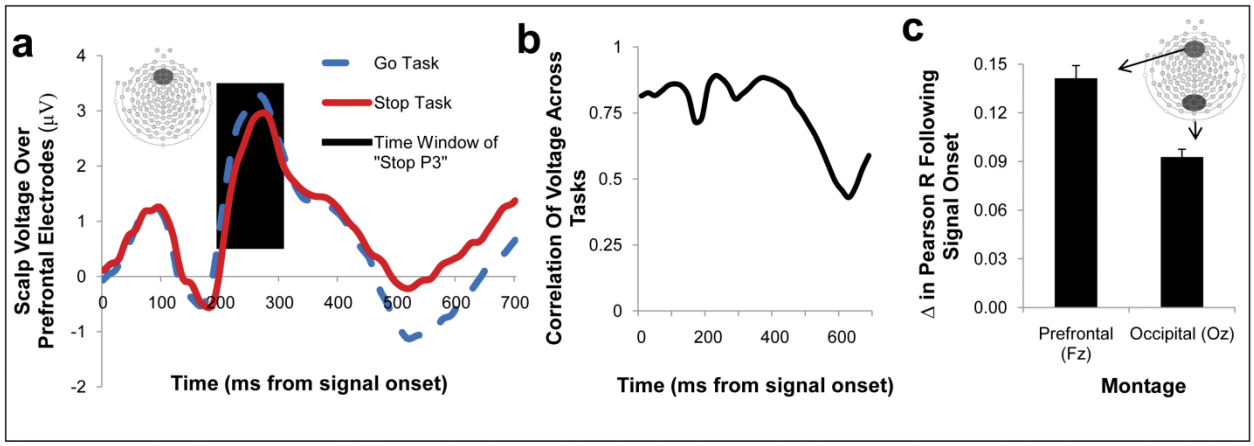


Figure 5. A prefrontal positivity peaking around 300ms, known as the "Stop P3," has been previously associated with stopping, but this component (darkened region of A) was significantly enhanced in the (Double) Go task. Individual differences in voltage were also highly correlated across tasks, indicating substantial overlap in the underlying cortical processes (B). Moreover, prefrontal correlations between the scalp voltage recorded across tasks were disproportionately increased following the presentation of the signal, relative to the increase in occipital correlations

observed at the same time (C). This difference indicates increased cross-task similarity in prefrontal processing specifically at signal onset.

These ERPs from the Stop and Double Go tasks were not two distinct potentials masquerading as the same; individual differences in this ERP were also highly correlated between tasks (Fig.5B). Critically, correlations between the ERPs elicited by each task were disproportionately increased over prefrontal electrodes, relative to occipital electrodes, following signal onset, when context-monitoring is most required but stopping demands differ most (Fig.5C; $F(1,98)=12.59$, $p=.001$).

Pupillometry

We also found that context-monitoring, not motoric stopping, explains the patterns of mental effort elicited during our tasks. We measured pupil diameter, a psychophysiological index of mental effort (Kahneman & Beatty, 1967; Beatty & Lucero-Wagner, 2000), following the onset of a signal (or the average signal onset time in the case of No-Signal trials; Fig.6). Averaging across all time points, mental effort was less for stopping than for monitoring for signals that fail to appear ($Stop_{Signal}<Stop_{No\text{-}Signal}$ $t(85)=7.00$, $p<.001$; $Stop_{Signal}<$Double $Go_{No\text{-}Signal}$ $t(85)=2.07$, $p<.05$). Mental effort was also less for context monitoring and motoric stopping than for context monitoring and an additional act of going ($Stop_{Signal}<$Double Go_{Signal} $t(85)=13.67$, $p<.001$). Finally, mental effort was greater when monitoring for signals that would require a change to the planned response than when monitoring for those that would not ($Stop_{No\text{-}Signal}$ >Double $Go_{No\text{-}Signal}$ $t(85)=10.25$, $p<.001$), a result which also rules out global reductions in effort during the Stop task (e.g. from fatigue; see also supporting discussion). Thus, motoric stopping

is not itself associated with any effort beyond that required for the processes involved in other trial types, contrary to the idea that motoric stopping of a response constitutes the cognitively-controlled, and therefore effortful, component of response inhibition. Instead, context-monitoring demands are more central to mental effort, and this relationship is modulated by the relevance of the monitored stimulus to the planned response.

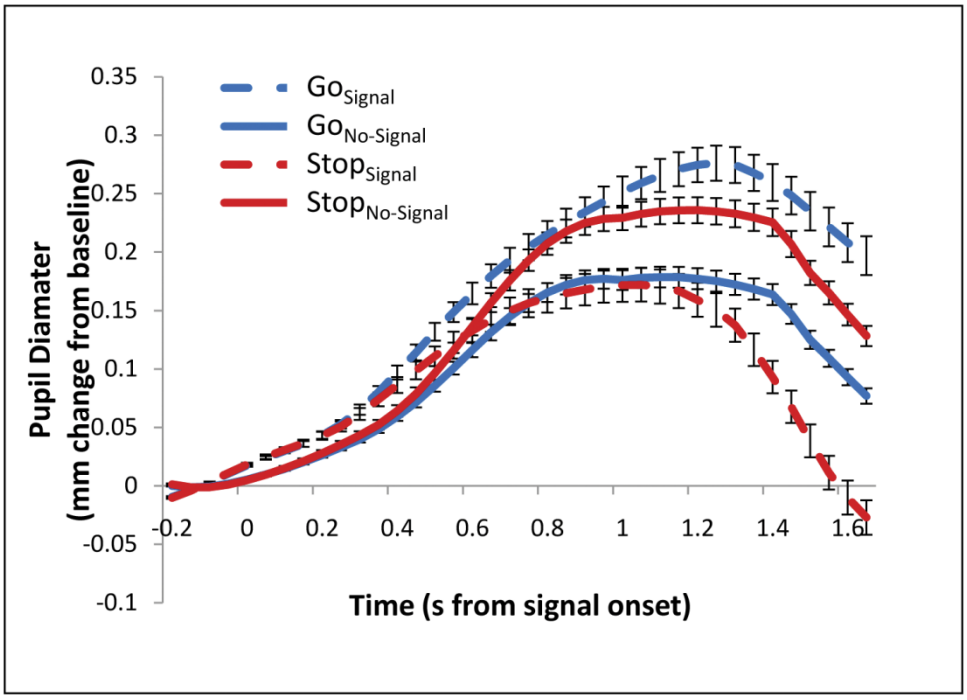


Figure 6. Patterns of mental effort assessed via pupillometry indicate that effort matches demands on context-monitoring, not stopping, and is modulated by the relevance of the infrequent stimulus to the planned response. In particular, stopping a response ($Stop_{Signal}$ trials) was associated with more mental effort was required by monitoring for the appearance of stimuli that would demand stopping ($Stop_{No\text{-}Signal}$ trials) than by stopping itself ($Stop_{Signal}$ trials) or by monitoring for the appearance of stimuli that would demand an additional act of going ($Go_{No\text{-}Signal}$ trials).

Model-based decomposition of behavior and correlations with brain activity

Stopping is not associated with differential mental effort or prefrontal recruitment, contrary to stopping-centric accounts of cognitive control. This pattern of results could imply

that motoric stopping is not a cognitively-controlled process, in the sense that it may not be employed in a goal-directed manner, given the widely-held assumption that goal directed behavior recruits the prefrontal cortex and requires mental effort. Consistent with this idea, subjects appeared to engage transient stopping on the Double Go task even though such stopping runs contrary to goals in this task. Specifically, although Double Go_{Signal} trials require that subjects commit a subset of the motor responses required on Double $Go_{No\text{-}Signal}$ trials, subjects were nonetheless slower to provide a response to stimuli when they were followed by the signal than when they appeared alone (Double $Go_{Signal}^{1st\ RT}$>Double $Go_{NoSignal}^{Only\ RT}$; $t(148)=9.59$, $p<.0005$; Fig.7A). To the extent that this behavioral slowing in the Double Go task reflects some transient stopping, it runs contrary to subjects' goals in the Double Go task and therefore may not be engaged in a controlled manner.

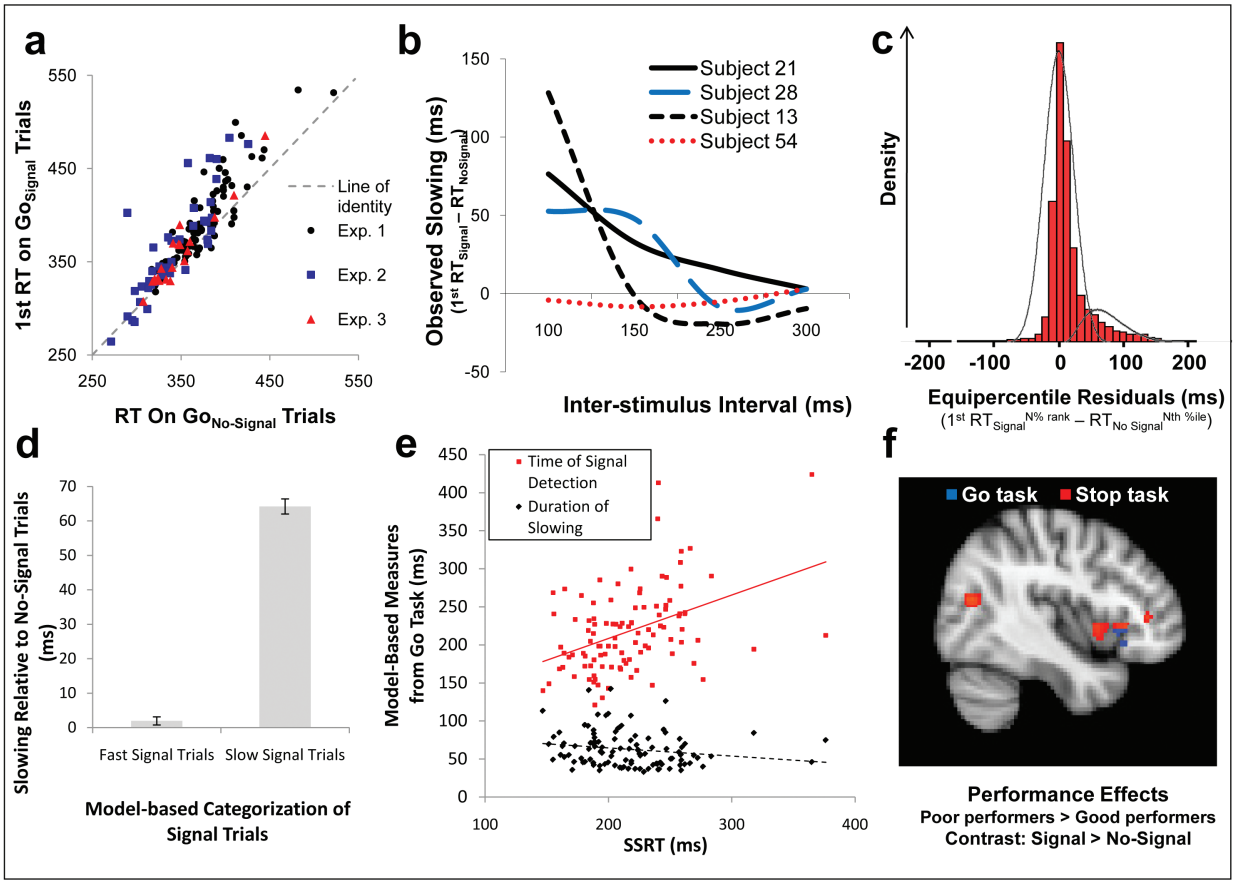


Figure 7. Response slowing was observed in the Double Go task (A), perhaps suggesting that stopping is not associated with differential mental effort or prefrontal activity because it is an automatic consequence of detecting an infrequent stimulus. Critically, this slowing was dependent on ISI; indeed, large individual differences were observed in the shortest ISI to yield zero slowing (B contains data from four representative subjects). A subtraction of reaction times on (Double) $Go_{No\text{-}Signal}$ trials from those with a corresponding percent rank on (Double) Go_{Signal} trials reveals a pronounced positive skew to these equipercentile residuals (C), indicating that some proportion of reaction times on Go_{Signal} trials are disproportionately delayed. Trials undergoing this slowing were identified as those more likely to come from a distribution not centered on zero, as determined through a two-component mixture model (see overlaid lines on histogram in C). This procedure adequately separated the slowed and unslowed distributions, as revealed by zero significant difference between Go_{Signal} trials categorized as unslowed and their corresponding reaction times in the $Go_{No\text{-}Signal}$ distribution, but a large difference between Go_{Signal} trials categorized as slowed and their corresponding reaction times in the $Go_{No\text{-}Signal}$ distribution (D). From this we estimated two individual differences: how long subjects are slowed (duration of slowing; DoS) and the time at which signals are detected (time of signal detection; TOSD). Only TOSD positively correlated with SSRT, whereas DoS showed a slight negative correlation, indicating that the slowing experienced by subjects in the Double Go task cannot be the source of shared variance between the Stop and Double Go tasks (E). Brain-behavior correlations

confirmed this conclusion: SSRT and TOSD, but not DoS, overlapped in their correlations with neural activity only in the rVLPFC (F).

An alternative interpretation of this slowing is that it does in fact reflect a controlled and goal-directed process: it may be an attempt to stop or replace the motor plan required on Double $Go_{No\text{-}Signal}$ trials (i.e. the motor plan for "respond once" is stopped or replaced with the motor plan for "respond twice"). We assessed this possibility with a model-based decomposition of subjects' behavior; however, the results of this analysis argue against this possibility, and further show that the efficiency of subjects' context-monitoring, rather than the efficiency of motoric stopping or motor plan replacement, shares a closer relationship with SSRT.

To assess the alternative accounts, we developed a formal model of context-monitoring and stopping by building on the classic race model of the Stop task (Logan & Cowan, 1984) in order to precisely estimate the duration of motoric slowing experienced by subjects in the Double Go task, as well as exactly which trials underwent such slowing. The race model of the Stop task posits that responses undergo inhibition when a stopping process, triggered by the onset of the Stop signal, completes *before* the "going" processes triggered by the onset of the 2AFC stimulus. The race between stopping and going processes is the model's namesake, and is supported by the monotonically-decreasing relationship of interstimulus interval (ISI) to successful inhibition: larger ISIs give the "going" process an increasing advantage in the race, and thus leads to less successful inhibition. We observed a similar phenomenon in our Double Go task, such that increasing ISIs led to less slowing of first responses; this effect was visible at the group level (see additional results in Supporting Information) but also even at the level of individual subjects (Fig. 7B), who showed substantial variability in the earliest ISI to yield zero observable slowing.

We utilized this behavioral variability to estimate individual differences in Double Go task performance. First, we estimated the probability that each trial belonged to either the "slowed" or "unslowed" distribution of reaction times. This categorization was accomplished by fitting a mixture model to the difference between reaction times of Double $Go_{Signal}^{1st\ RT}$ and Double $Go_{NoSignal}^{Only\ RT}$ trials of corresponding percent rank. To the extent these reaction times come from the same (i.e., unslowed) distribution, these equipercentile residuals should be centered on zero; however, there was pronounced positive skew (Fig. 7C), indicating that a substantial proportion of trials did undergo slowing. We considered as "slowed" those trials that were marginally less likely to come from a Gaussian distribution centered on zero, relative to an alternative distribution with a positive mean (see overlaid curves on Fig 7C, and supporting methods). This method clearly separated "slowed" from "unslowed" trials on the basis of the first RT on Double Go_{Signal} trials: "unslowed" trials showed approximately zero slowing relative to corresponding trials within the No Signal distribution, whereas "slowed" trials were significantly longer than corresponding trials within the No Signal distribution (Fig. 7D).

Next, we estimated for each subject the amount of time that must elapse after signal presentation until responses are categorized as "slowed" (yielding the time of signal detection [TOSD], our measure of context-monitoring), and the difference between that subjects' "slowed" and "unslowed" reaction times (yielding the duration of slowing [DoS], our measure of stopping from the Double Go task). If motoric stopping (or, equivalently, motor plan replacement) is controlled, and initiated in this controlled fashion in the Double Go task, then the process of motoric stopping or motor plan replacement should should cease (as estimated by DoS, in the Double Go task) in proportion to how quickly competing motor plans can be stopped, as assessed

by SSRT in the Stop task. That is, the "controlled motoric stopping" and "controlled motor plan replacement" accounts both predict that DoS and SSRT should be positively correlated.

However, DoS and SSRT were not positively correlated – they instead showed a weak negative correlation (Pearson R=-.188, p=.048; Fig 7E), in direct contradiction to the prediction motivated by these alternative accounts. SSRT was instead positively correlated only with TOSD – i.e., the efficiency with which signals could be detected (Fig. 7E; R=.418, p<.0005) – as predicted by accounts which posit that context-monitoring underlies the commonalities of the Double Go and Stop Signal tasks. This positive relationship persisted when controlling for DoS (R=.410, p<.0005), indicating that the overlapping variance in TOSD and SSRT does not reflect motoric stopping or motor plan replacement. Strikingly, this relationship of context-monitoring to SSRT was also regionally-specific: SSRT and TOSD overlapped in their relationship to hemodynamics only within the rVLPFC (Fig.7F).

A second, independent assessment of the origin of the observed commonalities across our tasks is also enabled by our formal model. Specifically, the model identifies exactly which trials undergo motoric stopping/slowing within the Double Go task, and thus permits these trials to be excluded from analysis. To the extent that similar hemodynamic, electroencephalographic, and pupillometric patterns are observed when these "slowed" trials are excluded, it would suggest that the commonalities across our tasks do not reflect a motoric stopping process common to these tasks.

Consistent with the claim that a common and cognitively-controlled process of context-monitoring – and not a common process of motoric stopping – underlies the commonalities of our tasks, a complete re-analysis of the data without such "slowed" trials replicated all of our primary results: the increased transient hemodynamic response in the rVLPFC during the Double

Go task, the prominent sustained hemodynamic activity observed in that task, the multivariate hemodynamic commonalities across tasks, the increased Stop P3 response in the Double Go task, the strong correlations of scalp voltage across tasks as well as the selective increase in those correlations over frontal electrodes following signal onset, and yields qualitatively similar patterns of mental effort (see Supporting Methods and Supporting Table S4). This analysis further substantiates our conclusion that context-monitoring, not motoric stopping, reflects the cognitively-controlled component of this canonical response inhibition task.

Discussion

By matching our tasks on all characteristics except motoric stopping demands, we find that monitoring context for behaviorally-relevant signals, not stopping, is an effortful, controlled, and prefrontal process that explains individual differences in cognitive control during response inhibition. We were able to replicate all of our primary results when utilizing only the trials that were categorized as "unslowed" from the Double Go task, indicating that the slowing in that task was not the origin of the hemodynamic, electroencephalographic, and pupillometric commonalities of the Stop and Double Go task. This conclusion is consistent with recent evidence that the behavioral slowing expressed in context-monitoring tasks is not related to hemodynamics in rVLPFC, nor to that in any portion of lateral prefrontal cortex (Sharp et al., 2010). In contrast, SSRT was instead more closely related to our measure of context monitoring.

More broadly, our conclusions are also largely concordant with comparisons of the same Double Go task we used above (Dodds et al., 2011) and alternative context monitoring tasks (Sharp et al., 2010; Cai & Leung 2011; Hampshire et al., 2010) with the Stop Signal task, as

described below. It nonetheless remains possible that the prefrontal cortex could subserve some form of motoric stopping, or motor plan replacement, or that motoric stopping could in some cases be cognitively controlled. Our results indicate only that there is no need to assume that motoric stopping occurs in a cognitively controlled fashion within the canonical task of response inhibition, the Stop task. Instead, many of the phenomena from this task – including both transient and sustained hemodynamics, multivariate patterns in those hemodynamics, event-related potentials, mental effort as quantified through pupillometry, and the primary behavioral measure from this task – seem to primarily reflect this task's demands on context monitoring processes.

Relation to Recent Work.

Our study addresses the evolving debate on the functional specialization of the rVLPFC in three ways: by developing a formal model, by distinguishing subprocesses within these tasks that may have led to otherwise unresolved discrepancies across previous findings, and by testing a question of different scope: whether by any major criteria, motoric stopping could be considered a specific cognitive control mechanism utilized during response inhibition. Previous formal models did not separately account for motoric stopping and monitoring, or even explicitly distinguish between them (Logan & Cowan, 1984). Previous empirical attempts to dissociate monitoring and motoric stopping yielded conflicting results: less, more, or equivalent recruitment of either the same or separable subregions of rVLPFC (Sharp et al., 2010, Hampshire et al., 2010; Dodds et al., 2011). Finally, previous neuroimaging work has largely focused only on transient prefrontal hemodynamics in context-monitoring and response inhibition tasks. By investigating not only transient but also the sustained and effortful components to inhibitory control, their goal-directedness, and the extent to which they drive individual differences in

behavior, event-related potentials, and multivariate hemodynamics, we demonstrate the importance of context-monitoring as a mechanism of cognitive control.

Nonetheless, our results may seem to stand in contrast to some conflicting findings of previous work. Below, we step through these findings and describe how the context-monitoring account may offer a consistent way to understand these otherwise contradictory results, with respect to issues of statistical efficiency in estimating the hemodynamic response, potential dissociations between inferior and superior rVLPFC, the role of arousal, and the potential importance of goal replacement in the Stop task.

Statistical Efficiency.

In previous neuroimaging work that included two types of Signal trials within the Stop task – signal trials that require stopping and "distractor" signal trials which indicate no change to the planned response – the latter activated parts of the rVLPFC that are either spatially indistinguishable or spatially distinct from the portions of the rVLPFC recruited by the former. These conflicting results may indicate that the hemodynamic responses to these two types of signal trials were estimated with differential efficiency.

Indeed, when statistical efficiency is precisely matched across tasks, as in one previous study (Hampshire et al., 2010), response inhibition in the Stop task is actually associated with a decreased transient hemodynamic response relative to tasks involving response commission (consistent with our ERP and fMRI results). This previous study also demonstrated that rVLPFC activity was increased to infrequent stimuli that required either response commission or response inhibition, relative to infrequent stimuli that required no overt behavior. This result can be

viewed as consistent with context-monitoring, which posits the rVLPFC is involved in the detection and interpretation of behaviorally-relevant stimuli to guide the selection of action.

Dissociations between inferior/superior lateral PFC.

Some studies suggest that a region of superior rVLPFC may be more crucial for processes analogous to context-monitoring than an inferior section of rVLPFC, which is more crucial for stopping (Chikazoe et al., 2009; Verbruggen et al., 2010; Cai & Leung, 2011). In our results, these inferior and superior rVLPFC regions were distinctly activated, but both foci were more strongly activated on Double Go_{Signal} trials than $Stop_{Signal}$ trials (Fig. 3B). Thus, while inferior and superior rVLPFC could differentiate in principle, a simple monitoring vs. motoric stopping dichotomy is not sufficient for explaining the patterns observed here.

Instead, the apparent dissociations between superior and inferior rVLPFC observed previously may represent differences in efficiency across trial types, as described above, or the absence of a viable model to analyze behavior in paradigms with infrequent response commission trials. Specifically, the behavioral model used in one recent TMS study of the functional specialization of the rVLPFC (Verbruggen et al., 2010) assumes that the first response is unaffected by the appearance of the signal. However, we found that the first response is slowed by the appearance of a signal (Fig. 7A), and that measures extracted from these dynamics correlate selectively with rVLPFC (Fig. 7F). These results challenge the assumptions of the behavioral model in the TMS study, and the associated claims of dissociations between inferior and superior rVLPFC.

A second recent TMS study applied a "conditioning" pulse to the rVLPFC prior to a "test" pulse to primary motor cortex, and demonstrated a reduction in the observed motor-evoked potential (MEP) (Neubert et al., 2010). The authors interpreted their results to reflect a direct

inhibitory role of the rVLPFC on M1. However, similar effects have been observed with conditioning pulses to dorsal prefrontal cortex (Civardi et al., 2011). It is thus likely that these TMS effects reflect relatively general mechanisms (e.g., short- or long-interval intracortical inhibition) that are not functionally specific to the rVLPFC.

The role of arousal.

Previous work has reported a null effect of TMS to rVLPFC on pupil diameter using a relatively small sample of 17 subjects (Chambers et al., 2006), contrary to what might be expected if effortful monitoring processes were disrupted by rVLPFC TMS. However, we note that TMS did lead to a consistent reduction in pupil diameter on correctly inhibited trials across all time points (Fig. 6D of Chambers et al, 2006). Moreover, subsequent work has identified a generalized arousal effect of TMS (Verbruggen et al., 2010), which may have increased pupil diameter and thus masked decreases in pupil diameter after disruption of context-monitoring. Thus, to the extent any conclusions can be drawn on a null effect in a small sample, this previous result may suggest that the disruption of effortful context-monitoring processes (and consequently, decreases in pupil diameter) was only partially offset by the generalized arousal resulting from rVLPFC TMS.

The relationship of response inhibition to goal-switching.

Some theoretical accounts of Stop task performance emphasize the importance of goal-switching, such that a "Go goal" must be replaced with a "Stop goal" on $Stop_{Signal}$ trials (Verbruggen & Logan, 2009). However, factor analytic studies of individual differences in response inhibition (including the Stop task) and those in task-switching and working memory updating indicate that there is significant switching- and updating-specific variance in individual differences that is non-overlapping with that in performance on response inhibition tasks

(Friedman et al., 2008). The prevailing interpretation of these findings is that performance on response inhibition tasks is primarily driven by cognitive control processes supported by active maintenance mechanisms, and are thus common across all executive function tasks, rather than being specific to goal- or task-switching, or to working memory updating processes (Friedman et al., 2008). Our ongoing neurocomputational modeling work indicates that such active maintenance mechanisms, used in the service of context-monitoring, may indeed be sufficient for explaining detailed patterns of performance on the Stop task. This conclusion is further consistent with the fact that task-switching and Stop task performance do not influence one another (i.e., their effects are additive) (Verbruggen, Liefhooge, Szmalec, & Vandierendonck, 2005), a result which argues against the idea that a controlled goal-switching process is operative during the Stop task.

Broader Implications

Our results contradict a long-standing and currently-dominant account of cognitive control in response inhibition tasks and demonstrate that a role for the rVLPFC in transient stimulus processing is not mutually-exclusive with the sustained prefrontal dynamics emphasized by neuropsychological and neurocomputational theories (Stuss & Alexaner, 2007; Miller & Cohen, 2001; O'Reilly, 2010).

The context-monitoring account of rVLPFC function also supports emerging taxonomies of prefrontal organization. According to one recent taxonomy (O'Reilly, 2010), ventral areas of the prefrontal cortex are particularly important for contextual processing of stimulus significance (broadly speaking, "what" processing) whereas more dorsal areas may be particularly important for the processing contextually-appropriate responses to those stimuli ("how" processing"). The putative localization of context-monitoring to rVLPFC is fully compatible with this framework.

To the extent that response-related "stopping" processes have a dedicated prefrontal substrate, this taxonomy predicts that such processes should localize to areas dorsal to the rVLPFC, some of which project to the STN with similar or greater density (Nambu et al., 2002). However, our results do not suggest that dorsal areas are differentially associated with stopping, consistent with the wider literature, and further supporting our conclusion that motoric stopping does not have a dedicated lateral prefrontal substrate within the Stop task.

The context-monitoring role of rVLPFC may also be understood as arising from the proximity of rVLPFC to the anterior insula, which appears to monitor interoceptive information (Craig, 2009), in some cases proactively (Lovero, Simmons, Aron & Paulus, 2009). The anterior insula also shows greater hemodynamic responses to demands on action selection than to demands on motoric stopping, and is thought to be tightly integrated with the rVLPFC (Lim, Padmala & Pessoa, 2009). Thus, a basic mechanism in anterior insula for monitoring the internal significance of upcoming stimuli may have been evolutionarily adapted for use in monitoring their relevance for action selection in the nearby rVLPFC. These representations may even be tightly integrated, such that context monitoring can be effectively recruited when it counts most – under conditions of threat or pain. Indeed, target detection is improved when the targets are predictive of pain, an effect that is associated with greater activity in both rVLPFC and anterior insula (Lim, Padmala & Pessoe, 2009).

The context-monitoring account is also compatible with recent revisions to a classical taxonomy of the effects of prefrontal insult, in which the inhibitory deficits arising from right lateral prefrontal damage are now explained as monitoring deficits instead (Stuss & Alexander, 2007). The match between our findings and those motivating this taxonomic revision may indicate the need to rethink a broad range of putative inhibitory deficits. For example, focal

rVLPFC damage can lead to poor target detection, such that even when the location of an upcoming target is cued before trial onset, this location is not effectively monitored following the onset of any stimulus (Michael, et al., 2006) . While this deficit might reflect problems with inhibiting locations in space, our results suggest this patient's focal rVLPFC damage may have yielded a deficit in monitoring contextually-appropriate locations in the service of target detection and action selection.

In addition to the significance of our result for understanding neural insult, our result may also impact the treatment of pathological impulse control deficits (e.g. as in substance abuse or attention-deficit hyperactivity disorder [ADHD]). Specifically, our result suggests that pathological impulse control deficits might not reflect a failure to stop in particular, but rather the more effortful and prefrontal processes involved in context-monitoring. For example, ADHD may be associated with a monitoring deficit in which many stimuli, regardless of their behavioral-relevance, are thought to warrant attention. This prediction is supported by the finding that ADHD is more strongly associated with increased reaction time variability, as might result from a context-monitoring deficit, than with deficits in tasks that require stopping (Castellanos et al., 2006). Relatedly, the resistance of response inhibition to improvement via training (Thorell et al., 2009) may reflect that monitoring context for contingent action selection, not the act of stopping, is the controlled process to be targeted for effective intervention.

CHAPTER 3: A NEUROCOMPUTATIONAL MODEL OF MONITORING AND STOPPING IN RESPONSE INHIBITION

Abstract

Although the importance of motoric stopping has long been emphasized in the domain of response inhibition, recent evidence suggests that this task is subserved by highly domain-general attentional mechanisms. In particular, it has been argued that the role of the prefrontal cortex in the canonical test of response inhibition – the Stop task – is to monitor the environmental context for behaviorally-relevant stimuli, and further that the prefrontal cortex subserves no stopping-specific function in this task. Here we assess this hypothesis with a computational model of the Stop task using the Prefrontal Basal Ganglia Working Memory architecture, which is here extended to include a simulated subthalamic nucleus. Our model matches data from behavioral, hemodynamic, and pharmacological studies of response inhibition, and further confirms recent claims that monitoring (rather than stopping) better characterize the computational role of prefrontal cortex in this classic test of response inhibition.

Introduction

The construct of inhibition has been fractionated into multiple heterogeneous abilities (Nigg, 2000; Friedman & Miyake, 2004; Aron; 2007), but modern work has focused particularly intensively on the processes supporting the inhibition of responses. The processes are thought to most conspicuously include the controlled stopping of unwanted motor actions (Logan &

Cowan, 1984). Deficits in this ability in particular have been argued to underlie a variety of psychopathological disorders (e.g., Barkley et al., 1994; Shanahan, Pennington & Willcutt, 2008) and to be a central characteristic of frontal lobe injury and immaturity (e.g., Diamond et al., 1990).

One task has emerged as a canonical test of response inhibition because it permits a precise estimate of the latency for response inhibition to take place (Logan & Cowan, 1984). In the Stop task, subjects are presented with an imperative stimulus that requires a response (typically a 2-alternative forced choice; 2AFC). On a subset of trials (so-called "Signal" trials), this 2AFC is followed with a variable delay by a stop signal indicating that no motor response should be executed on that trial. By titrating the delay between this 2AFC stimulus and the stop signal, one can estimate the latency of the latent "stopping process" (an estimate known as the Stop Signal Reaction Time, or SSRT) under the assumption that this stopping process initiated on Signal trials is engaged in a race with a "going" process that is similarly initiated on Signal and No Signal trials alike. The assumptions underlying this calculation have been formalized as a race model (Logan & Cowan, 1984), which can in turn be captured as a diffusion model (Shenoy & Yu, 2011; Verbruggen & Logan, 2009) and as a two-unit neural network (Boucher et al., 2007).

Central to the theories underlying these successful models, and the models themselves, is a mechanism that is explicitly dedicated for motoric stopping. Indeed, it is often presumed that the involvement of this motoric stopping mechanism is the defining feature of response inhibition paradigms. Both forward and reverse inferences have been common in this domain: If a task demands motoric stopping, then neural regions recruited by that task are often interpreted to reflect motoric stopping processes; when other tasks activate that region, they are sometimes

presumed to involve a motoric stopping process as well. Importantly, the models underlying these inferences have been "built to order" for the underlying theories – they are not extensions to more domain-general systems that can also perform tasks which do not explicitly require motoric stopping.

Accumulating empirical work, however, suggests that such dedicated motoric stopping mechanisms may not be of such central importance to response inhibition tasks. Studies of individual differences suggest that the higher cognitive abilities that drive performance on the Stop task, and other tasks involving motoric stopping, are overlapping with those driving performance on tasks with little apparent demands on motoric stopping (Friedman et al., 2008; Friedman & Miyake, 2004). Likewise, experimental work also indicates that common mechanisms drive performance on response inhibition tasks and those that do not involve motoric stopping (Kramer, Humphrey, Larish, Logan & Strayer 1994; Ridderinkhof, Band & Logan, 1999; Verbruggen, Liefooghe, Notebaert & Vandierendonck, 2005;Verbruggen, Liefooghe & Vandierendonck, 2004, 2006; van den Wildenberg & van der Molen, 2004).

Functional magnetic resonance imaging (fMRI) indicates that these domain-general mechanisms may in fact have direct correlates at the implementational level (Marr, 1982). The right ventrolateral prefrontal cortex (rVLPFC) is the cortical area most consistently associated with inhibition across numerous domains (Cohen, Berkman & Lieberman, 2010), and is particularly reliably associated with tasks that demand motoric stopping (Aron et al., 2007) but is now also recognized as performing a more general function than motoric stopping (Munakata et al., in press). For example, the rVLPFC is often more strongly activated by tasks that require the infrequent commission of a response than by those that require the inhibition of a response (Chatham et al., 2011a, Dobbs et al., 2010; Hampshire et al., 2010). Similarly, when transcranial

magnetic stimulation (TMS) is applied to this region, the induced deficits in inhibitory performance are paralleled by equivalent deficits in target detection, even in the absence of distracting sensory information (Verbruggen et al., 2010). These results suggest that the rVLPFC is not specific to tasks that demand motoric stopping, but may instead subserve more abstract goals, such as those that involve monitoring for behaviorally-relevant information in the current context (so-called "context monitoring"). This ability may be particularly important for Stop task performance (Chatham et al., 2011a) as well as for performance in domains to which performance on the Stop task relates.

Recently, multivariate pattern analysis has demonstrated evidence uniquely consistent with the use of this kind of domain-general context monitoring process. In particular, a multivariate classifier was trained to identify individual subjects on a task that involves similar context monitoring processes as the Stop task, but does not require motoric stopping. This classifier was then able to generalize this learning to identify individual subjects performing the Stop task on the basis of hemodynamic responses in the rVLPFC, and did so better on the basis of rVLPFC responses than on the basis of responses in primary motor cortex. This pattern was more pronounced on Signal trials in particular, relative to the more frequent No Signal trials, indicating that such "hemodynamic fingerprints" are highly consistent across tasks regardless of the demands of Signal trials on motoric stopping mechanisms. This evidence could imply that context-monitoring processes are a special case of more abstract goal maintenance mechanisms within the prefrontal cortex (Miller & Cohen, 2001), such that the rVLPFC-represented goal in these tasks is effectively "look for behaviorally-relevant stimuli."

Here we provide a computational test of this hypothesis, as well as the idea that the role of the prefrontal cortex in this task – putatively for "context-monitoring" (Chatham et al., 2011)

– can be understood as a particular instantiation of goal maintenance. We do so in the context of simulated subcortical mechanisms, including the subthalamic nucleus, which have been widely associated with response inhibition. This framework also allows us to assess whether some functionally specific motoric stopping process would emerge in the prefrontal layers of our model – offering a further test of the idea that such stopping-specific mechanisms are not localized to the prefrontal cortex (Chatham et al., 2011). If correct, our hypothesis would significantly circumscribe the currently central role of specialized motoric stopping mechanisms in extant theorizing on response inhibition, and in associated "built to order" computational models.

To test these predictions, we implemented a neural network model of the Stop task using an adaptation of the Prefrontal/Basal Ganglia Working Memory (PBWM) architecture (O'Reilly & Frank 2006; Hazy, Frank & O'Reilly 2007). The core components of this architecture are illustrated in Figure 8. PBWM utilizes a combination of Hebbian and error-driven learning rules to adapt the connection weights throughout a series of layers that simulate both posterior cortical sensorimotor processing, as well as prefrontal mechanisms for active maintenance. These prefrontal layers are unique relative to others in the model insofar as they are capable of stably maintaining such information over time (by virtue of self-excitatory connections and intrinsic maintenance currents). They are also unique insofar as they are also capable of flexibly updating this information based on activity elsewhere in the model, if a "gating" signal is triggered by subcortical striatal mechanisms. These subcortical mechanisms, implemented as distinct layers in the model, are trained to initiate such gating signals via a biologically-plausible reinforcement learning rule based on subcortical dopamine function (the Primary Value/Learned Value algorithm; PVLV). PBWM models have simulated a wide range of tasks, including those

involving task-switching (O'Reilly & Frank, 2006) and working memory updating (O'Reilly & Frank, 2006; Chatham et al., 2011), suggesting that they should be sufficient for capturing phenomena in response inhibition tasks.

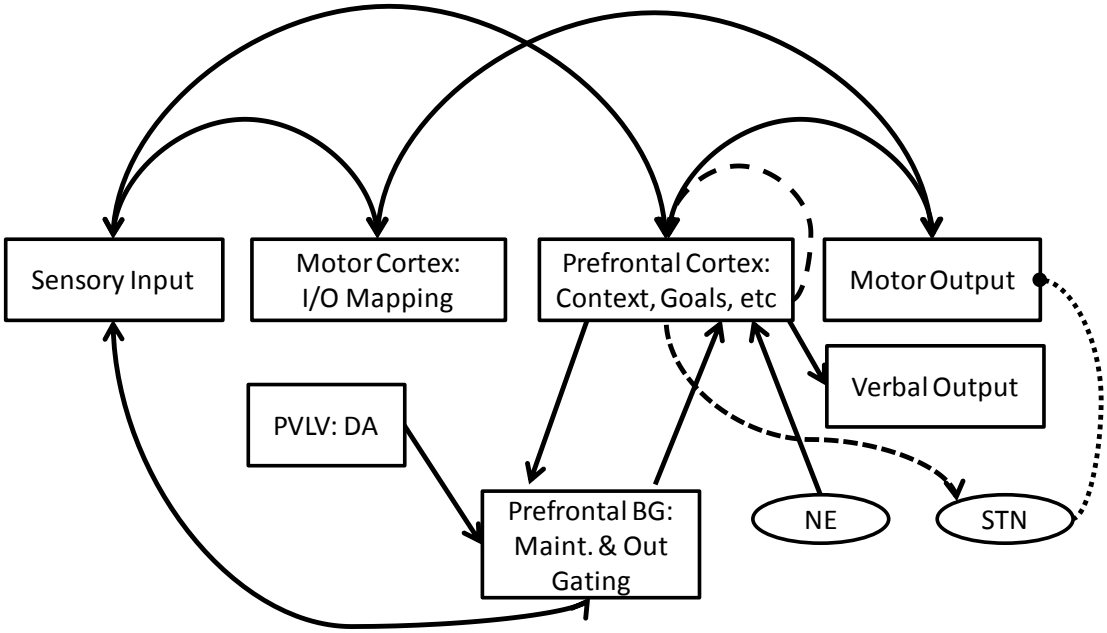


Figure 8. Schematic illustration of extension to the PBWM architecture, in which prefrontal context representations of relevant prior information and current goals bias the sensory-motor mappings that are learned by posterior cortical "hidden" layers. The prefrontal context representations are updated via dynamic gating by the basal ganglia. These gating functions are learned by the basal ganglia on the basis of input from the PVLV system, which provides modulatory dopaminergic input depending on the reward value of the actions performed by the basal ganglia.

One way in which we extended the PBWM model was to include a layer corresponding to the subthalamic nucleus (STN). The STN is part of an inhibitory subcortical circuit, such that its activation has a rapid and net inhibitory effect on thalamic output (e.g., Nambu et al., 2002). Computational modeling and empirical work both suggest that the role of the STN may be to dynamically adjust decision and response thresholds (e.g., Frank et al., 2007). Here, we opted for an abstract implementation of this computational function, such that activation in an STN layer

directly determined the response threshold of the network. By this scheme, even maximal output unit activation was considered insufficient for generating a response when the STN was highly activated; conversely, when the STN was minimally active, a relatively small amount of activation in the output layer was considered sufficient for generating a response. This simulated STN receives projections from the prefrontal layers of the network, consistent with the widespread projections to STN from numerous prefrontal areas (including ventrolateral, dorsalateral, and dorsomedial prefrontal cortex; Nambu et al., 1997; Nambu et al., 2002).

A second extension to the PBWM model involved a neuromodulatory input to prefrontal cortex intended to simulate the effect of prefrontal norepinephrine (NE) neuromodulation. NE mechanisms are important for Stop task performance (Aston-Jones & Gold, 2009; Chamberlain et al., 2009; Overtoom, 2003), and may selectively activate the rVLPFC during inhibitory control (e.g., Chamberlain, et al 2009). The possible computational role of NE mechanisms is less well understood than that of the STN, but it has been suggested that phasic NE release may accompany the presentation of Stop signals due to their low frequency or unexpectedness (Dayan & Yu, 2006; Frank, Santamaria, O'Reilly & Wilcutt, 2007). Here we simply assumed that such phasic NE release occurs during the presentation of Stop Signals and provides an additional excitatory signal to prefrontal areas. We then tested whether the model would reproduce the observed reduction in SSRT that accompanies potentiation of NE mechanisms (Chamberlain et al., 2009).

To foreshadow our results, the model demonstrates that the relatively domain-general mechanisms implemented by PBWM are indeed capable of simulating a wide variety of data from the Stop task, and do so in a way that matches a number of detailed findings. After demonstrating the ability of the model to capture these hallmark phenomena from the Stop

Signal task (including a number of behavioral effects, as well as canonical results from neuroimaging and psychopharmacological manipulations of noradrenergic mechanisms), we present a descriptive analysis of the model's behavior to illuminate how such benchmark phenomena naturally arise from these domain-general mechanisms.

Methods

Implementation.

Our model is implemented using the Local, Error-driven and Associative Biologically Realistic Algorithm (Leabra) framework (O'Reilly, 2001). In Leabra, neural processing occurs across a series of interconnected units each of which has a membrane potential with separate excitatory, inhibitory and leak conductances. A rate-coded output is derived from fluctuations in this membrane potential and this output contributes to the excitatory conductance of all units to which a given unit is connected, in proportion to the connection weight between them. Although initially randomized, these connection weights are adjusted over training according to Hebbian, reward-driven, and biologically realistic error-driven learning rules (see Appendix I). Units are further grouped into layers that undergo a k-winners-take-all function for simulating the influence of local inhibitory interneurons. Leabra has been used in over 40 models of a variety of cognitive phenomena (e.g., O'Reilly & Munakata, 2000) and thus constitutes a biologically-constrained formalism that yields human-like performance across many domains.

Each named layer of the implemented model contains features which uniquely associate its layers with the identified brain regions. For example, prefrontal layers are unique due to recurrent connections and an excitatory hysteresis current, as well as the stripe-wise organization that is paralleled by stripe-wise organization in striatal layers. Striatal layers are unique due to

their dopamine-driven reinforcement learning via PVLV (see Appendix). The posterior layer is distinct because it contains none of the unique features above, but only the more general mechanisms implemented by Leabra and thought to apply to neocortex in general. Finally, the connectivity among these layers is based on known neurobiology (Hazy, Frank & O'Reilly, 2006, 2007, 2010; O'Reilly & Frank, 2006).

The subthalamic nucleus was implemented as a single-unit layer whose activation determined the precise level of activation in the output layer that was sufficient for generating a response. To enable a match to the phasic bursting dynamics of the subthalamic nucleus, we utilized the accommodation currents of Leabra. This separately simulated ion channel within Leabra is intended to capture the influence of calcium-activated potassium channel dynamics, which are known to be important in driving the phasic nature of STN bursting (e.g., Gillies & Willshaw, 2005). In the model, increases in the basis variable driving this channel (i.e., calcium) accumulated at a rate of .005, and decreased at a rate of .001; once the basis variable passed a value of .05, the accommodation current was turned on, with a possible total conductance of 1, and not turned off until the basis variable passed a value of .001. These parameter settings have the effect of allowing for a rapid short burst of firing within the STN, which then strongly and persistently accommodates.

The inputs and outputs presented to the network are depicted in Figure 9. The network includes an input layer with units corresponding to the 2AFC stimuli as well as to the stop signal (and to other inputs, corresponding to tasks not presented in the current report). The network also includes a distinct NE input layer, consisting of a single unit, which is activated when the Stop Signal is present in the environment. This layer projects specifically to the prefrontal layers of the model. This single-unit input layer effectively duplicates the sensory representation of the

Signal that exists in the input layer but, as mentioned above, acts as an additional neuromodulatory signal to the prefrontal cortex that may correspond to the pronounced NE modulation of rVLPFC during response inhibition (Chamberlain et al., 2009). Finally, the output layer of the network consists of units corresponding to the 2AFC responses (as well as to alternative outputs corresponding to tasks not presented in the current report).

Training & Testing.

All models were run in batches of 25 networks, and each network was initialized with random patterns of connection weights. Training proceeded in epochs of 500 events, 30% of which were Signal trials, and the remainder of which required only a 2AFC, until networks had performed at between 30 and 80% accuracy on the Signal trials for four consecutive epochs or 9 epochs total, whichever occurred first. The last epoch was used for analysis, and any experimental manipulations of the network were made only on that final epoch (e.g., NE layer projection strength, Signal trial frequency, or lesions to the prefrontal layer).

On Signal trials, networks were considered to have responded correctly if no unit of the output layer was activated above the response threshold, which was determined in an online fashion as $[.25 + STN_{activation}]$. Similarly, on No Signal trials, networks were considered to have responded correctly if only a single output unit corresponding to the 2AFC stimulus was activated above the current threshold value, calculated in the same way as on Signal trials. The stop signal delay was adjusted using an adaptive staircase algorithm such that if the network failed to respond correctly, the stop signal delay was shortened by a single processing cycle; if the network responded correctly, the stop signal delay was lengthened by a single processing cycle.

Individual differences were simulated as arising from a combination of sources: not only

those effects of initial weight randomization, idiosyncracies in the precise training events presented to each network, and any influences of noise that persisted throughout training, but also as the effects of differences between networks in the activation gain of units in the prefrontal layers. For consistency these gain values were kept at their default levels for the PBWM models applied to working memory updating and task-switching paradigms. In these models, which will be the subject of a future report, differences in gain appear to underlie individual differences in human subjects performing executive function tasks (e.g., Chatham et al., 2011b). Here such individual differences were of interest primarily to verify that SSRT was uncorrelated with No Signal reaction times given a plausible set of underlying individual differences. This prediction must be met for SSRT to be reliably extracted (Logan & Cowan, 1984; Band et al., 2003).

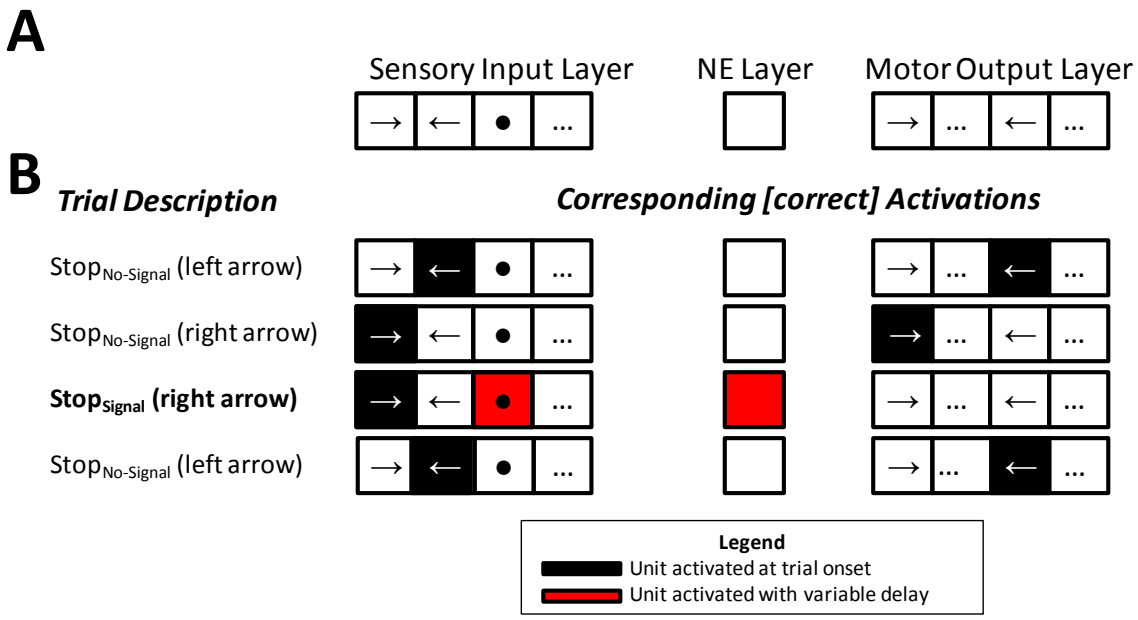


Figure 9. Inputs to the network and corresponding outputs (A) for an example trial sequence (B). A. Inputs to the model include a sensory input layer with units corresponding to the 2AFC stimuli (left and right arrows, above), the Stop Signal (circle, above), and other units not utilized in the current task, as well as a single-unit layer corresponding to norepinephrine-based neuromodulation (which is activated with the signal). The network must learn to produce the appropriate outputs given a particular set of inputs, where successful inhibition corresponds to no above-threshold activation in the output layer. B. Trials proceeded serially, such that the model was presented with 70% No Signal trials that consisted merely of one of the 2AFC units being activated. The remaining 30% of trials consisted of Signal trials, in which the stop signal was activated (and the 2AFC stimulus extinguished) after a variable delay (red/shaded above). This variable delay was determined according to the model's ongoing performance in the task utilizing an adaptive staircase algorithm (see Methods).

Results

Our results indicate that the relatively domain-general mechanisms of PBWM cooperate through learning to give rise to a number of hallmark phenomena from the Stop task. Furthermore, they do so in a way that sheds light on both neuroimaging results and pharmacological manipulations, as well as reliance of the Stop task on a computational "division of labor" across multiple neural substrates.

First we examined the ability of the model to reproduce the most widely-reported

behavioral findings from the Stop task. Foremost among these is the fact that response inhibition can be made more successful by providing the stop signal more quickly after the onset of the 2AFC stimulus; conversely, it can be made less successful by providing the stop signal relatively later after the 2AFC stimulus onset. This pattern is empirically confirmed by utilizing an adaptive algorithm based on subject performance, whereby the stop signal is presented earlier or later depending on the subject's performance from the last signal trial, and testing whether this procedure leads to approximately 50% accuracy across subjects. This widely-observed pattern was reproduced by our model, such that iterative adjustment of the delay between the 2AFC stimulus and the Stop signal yielded approximately 50% accuracy on Signal trials (Signal trial mean accuracy = .50, one-sample t-test against 50%: $t(99)=.221$, $p>.82$), while performance on No Signal trials was at ceiling (Figure 10A; No Signal trial mean accuracy = .96, t-test against Signal trial accuracy: $t(99)=107.35$, $p<.001$). This effect occurs in our model because there is a certain minimum number of processing cycles which must elapse until the presentation of the stop signal can influence the processing taking place in the posterior cortical layer, where the simple sensory-motor mappings involved on No Signal trials are executed. Earlier presentation of the stop signal simply leads to an earlier initiation of such processes.

In the model and in humans alike, this phenomenon can be understood as a race between those processes that support stopping and those that support going. A test of this canonical interpretation of the Stop task is that trials in which inhibition failed should be precisely those for which a reaction was emitted too quickly for said inhibition to occur. In other words, reaction times (RTs) on Signal trials in which a response was erroneously committed should be quicker than the mean RT on No Signal trials. We were also able to verify this phenomenon in our model (Figure 10B; $t(99)=11.23$, $p<.001$), confirming again that a race-like process unfolded in

the model.

For the primary dependent measure of the Stop task – the Stop Signal Reaction Time (SSRT) – to be reliably extracted, there must be "stochastic independence" (Logan & Cowan, 1984) such that the duration of stopping process are not correlated with the duration of going processes. In our model, SSRT and No Signal RTs showed the absence of a correlation that is widely reported in human subjects, and taken as a confirmation of this assumption (R=-.02, p>.5; Figure 10C). This stochastic independence may seem to be a surprising feature of a massively interactive architecture like our model; however, some aspects of the model's mechanisms are more important on Signal trials (e.g., the prefrontal layers) than on No Signal trials (for which the simple sensori-motor mappings in posterior cortex are more crucial). This division of labor can give rise to stochastic independence, as reflected in the near-zero correlation between SSRT and GoRT, even though the underlying processes are in fact interactive.

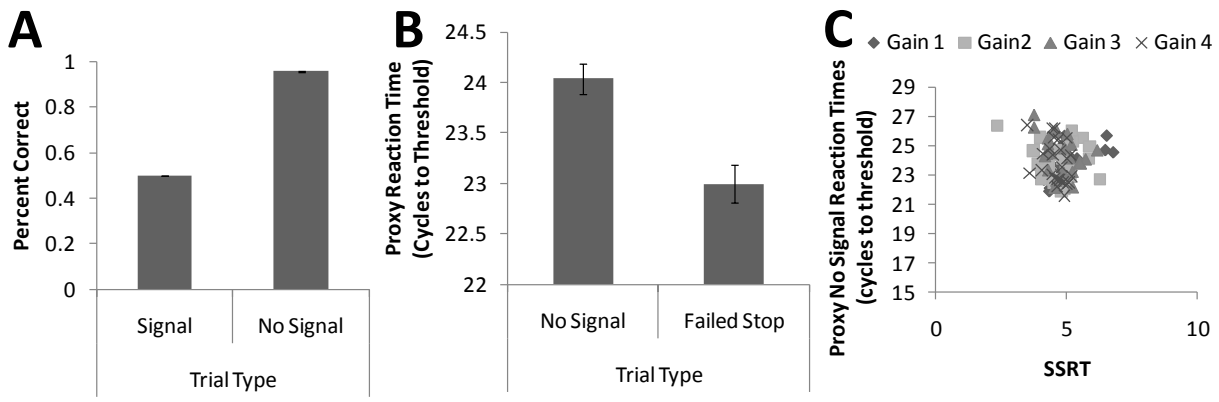


Figure 10. Benchmark phenomena from the Stop task. Based on the principles of the race model of the Stop task, the use of an adaptive algorithm should yield approximately 50% accuracy on Signal trials without affecting performance on the No Signal trial type; this result that was captured by our model (A). The race model also predicts that reaction times on incorrect signal trials – those where the subject failed to stop – should be faster than the average reaction time on No Signal trials; our model also captured this result (B). Finally, the race model predicts that reaction times on No Signal trials and the dependent measure of stopping latency, SSRT, should be uncorrelated; our model also reproduced this result (C; R=-.02, p>.5).

Our model also captured a number of more detailed phenomena reported from the Stop task. For example, subjects are slowed on No Signal trials that follow Signal trials, a finding that was also captured by our model (Figure 11A; $t(99)=11.68$, $p<.001$). In human subjects, this slowing is exacerbated following unsuccessfully stopped trials; this additional nuance was also present in our model (Figure 11B; $t(99)=2.92$, $p<.01$). Stop signals prolong subsequent reaction times in part because Stop Signals strongly activate the prefrontal layers of the model, which in turn activate the subthalamic nucleus (and thus increase response thresholds). Such activations will generally continue into the following trial. Because subthalamic activation tends to occur relatively later in unsuccessfully stopped trials (as discussed below), it will also more strongly persist into subsequent trials – thereby lead to an exacerbated slowing effect following those trials.

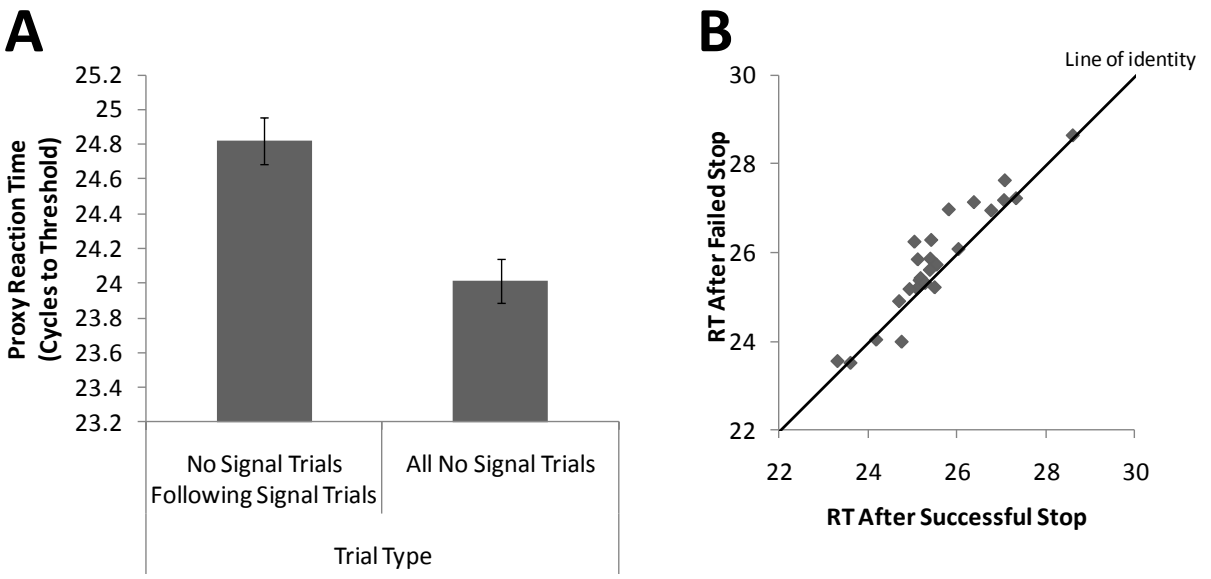


Figure 11. Reaction time slowing due to the presence of Signal trials (A), and in particular unsuccessfully-inhibited Signal trials (B), were also captured by the model. The model was slowed to produce a response on No Signal trials that followed Signal trials (A), as observed in humans. This pattern was also exacerbated on No Signal trials that followed failed Stop trials, relative to those that followed successful Stop trials (B), as illustrated by the fact that most points (each corresponding to a model) lie above the line of identity.

Another widely-reported finding from the Stop task is that SSRT is unaffected by the frequency of Stop Signals (e.g., Logan & Cowan, 1984). Our model once again reproduced that result (Figure 12; $F(1,14)=.33$, $p>.57$), while also giving rise to the additional common finding that No Signal trial RTs are lengthened when Stop signals were presented more frequently ($F(2,14)=74.14$, $p<.001$). In contrast to the null effect of Stop Signal frequency on SSRT, the strong effect of stop signal frequency on No Signal reaction times in the model reflects the increasing contribution of post-signal (and post-error) slowing to this observed mean when stop signals are more frequent.

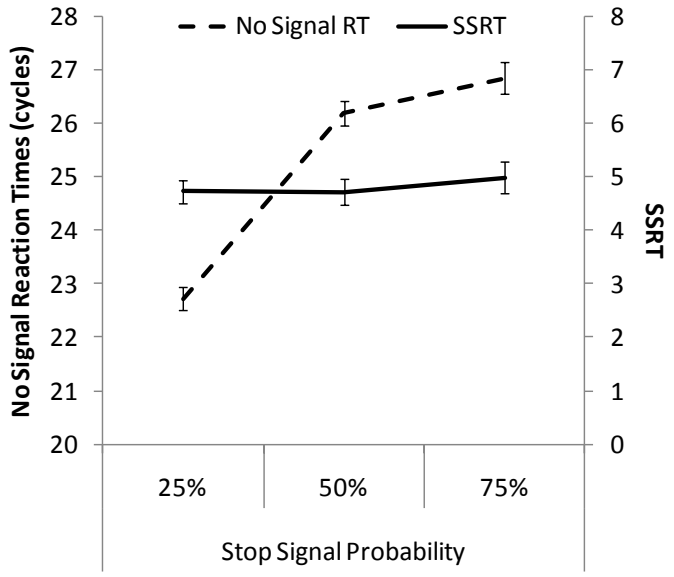


Figure 12. Increasing the frequency of stop signals led to an increase in No Signal trial RTs, but left SSRT largely unaffected, consistent with empirical observation. In the model, No Signal reaction times are lengthened when stop signals occur with greater probability, in part because of the increasing contribution of post-error and post-signal slowing to these observed means (e.g., Figure 11).

Our model also reproduced a number of widely-reported neuroscientific findings from the Stop task. For example, neuroimaging experiments have repeatedly demonstrated that the rVLPFC is more strongly recruited on Signal than No Signal trials. Our model reproduced that result, insofar as activation within the units allowed to surpass threshold was higher on Signal than No Signal trials on average (Figure 13A; $t(99)=13.86$, $p<.001$). The model gives rise to this pattern because error-driven learning mechanisms support a representation of the "Signal" stimulus within the prefrontal layers on Signal trials, and this representation is then maintained across trials due to the excitatory currents within prefrontal layers. These representations are further strengthened in the presence of congruent bottom-up input (i.e., when the Signal is actually present in the environment).

Activation in these prefrontal layers was negatively correlated with SSRT (Figure 13B;

R=-.34, p=.001), as has been observed with fMRI for the rVLPFC. In the model, this correlation arises because prefrontal units support efficient processing of the Signal, by virtue of maintaining that stimulus's identity, and thereby provide a top-down biasing signal to the rest of the model for streamlining processing of the signal. The importance of this prefrontal active maintenance mechanism is further demonstrated by the increased SSRT observed when the prefrontal layers of the model were lesioned after learning (Figure 13C; $t(99)=12.60$, $p<.001$), which accords with the finding that rVLPFC damage (Aron et al., 2003) or transient disruption (Chambers et al., 2006) increases SSRT.

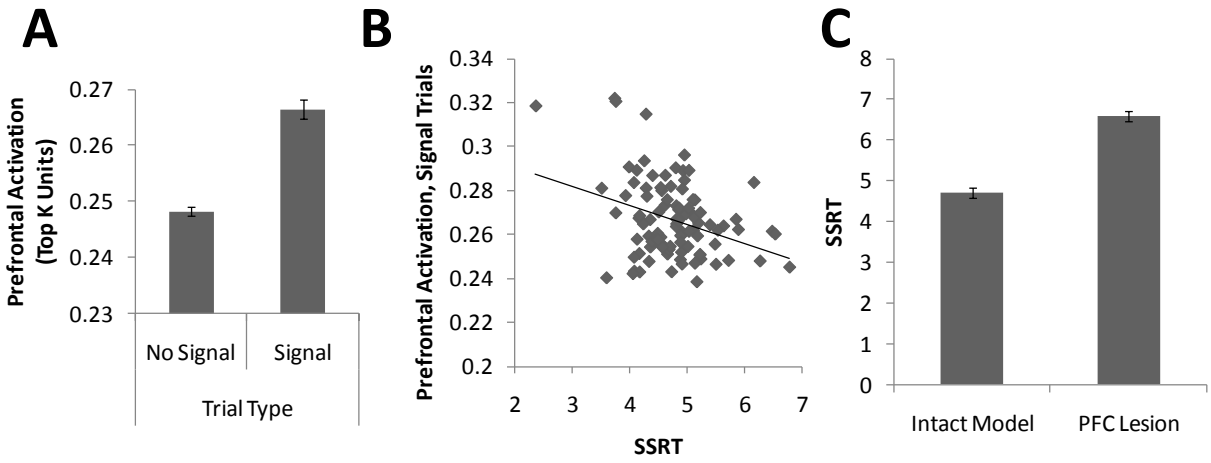


Figure 13. The importance of prefrontal layers in the model. A. Consistent with extant neuroimaging of the Stop task, activation in the prefrontal layers was significantly increased on Signal trials, relative to No Signal trials. (B) This activation was predictive of performance, such that greater prefrontal responses on Signal trials led to more effective stopping. (C). These activation patterns are not merely correlated with SSRT; they are an important causal factor: when prefrontal layers were lesioned, SSRT was significantly increased. This effect matches evidence from both TMS and acquired neurological insult to rVLPFC (Aron et al ., 2003; Chambers et al., 2006).

The importance of the prefrontal cortex for Stop task performance arises in part due to its

modulation by NE. Support for this hypothesis comes from the finding that NE reuptake inhibitors shorten SSRT (Chamberlain et al., 2009). Indeed, after increasing the relative weight of the projections from the NE layer to the prefrontal layers as a proxy for such NE reuptake inhibition, SSRT was also reduced in our model (Figure 14; $t(99)=2.69$, $p<.001$). The additional excitatory input conveyed by this layer supports enhanced prefrontal processing of the Stop Signal when it appears, and thereby supports reductions in SSRT when potentiated.

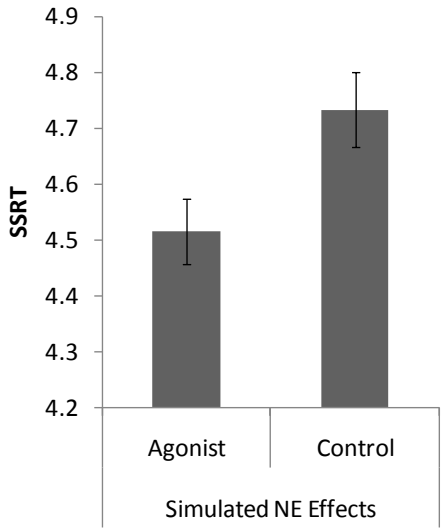


Figure 14. Potentiation of the simulated NE layer in our model leads to a reduction in SSRT, similar to that observed in humans undergoing NE reuptake inhibition.

A number of phenomena related to primary motor cortex have also been reported in the Stop task literature. For example, the excitability of motor cortex – as determined through single-pulse TMS – differs as a function of both trial type and performance: reductions in motor cortex excitability on Signal trials, relative to corresponding No Signal trials, are present only when TMS is provided relatively late (van den Wildenberg et al., 2009). In addition, these reductions are more pronounced on successfully inhibited trials than unsuccessfully inhibited trials.

We observed conceptually similar phenomena in the activation patterns occurring in the model's output layer. Specifically, the accumulation of excitatory activation in this output layer on Signal trials proceeded in a manner that was similar to that occurring on No Signal trials, until a relatively late stage. At that point, reductions in excitatory activation became more pronounced on correct Signal trials, relative to incorrect Signal trials (Figure 15A). In addition, a broad difference between correct and incorrect Signal trials was observed in terms of the timing of excitatory activation, reflecting the fact that incorrect Signal trials are precisely those where motor outputs race to threshold more quickly.

The subthalamic nucleus plays a determining role in whether these activation dynamics lead to a registered response: Activation in the subthalamic layer of our model determines what level of activation in the output layer is sufficient for actually executing a response. In the model, subthalamic layer activation proceeds very similarly on No Signal trials and on incorrect Signal trials, but is considerably heightened on correct Signal trials (Figure 15B). A compensatory increase in subthalamic activation is observed on incorrect Signal trials, but occurs too late to successfully raise the threshold on response execution.

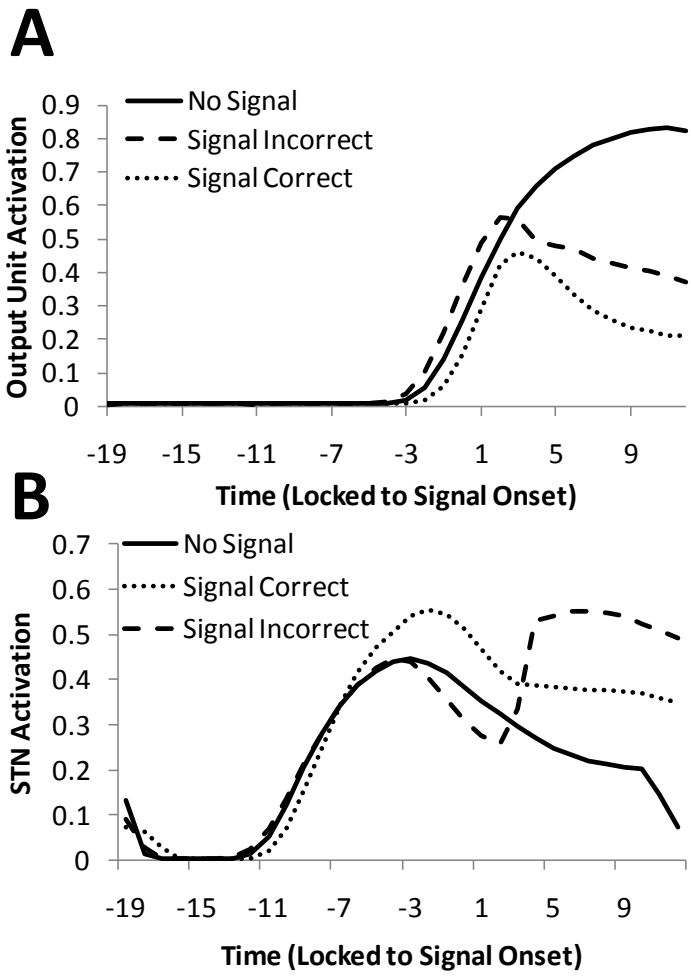


Figure 15. Activation dynamics of the output layers (A) and of the subthalamic nucleus layer (B) during various trials of the Stop task. A. The activation of the output layer proceeds more quickly on incorrect Signal trials than on No Signal trials, consistent with the idea that unsuccessfully inhibited trials are those where responses were emitted relatively quickly. Correct signal trials, by contrast, tend to show a slightly later accumulation of activation in the output layer; this delay provides an additional opportunity for these responses to be cancelled. B. The subthalamic nucleus layer is activated on all trials, but particularly on successfully inhibited trials. In contrast, unsuccessfully inhibited trials show a more delayed response in the subthalamic layer.

Having verified that our model gives rise to a number of hallmark phenomena from the Stop task, including those from investigations of behavior, hemodynamics, neuropsychological insult, pharmacological manipulations, and corticomotor excitability, we next assessed the precise computational role of the prefrontal cortex in the Stop task. It has often been assumed

that the role of the prefrontal cortex in this task is to engage motoric stopping processes, but we found no evidence for this in the model. Specifically, those units in the prefrontal layer that were most active before the signal onset, and before the average signal onset time on No Signal trials, were also those that showed a large increase following signal onset on Signal trials (Figure 16A; $t(56)=6.70$, $p<.001$). Because motoric stopping is only required in the latter case, this phenomenon would seem to suggest that the role of the prefrontal cortex is more general than simply initiating a motoric stopping process. Had that been the case, a distinct set of prefrontal units should have been recruited when motoric stopping was required (i.e., on Signal trials).

It could nonetheless be argued that these units are those which engage motoric stopping, and that this function is also recruited (albeit to a lesser extent) on the No Signal trials. The use of a motoric stopping process on No Signal trials could then be understood as a proactively-controlled form of response inhibition, akin to "responding with restraint". However, and contrary to this account, the units showing this activation dynamic were no more strongly connected with the subthalamic nucleus than the units that failed to show this activation dynamic (Figure 16B; $t(56)=.55$, $p>.58$). In other words, these units clearly did not develop any functional specialization for motoric stopping, although they did clearly develop a functional specialization for detecting the Stop signal.

This effect substantiates accumulating evidence that the computational role of the prefrontal cortex in the Stop task is more general than simply engaging motoric stopping. Instead, it appears to subserve the more general process underlying detection of behaviorally-relevant stimuli within the environmental context (Chatham et al., 2011a). The emergence of such a function within the model clearly establishes that context monitoring can be understood as a particular instantiation of the domain-general active maintenance mechanisms implemented in

our model. It is not a privileged "stopping-specific" subset of prefrontal units which drive activation of the subthalamic layer; instead, such subthalamic activation is simply driven by more global changes in prefrontal activation overall.

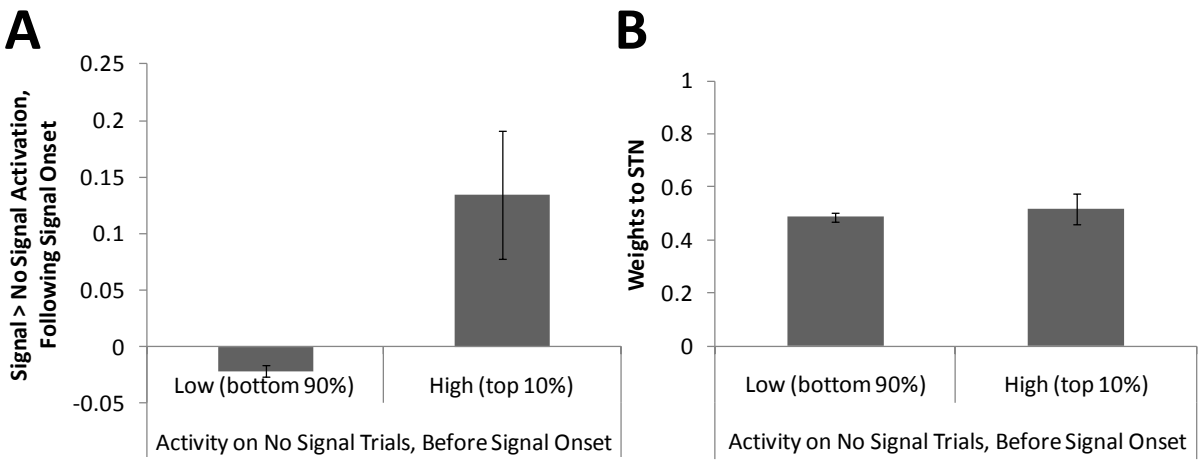


Figure 16. Evidence for a domain-general, and not stopping-specific, function of prefrontal cortex. (A) Units which showed the strongest activity on No Signal trials prior to the average onset time of Signals were precisely those which also showed a further increase in activation when Signals were actually subsequently presented. The recruitment of these units could in principle reflect a proactive engagement of motoric stopping mechanisms on No Signal trials, but in actuality these units were no more strongly interconnected with the subthalamic layer than the remainder of the units (B). Together, this evidence indicates that subthalamic activation occurs as a relatively general side-effect of prefrontal activation, and that while the Stop task crucially relies on a subset of prefrontal units that are specialized for detecting the Stop signal, they are not specialized for engaging motoric stopping.

Discussion

Our results demonstrate that the relatively domain-general mechanisms supported by the PBWM architecture give rise to a number of detailed phenomena from the canonical test of response inhibition, the Stop task. Specifically, our model's behavior matches the predictions of a more abstract race model that accurately characterizes the Stop task, including the convergence of the model's performance to approximately 50% accuracy following the use of an adaptive algorithm for determining Signal onset times, the relatively fast mean RT on failed Signal trials,

and the stochastic independence of stopping and going processes. The model also exhibits both the post-error slowing and the post-signal slowing commonly observed in humans. The effects of signal frequency manipulations match those observed empirically, including the sensitivity of No Signal RT to such manipulations and the specificity of this effect to No Signal RTs in particular (i.e., SSRT was unaffected).

Our model also gives rise to a wide array of phenomena that have been reported in the neuroscientific literature. Specifically, the model demonstrated an increased recruitment of prefrontal layers on Signal trials, similar to the increased hemodynamic recruitment to Signal trials observed empirically. This activation negatively correlated with SSRT, as it does in humans. Lesions to the prefrontal layer of the model led to increases in SSRT that were similar to those yielded by rVLPFC lesions in humans. Conversely, SSRT was decreased by potentiation of the simulated NE inputs to these prefrontal layers, consistent with the effects of NE reuptake inhibition in humans. Patterns of excitability in the output layer of our model were also broadly consistent with patterns of corticomotor excitability in the Stop task probed with single pulse TMS.

In addition to capturing these effects, the model also provides novel support for an emerging perspective on the apparently domain-general components of response inhibition. In particular, we found that the prefrontal units of the model which responded most strongly to demands on motoric stopping were also those which were engaged most strongly on No Signal trials, as though supporting a process involved in monitoring the environment for the behaviorally-relevant "signal" stimuli regardless of whether motoric stopping was required. These units demonstrated no unique specialization for motoric stopping demands: they manifested no differential interconnectivity with the subthalamic layer of our model.

Conceptually similar conclusions have been reached previously, such that the role of the rVLPFC in response inhibition is now interpreted to be substantially more general than what would be expected if rVLPFC were dedicated for motoric stopping (Chatham et al., 2011; Dodds et al., 2011; Hampshire et al., 2010; Sharp et al., 2010; Verbruggen et al., 2011).

The current model goes beyond such work to clarify why the cognitive abilities involved in response inhibition tasks may not be clearly differentiable from those involved in a much wider array of tasks, including those tasks which have little apparent demands on motoric stopping processes. Such tasks all rely on specific instantiations of the active maintenance processes that lie at the core of prefrontal function (Miller & Cohen, 2001) and at the core of the PBWM architecture. Additional quantitative tests of this model and its ability to give rise to individual differences across a battery of executive function tasks (Miyake et al., 2000) are currently underway.

The model also supports a deeper understanding of how such active maintenance mechanisms may be utilized in the Stop task, particularly with respect to rVLPFC function. Recruitment of the rVLPFC is strongly associated with tasks that involve the detection of stimuli that are relevant for behavior in general, regardless of whether they specifically demand motoric stopping. This finding is consistent with the fact that prefrontal interconnectivity with subcortical basal ganglia and subthalamic circuits is a relatively general feature of prefrontal cortex, and not particularly associated with rVLPFC anatomically (Nambu et al., 2002). This widespread pattern of connectivity could reflect the phylogenetic origins of prefrontal cortex for motoric behaviors in general, for which subthalamic and other subcortical circuits are widely regarded as crucial.

The function of the subthalamic nucleus in this circuit is sometimes thought to be cooperative, rather than competitive, with motor initiation processes (Nambu et al., 2002). In essence, the dynamic modulation of response thresholds can be facilitative for producing robust motoric output, such that motor actions commence only when under robust cortical drive. This perspective could explain why dorsal regions of prefrontal cortex – which are in general more strongly associated with motoric processes than rVLPFC (O'Reilly, 2010) – are also more strongly interconnected with subthalamic areas than rVLPFC (Nambu et al., 2002).

On the other hand, the rVLPFC is more strongly interconnected with the anterior temporal regions thought to subserve high-level stimulus identification (Fernandez-Miranda et al., 2008) relative to other prefrontal areas. This connectivity implies that high-level visual information extracted by anterior temporal cortex might then be passed to rVLPFC, which would in turn evaluate this information with respect to its possible relevance for actions, behavior, and goals. For particularly rapid interactions with anterior temporal lobe, this information might partially bypass the "gating" process subserved by subcortical regions. Consistent with these claims, the gating functions subserved by the striatal mechanisms of our model are not used to any large extent in the current task. Models of the homologous region on the left hemisphere have captured a similarly wide array of data without the use of such mechanisms (Snyder et al., 2010). The rather secondary role of such subcortical mechanisms is also consistent with the notable absence of effects from subcortical dopaminergic manipulations or individual differences in subcortical dopaminergic function on Stop task performance (Bari et al., 2011). Finally, gating deficits as would be caused by dopaminergic dysfunction do not seem to be identifiable with pure ADHD (Castellanos, Fine, Kaysen, Marsh, Rapoport & Hallet, 1996), despite that disorder's association with impaired response inhibition (e.g., Barkley et al., 1994). Conversely,

psychopathologies that do show gating deficits can be unimpaired on response inhibition tasks (e.g., Kriete PhD Dissertation). Thus, these "natural experiments" as well as those in conducted in the laboratory provide further support for the aforementioned computational arguments that gating effects are relatively unimportant to response inhibition phenomena.

Rather, the diffuse connectivity of STN with prefrontal layers, and the absence of preferential recruitment of prefrontal units that were strongly interconnected with STN, could suggest that subthalamic activation is a very general or even obligatory consequence of prefrontal recruitment. Taken to its extreme, this position holds that if any experience activates prefrontal cortex it may also activate some corresponding area of the subthalamic nucleus. Although speculative, this hypothesis offers a clear explanation of the commonly observed psychomotor slowing that occurs under intense working memory load (e.g., Barch et al., 1997), a phenomenon which has not been previously explained at the implementational level. This hypothesis also leads naturally to several falsifiable predictions.

First, to the extent any task strongly recruits areas of prefrontal cortex at the same level along the rostro-caudal hierarchy as the rVLPFC, it may also produce motoric slowing by way of activating the subthalamic nucleus. Careful non-parametric analyses of reaction time distributions (e.g., Chatham et al., 2011a) will be required to test this hypothesis.

Second, task-evoked recruitment of more rostral prefrontal areas may in turn activate correspondingly more rostral areas of the subthalamic nucleus, and thereby yield more abstract or cognitive (i.e., less motoric) "slowing" of rostral cortico-thalamic loops. This kind of obligatory subthalamic recruitment following prefrontal activation could relate to disparate domains, such as the cognitive slowing observed during the psychological refractory period, or the attentional blink. Both these paradigms relate to rVLPFC recruitment (Marois et al., 2006;

Verbruggen et al., 2011), but our model further predicts these paradigms should be related to performance on the Stop task.

Third, obligatory subthalamic activation following prefrontal recruitment could have notable computational advantages. Activation of prefrontal cortex is widely associated with difficulty and conflict (e.g., Barch et al., 1997). The activation of prefrontal cortex could thus be a very ecologically-valid signal that deliberation, rather than action, is a more advantageous behavior in the current context (e.g., Frank et al., 2007). If prefrontal areas tend to innervate more caudal subcortical areas, as supported by emerging neuroimaging data and computational investigations (e.g., Badre & Frank, in press; Frank & Badre, in press) as well as established neuroanatomical data (e.g., Alexander, 2007), this kind of cascading architecture may extend to subthalamic innervation as well. In this case, there would be a clear computational benefit to pausing the downstream output of more caudal prefrontal regions: their representations might be first be relevantly informed by the representations of more rostral prefrontal areas.

Verification of these predictions could provide novel support for the current neurocomputational account of response inhibition, but there are also more direct tests of the model. Nearly 25 years of research has accumulated on the Stop task, and there are numerous phenomena which could fall within the scope of an extended model. These systematic extensions could further establish the viability and generality of our account.

For example, a number of detailed phenomena have been reported from Stop tasks with more highly structured demands on motor processes, including 4AFC tasks (Bissett & Logan, 2011) and bimanual dual-2AFC tasks (e.g., Aron & Verbruggen, 2008). The current model is capable of performing both such tasks, but their hierarchical motoric demands may require a more biologically-realistic motor circuit. Specifically, our current output layer is merely a proxy

for the motor strip and an interconnected array of basal ganglia mechanisms. We have abstracted over these substrates here, but investigation of such hierarchical architectures is crucially important, and is the subject of much ongoing work.

At the more cognitive level of analysis, our model also does not suffer from "goal neglect," which we hypothesize to be an important feature driving the hemodynamic differences between successful and unsuccessful Signal trials. Goal neglect has sometimes been simulated as arising from stochastic dopamine fluctuations (e.g., Reynolds et al., 2006). These fluctuations are present in PBWM but are apparently insufficient for generating these differences in the Stop task. This failure of the model is nonetheless informative (Chatham, Yerys & Munakata, under review): it suggests that goal neglect in the Stop task could take the form of stochastic dips in NE rather than fluctuations in dopamine.

While this extension would clearly be sufficient for generating the desired hemodynamic effect in the model, we abstain from implementing these NE dynamics here. Current data on NE function is simply too underconstrained for this to represent a strong biological constraint on the model. For example, a wide variety of current hypotheses regarding the computational role of NE release have been proposed; these hypothesis range from a NE gain effect (Aston-Jones & Cohen, 2005, although cortical dopamine is sometimes hypothesized to do the same thing; e.g., Dursteiweitz & Seamans, 2008) to its near opposite – supporting a broadening, rather than sharpening, of representational attractors (Bowman, Wyble, Chennu & Craston, 2008). A systematic extension of the model to better capture these possible functions of NE thus awaits more hypothesis-driven empirical work.

Finally, the current model is trained to perform the Stop task alone. Clearly, humans are (generally) capable of performing the Stop task but do not require the extensive task-specific

training that is required by our model. This difference is a widespread issue for numerous models, particularly those that are "built to order" for a particular task. While PBWM was certainly not "built to order" for the Stop task – unlike previous models of the Stop task – the ability of PBWM to perform more ecologically-valid tasks awaits future assessment.

Conclusion

The centrality of controlled motoric stopping to response inhibition in extant models and much theorizing is here significantly circumscribed: we demonstrate that a domain-general architecture succeeds in capturing numerous phenomena from the canonical test of response inhibition, the Stop task. The model matches well-established behavioral, neuroimaging, and psychopharmacological phenomena, and also accords with findings that the primary frontal substrate of response inhibition, the rVLPFC, has a domain-general rather than stopping-specific role. The model provides insight into the possible computations of this area, suggesting that its role in detecting behaviorally-relevant stimuli in the environment may be one instantiation of active maintenance mechanisms. Our work provides a quantitative existence proof that such mechanisms could underlie a number of findings, including the commonalities across tasks that do not appear to demand motoric stopping.

CHAPTER 4: STRATEGIC STOPPING AND FOREKNOWLEGDE IN RESPONSE INHIBITION

Abstract

Recent evidence suggests that subjects can exert strategic control over the precise subcortical pathways that support response inhibition, utilizing a slower but more selective form of motoric stopping when it is advantageous for the task at hand. Such accounts contrast with arguments that motoric stopping is a uniformly global and automatic outcome of more general attentional mechanisms. Utilizing a large sample of children and a control group of healthy adults, we find a double dissociation in the phenomena previously taken as diagnostic of strategic control over stopping (namely, an increase in the latency to stop a response, and a concomitant increase in the selectivity of that stopping). The results circumscribe current claims that subjects exert strategic control over stopping, and may point towards a counterintuitive effect of foreknowledge on the detection of behaviorally-relevant stimuli.

Introduction

The ability to stop or inhibit unwanted responses has long been considered central to higher level cognition (Ferrier, 1885). However, recent work casts doubt on the centrality of motoric stopping processes to the cognitive demands of response inhibition. Specifically, it has been argued that the cognitive control processes engaged by at least one canonical test of

response inhibition, the Stop task, may be better characterized as involving the active maintenance of task goals (Friedman & Miyake, 2004; Friedman et al., 2008) and the ability to monitor the environmental context in support of these goals (Chatham et al,. in preparation). Motoric stopping mechanisms may instead be engaged relatively automatically, and fall outside the purview of cognitive control (Chatham, Claus, Kim, Curran, Banich & Munakata, 2011).

This account contrasts with claims that subjects can strategically engage different kinds of motoric stopping depending on the demands of the task (Aron & Verbruggen, 2008). In particular, when subjects must selectively stop only one of multiple concurrent actions, they are thought to do so by relying on the "indirect" pathway of the basal ganglia and its more focal inhibitory effects on thalamic output, rather than the more global "hyperdirect" subthalamic pathway that stops the execution of all actions. Two central pieces of evidence are used to support this claim. First, when subjects are provided with foreknowledge about which one of multiple responses may need to be stopped, this foreknowledge reduces the interference exerted by stopping one action on the performance of concurrent actions – consistent with the use of the slower but more selective "indirect" pathway. Second, the latency to stop a response (as estimated by Stop Signal Reaction Time, or SSRT; Logan & Cowan, 1984; Band et al., 2003) is prolonged given such foreknowledge, again consistent with the use of the slower and more selective "indirect" pathway.

While neuroanatomically plausible, the argument that subjects can cognitively control their use of these motoric stopping pathways seems to contradict previous claims that motoric stopping is a relatively uncontrolled phenomenon that follows automatically from the detection of behaviorally-relevant items in the environment (Chatham, Claus, Kim, Curran, Banich & Munakata, 2011). On the other hand, if these two effects arise from alternative mechanisms,

these data may be fully compatible with the idea that subjects have little cognitive control over these motoric stopping mechanisms. In fact, there are several reasons to believe that these two particular effects have rather little basis in cognitive control.

First, the reduction in interference caused by foreknowledge is extraordinarily small (only ~20ms; Aron & Verbruggen, 2008; Claffey et al., 2010). Moreover, concurrent actions still undergo substantial interference despite this foreknowledge (around 100ms; Aron & Verbruggen, 2008, Claffey et al., 2010). This large "leftover" effect is surprising because foreknowledge should allow one to commit the non-cued response with 100% confidence: it is never the case that the non-cued response will have to be stopped. Apparently, responses with one hand simply cannot be sufficiently insulated from the influence of stopping the other hand, even with perfect foreknowledge. If one interprets such small benefits, and such large leftovers, to be the outcome of strategic cognitive control processes, then cognitive control must also be regarded as particularly ineffective in this domain.

More persuasive physiological effects of selective stopping have occasionally been reported, but some of these effects have also been observed in tasks that do not require stopping – once again calling into question whether they reflect a control-induced dissociation of two subcortical pathways. For example, single-pulse transcranial magnetic stimulation has been used to demonstrate a specific decrease in the motor excitability of a hand that may soon need to be stopped (Cai, Oldenkamp & Aron, 2011). This effect has been interpreted as reflecting selective stopping, but a similar decrease in motor excitability is also observed following cues indicating that a hand may soon be involved in committing a response (Duque & Ivry, 2009). This finding thus raises doubts about the specificity of reduced motor excitability to strategic and controlled demands on selective stopping.

If, as argued above, these effects do not reflect strategic cognitive control over which pathway is used in support of motoric stopping, then what might then drive the effects of foreknowledge on behavioral measures of stopping? We propose that by virtue of providing behaviorally-relevant foreknowledge, cues in the selective stop task make more predictable the nature of the subsequent stop signal's relevance to behavior. The appearance of the stop signal then more weakly drives the prefrontal mechanisms that monitor for behaviorally-relevant stimuli, notably the right ventrolateral prefrontal cortex (rVLPFC; Chatham et al., 2011). Because the recruitment of these prefrontal mechanisms may be ensued by an automatic and global form of motoric stopping (Chatham et al., 2011), foreknowledge would tend to increase SSRT to the extent it weakens this prefrontal recruitment. Interference will also be thereby reduced: more responses will have already been emitted, unabated, before this global stopping takes effect.

This alternative hypothesis gathers some preliminary support from the extant literature on the effects of stimulus predictability. Relative to unpredictable rare stimuli, predictable rare stimuli lead to less pronounced behavioral slowing (Sussman, Winkler, Schroeger, 2003). If the mechanisms supporting such behavioral slowing overlap with those that support motoric stopping, and if stimuli are made more predictable in the selective stop task by providing foreknowledge regarding their impending behavioral-relevance, these patterns could explain both the slightly increased SSRT and slightly reduced interference that result from such foreknowledge.

Second, predictable rare stimuli lead to a reduced frontal event-related potential (ERP; the p3a) relative to unpredictable rare stimuli (Sussman, Winkler, Schroeger, 2003; Sandman, Donnelly, O'Halloran, Isenhart, 1990). The Stop task is also associated with a frontal p3a,

although it is nonspecific to demands on motoric stopping (Chatham et al., 2011); instead, the p3a may reflect the recruitment of prefrontal mechanisms that monitor for behaviorally-relevant stimuli. Changes in this p3 component consistently covary with changes in SSRT across numerous pharmacological manipulations and psychopathological disorders (Chatham et al., in preparation). Thus, because predictable rare stimuli reduce the frontal p3a, manipulations that make stop signals more predictable (e.g., foreknowledge) may also lead to an increase in SSRT.

Interestingly, the effects of manipulating stimulus predictability in healthy adults are significantly different than those observed among the elderly (Sandman, Donnelly, O'Halloran, Isenhart, 1990), among 7-8 year-old children (Wetzel, Widmann & Schroeger, 2009), and among patients with frontal lobe damage (Barcelo & Knight, 2007). In all of these cases, the typical reduction of the p3a to predictable stimuli often shows a trend towards reversal, with a slightly enhanced frontal p3a to predictable targets. This reversal raises a natural test of the hypothesis that foreknowledge makes stop signals more predictable, and that this drives the concomitant changes in SSRT and interference observed with foreknowledge.

If this hypothesis is correct, the effect of foreknowledge on SSRT and interference may show a qualitatively different pattern in young children, relative to healthy adults. To test this hypothesis, we adapted the selective stopping task for 6-year-old children, and administered the same tasks to a group of adult control subjects. We utilized a global stop task as well as an uncued selective stop task (in which subjects must stop only one of two responses, but have no foreknowledge about which response may have to be stopped), and a cued selective stop task (in which subjects are provided foreknowledge about which of two responses may have to be stopped; Aron & Verbruggen, 2008). Short practice tasks were also used to familiarize subjects with the experimental apparatus, including a 2 alternative forced choice task (Choice RT;

administered prior to the global stop task) and a dual simple response task (Double RT; where children had to press with both hands, administered prior to the uncued selective stop task).

Methods.

Participants.

76 6-year-olds (mean age of 72 months; SD=4 months, range 67-81 months) and 11 young adults successfully completed the tasks. Children were recruited from the Cognitive Development Center's recruitment pool at the University of Colorado, Boulder. Adult subjects were recruited from the University of York's undergraduate psychology pool. An additional eight children were excluded from all analyses because their accuracy on Signal trials fell below the 15% minimum required for reliable extraction of SSRT (n=2), because their extracted SSRT was negative for one of the three stop tasks (n=2), or because their accuracy on the No Signal trials of one of the three stopping task fell below 75% (n=4). One additional adult subject was excluded for zero accuracy on Signal trials of the cued Stop task.

All subjects completed the tasks in a set order: Choice RT, Global Stop task, Double RT, Selective Uncued Stop Task, Selective Cued Stop task. (Additional tasks were also administered following these five tasks, but will not be described here). The stimuli used in the first five tasks are illustrated in Figure 17 and described in detail below.

Choice RT.

In this task subjects are introduced to "George the Monkey" and told to press a button with their left hand if the banana appears to George's left, and with their right hand if the banana appears to George's right. The experimenter then demonstrates how and when to respond on two successive trials. The child completes 24 subsequent trials with a randomly-selected interstimulus interval of 1.2-2s. Upon completion of the task, the child is allowed a short break

to pick a small prize (one sticker). The median RT is then calculated, trials exceeding the median RT by a factor of 2.3 or more are excluded, and the median is finally recalculated. This median is used throughout subsequent tasks to ensure that subjects do not attempt to slow their responses in those tasks.

Global Stop.

Subjects are told that the bananas will now sometimes turn brown, and that because George does not like brown bananas, they should try to refrain from pressing the button on those trials. The experimenter demonstrates the task on 4 trials, 2 of which constitute signal trials, one with a signal delay of 50ms and the other with a delay of 600ms. Subjects are then given 24 practice trials. Signals are presented with 33% frequency, and with an initial signal delay of 600ms. This delay is adjusted according to an adaptive staircase algorithm, such that signal onset delay is lengthened by 50ms following a successful stop, and shortened by 50ms following unsuccessful stops. The child then completes 3 blocks of 48 trials each, with short breaks to select a prize in between each block. If on any trial the reaction time is 2.3 times the median RT calculated above, a "ding" sound is played by the computer and the experimenter encourages the child to press the buttons more quickly.

Double RT.

Subjects are told that now two bananas will appear at once, and so they should press simultaneously with both their left and right hands so that George can get as many bananas as possible. They complete 32 trials of this task. If the reaction times of the left and right hands differ by 200ms or more, a recorded voice says "Press the buttons at the same time." The purpose of this task is to give children some practice with responding with both hands, and to ensure that they understand the requirement to press both buttons relatively simultaneously.

Uncued Selective Stop.

Subjects are told that now two bananas will appear, but one of the two bananas may turn brown. They are instructed to press the buttons corresponding to those bananas that do not turn brown. The experimenter completes 4 demonstration trials (2 of which are signal trials); next, the subject completed 24 practice trials, followed by 3 blocks of 48 test trials. If at any time subjects do not press the buttons within 200ms of each other, a recorded voice says "Press the buttons at the same time." Similarly, if any reaction time is larger than 2.3 times the median calculated above, the computer makes a "ding" sound and the experimenter encourages the subject to press the buttons faster. Signal onset delays begin with the asymptotic value reached in the Global stop task and continue to be adjusted in this task, according to the same adaptive algorithm.

Cued Selective Stop.

Subjects are told George knows which one of the two bananas might turn brown, and that he will point to the banana that might turn brown. Otherwise, the task is identical to the uncued selective stop task: the experimenter completes 4 demo trials, and the child completes 24 practice trials followed by 3 blocks of 48 test trials. Signal onset delays are modified via the adaptive algorithm, and reaction times are monitored for inter-response intervals greater than 200ms and reaction times larger than 2.3 times the median.

Trimming and Preprocessing.

Incorrect No Signal trials were excluded from analysis. All correct No Signal trial reaction times were then subjected to an iterative trimming procedure, conducted separated for each task. Specifically, any reaction times falling beyond 2.5 standard deviations away from a child's mean No Signal reaction time for that task were omitted from analysis; the mean and standard deviation for that task was then recalculated. This procedure was repeated iteratively 100 times,

or until no additional trials were excluded, whichever came first. Trials were also excluded from the Selective Stop tasks if the reaction times for the two hands differed by more than 70ms (as in Aron & Verbruggen, 2008). SSRT was then calculated as the average SSD minus the nth percentile of No Signal reaction times, where n was the proportion of errors observed ($SSRT_{AV}$ in Band et al., 2003). In the selective stop task, SSRT was calculated separately for the two hands, and then averaged (again, as in Aron & Vebruggen, 2008). Interference was calculated as the difference between the response times on successfully stopped trials and those occurring on No Signal trials. Finally, two tailed independent t-tests without the assumption of equal variances were used for all comparisons of children and adults.

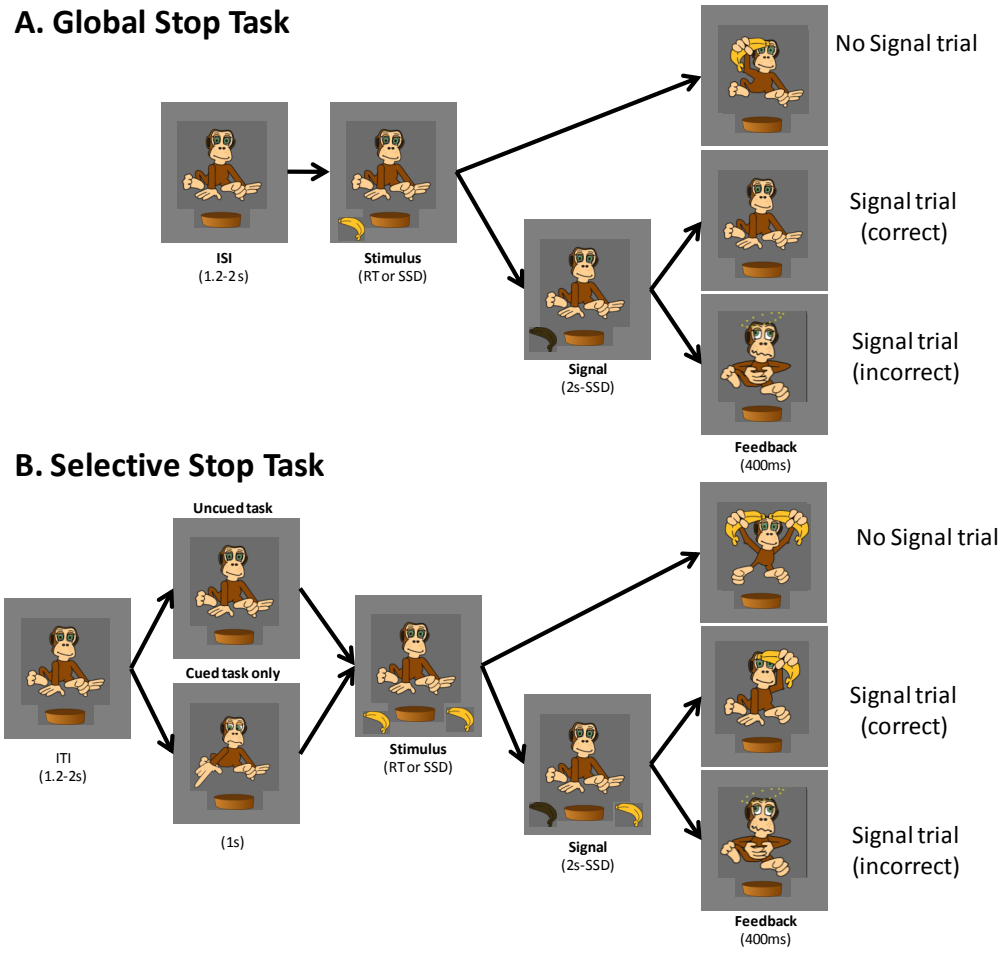


Figure 17. Stimuli used in the Stopping tasks. A. In the Global Stop Task, each trial begins with presentation of George without any bananas; following a variable intertrial interval, a yellow banana appears. On No Signal trials (and the preceding ChoiceRT task) this stimulus remains on screen until subjects press the corresponding button on a button box (with the left hand for bananas on the left and with the right hand for bananas on the right). On Signal trials, the banana turns brown with a variable stop signal delay (SSD) which remains on screen until subjects respond (in which case the monkey appears ill) or until the termination of the trial. B. As in A, trials begin with the presentation of George without any bananas. In the uncued selective stop task (as well as the Double RT task) the next stimulus is of George with two bananas. On No Signal trials (and the Double RT task), this stimulus remains onscreen until subjects press both buttons, at which point George is shown grasping the two bananas. On Signal trials, one of the two bananas turns brown with a variable SSD. If subjects press the button corresponding to the brown banana, George is shown to be ill; otherwise subjects are then shown an image of George grasping the yellow banana. The Cued Selective Stop Task differs only insofar as an additional stimulus is presented, for 1 second, which shows George pointing to the banana that may turn brown.

Results

Both children and adults performed with high accuracy on No Signal trials across the Global, Uncued Selective, and Cued Selective Stop tasks (children: M_{Global}= 93% M_{Uncued}=92% M_{Cued}=92%; adults: M_{Global}= 99% M_{Uncued}=99% M_{Cued}=99%). Furthermore, performance on the Signal trials across tasks was within the 15-85% range that is most sensitive for reliable extraction of SSRT, both for children (M_{Global}= 49% M_{Uncued}=44% M_{Cued}=42%) and for adults (M_{Global}= 50% M_{Uncued}=43% M_{Cued}=47%). Significant differences between children and adults in terms of Signal trial accuracy were achieved only on the Cued Selective task, where children had slightly lower accuracy (unequal variance $t(85)=17$, $p<.05$; Figure 18).

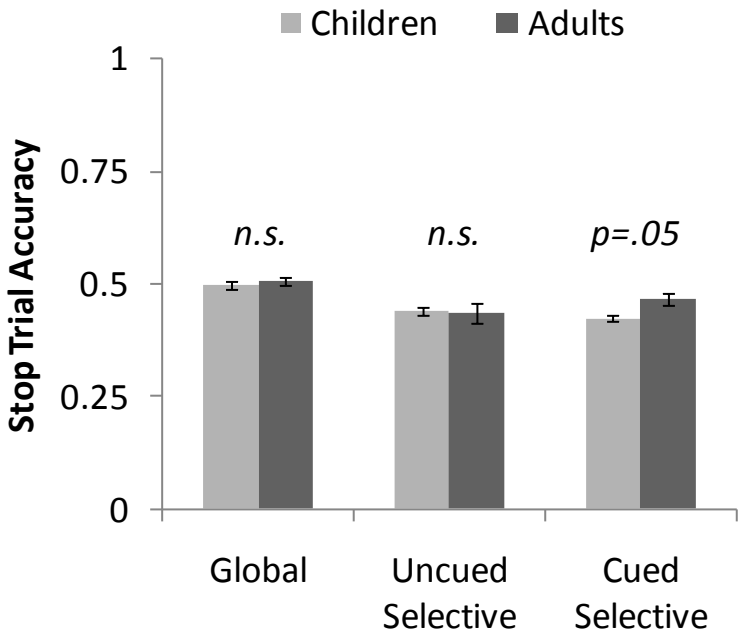


Figure 18. Accuracy on Stop Signal trials across tasks. For both children and adults, stop trial accuracy approximated 50%. A slight decrease in accuracy across tasks was observed; this decrease was significantly different between children and adults only during the Cued Selective stop task. In all tasks and subjects, accuracy on stop trials remained within the 15-85% window that is most sensitive for calculation of SSRT (Band et al., 2003).

The two further assumptions required for reliable extraction of SSRT were also met. First, failed stop trial reaction times were reliably faster than reaction times on No Signal trials across all three tasks, for both children (all t's>14.5, p's<.001; Figure 19A) and adults (all t's>2.7, p's<.02; Figure 19B), consistent with the assumptions of the independent race model that underlies the calculation of SSRT. Second, reaction times on No Signal trials were uncorrelated with SSRT on all three tasks, for children (all p's>.25) and adults alike (all p's >.05), again consistent with the assumptions of the independent race model that underlies the calculation of SSRT.

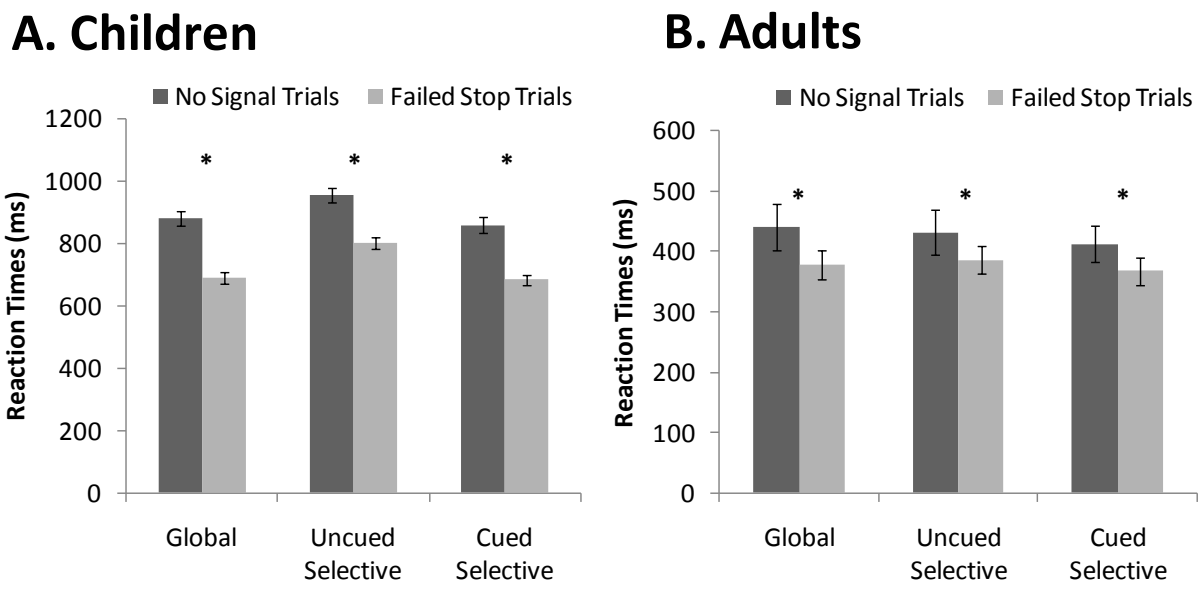


Figure 19. Across all tasks, Failed Stop trials were significantly faster than reaction times on No Signal trials, both for children (A) and for adults (B). One of the assumptions underlying the calculation of SSRT is that "going" and "stopping" processes are engaged in a race. Failed stop trials are presumed to be those in which the "going" process won this race, and thus were emitted too quickly for the "stopping" process to complete. This assumption was confirmed for both children and adults across all three tasks.

The latency to stop a response during the Global stop task, as estimated by SSRT, was significantly longer in children than adults (unequal variance $t(29)=2.9$, $p<.01$; Figure 4). This pattern is similar to that found in previous work, which has established a developmental decrease in the latency to stop a response (e.g., Ridderinkhof, Band & Logan, 1999).

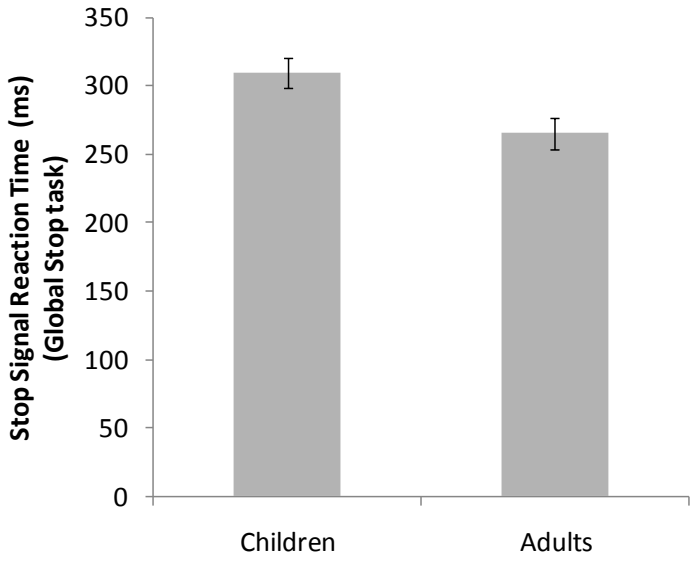


Figure 20. Children showed significantly increased SSRT for the Global stop task. Consistent with prior developmental work, we found that children required significantly more time to successfully stop a response, as estimated by SSRT (e.g., Ridderinkhof, Band & Logan, 1999). This difference is notably more minor than that observed for overt reaction times on No Signal and Failed Stop trials (c.f., Figure 19).

The effects of foreknowledge on interference were significantly different between children and adults. In particular, whereas adults showed a decrease in interference when provided with foreknowledge ($t(10)=3.1$, $p=.01$), children showed a nonsignificant effect in the opposite direction ($t(74)=.84$, $p=.4$). These patterns significantly differed by age (unequal variance $t(31)=2.8$, $p<.01$) (Figure 21A).

The effects of foreknowledge on SSRT were also significantly different between children and adults. Whereas children showed a significant increase in SSRT when provided with foreknowledge ($t(75)=2.6$, $p=.01$), contrary to the patterns previously observed in adults, our adult sample showed no such effect ($t(10)=.245$, $p=.81$). These effects significantly differed across age groups (unequal variance $t(65)=2.1$, $p<.05$ Figure 21B).

Finally, the opposite influences of foreknowledge on interference and SSRT observed in adults was altogether reversed in the younger age group. In particular, the foreknowledge

provided by cues decreased interference among adults (but not children), whereas it decreased SSRT among children (but not adults) ($t(57)=3.0$, $p<.005$; Figure 21C).

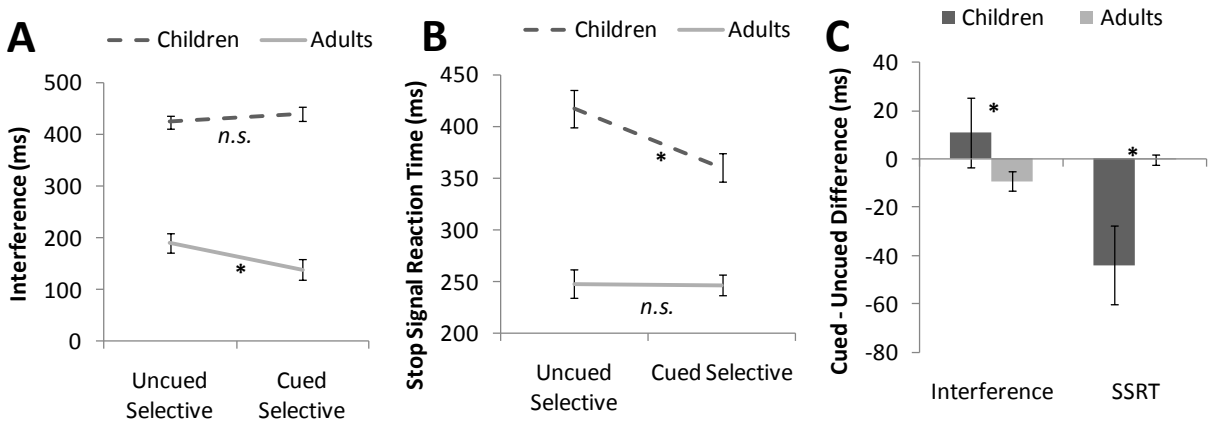


Figure 21. Foreknowledge as provided by cues affected both interference (A) and SSRT (B) in ways that were significantly different across children and adults (C). A. When adults were provided with foreknowledge via cues, they showed a reduction in the interference on concurrent actions that was exerted by stopping an action. In contrast, children demonstrated a nonsignificant effect in the opposite direction. B. When children were provided with foreknowledge via cues, they manifested a significant reduction in the latency to stop a response, as estimated by SSRT; adults showed no such effect. C. These effects of foreknowledge were significantly reversed as a function of age: foreknowledge decreased interference only for adults, and decreased SSRT only for children.

Discussion

It has been argued that the subcortical motor pathways supporting response inhibition can be differentially recruited depending on task demands. In particular, when only one of multiple concurrent actions must be stopped, and when subjects are provided with appropriate foreknowledge, they are thought to be capable of recruiting a slower but more selective pathway for stopping. Here we assessed whether these effects may not be specific to the controlled recruitment of motoric stopping mechanisms by examining whether foreknowledge might elicit the unusual developmental trajectory previously associated with manipulations of target predictability.

Consistent with this prior literature, we observed that foreknowledge had significantly opposing effects on children and adults. For children, foreknowledge significantly reduced the latency to stop a response, and numerically increased interference. In adults, such foreknowledge significantly reduced interference, and had no apparent effect on the latency to stop a response.

This pattern is not naturally accommodated by existing accounts of selective stopping. Such accounts predict that foreknowledge about the upcoming selectivity of response inhibition should encourage a strategic recruitment of the slower but more selective "indirect" pathway. Given the difficulty children have with proactive cue processing (Chatham, Frank & Munakata, 2009), these effects might be predicted to be merely less pronounced in children. Instead, we found significantly different patterns in children than in adults, whereby children showed an indication of a reversal of these effects of foreknowledge.

On the other hand, this reversal is not complete. Behaviorally-relevant foreknowledge failed to significantly increase SSRT in our adult group, contrary to previous findings (Verbruggen & Aron, 2008), although it did significantly decrease the observed interference of stopping on concurrent actions. The former null finding may reflect the small number of subjects in our adult group. Conversely, behaviorally-relevant foreknowledge failed to significantly increase interference in children, although it did significantly decrease the latency to stop a response. We suggest that this null finding may reflect the increased reaction time variability of children; indeed, the mean change in interference experienced by children was similar to that experienced by adults, but with a significantly larger amount of variability around that mean (Figure 21C).

Strictly speaking, the pattern we observed thus reflects only a double dissociation – by which foreknowledge affected only SSRT in children, and only interference in adults. While broadly consistent with the idea that foreknowledge has much more general attentional effects on stimulus predictability (see Introduction), we did not observe the full double crossover interaction predicted by this account – in which foreknowledge would have affected both SSRT and interference in both children and adults, but in opposite directions.

There are alternative interpretations of our results, but they seem even less viable. For example, the information extracted by children from the cues may have been less about the precise behavioral-relevance of a probabilistic stop signal, and rather more about the certain impending onset of a trial. Indeed, in our design (Figure 17A), the stop signal cues act as a reliable indicator of trial onset, and the Cued Selective Stop task was associated with faster reaction times than the Uncued Selective Stop task among children (Figure 19A), as would be expected if cues alerted children to the upcoming task's demands. This interpretation is unlikely to be the whole story, however, given two prominent features of our data. First, the same pattern was not observed among adults (Figure 3B), even though adults benefit from warnings about trial onset in other paradigms (e.g., Valessi, Shallice & Walsh, 2007). Second, there was no evidence that this difference in reaction times significantly influenced the changes in SSRT or interference induced by foreknowledge[1], contrary to the idea that this kind of "alerting" affect might explain the observed results.

$SSRT_{Cued}$, $SSRT_{Uncued}$, and the $SSRT_{Cued-Uncued}$ difference scores were all uncorrelated with the difference in reaction times on the Cued and Uncued blocks (R's <.14, p's>.2). Similarly, interference on the Cued task, the Uncued task, and the difference in interference between the cued and uncued tasks were also uncorrelated with the reaction time differences across these tasks (R's<.12, p's>.27). Controlling for the differences across tasks also revealed a continuing significant effect of foreknowlegde on SSRT among children. In all cases, the extent to which reaction times changed across tasks in children does not seem to be related to the effects we report.

Alternatively, the observed patterns may reflect the significantly reduced accuracy in the Cued Selective Stop task observed among children (Figure 18). However, the observed patterns were highly stable across multiple exclusion criteria; they actually increase in size when subjects with lower performance on the Cued Selective Stop task are excluded. This remains true even when such exclusion yields a significant accuracy advantage for children, relative to adults, on the Cued Selective task. Once again, such differences seem insufficient for explaining the observed patterns.

More broadly, it is possible that some other more general feature of the task design has given rise to the observed effects. For example, our child-friendly adaption of the selective stop paradigm utilizes a dual simple reaction time task (dual SRT; where each hand only had one possible response) as opposed to a dual 2-alternative forced choice (dual 2AFC; where each hand can make one of two possible responses, as in Aron & Verbruggen, 2008). We chose this design in part because children of this age had difficulty even with the dual SRT, particularly within the Cued Selective Stop task (where accuracy on No Signal trials dropped to as low as 57%), and in part because the Global Stop task has been successfully administered as a SRT (Logan & Cowan, 1984). Nonetheless, the use of a dual SRT could have reduced the number of possible stimulus-response mappings in a way that enabled adults (but not children) to more deftly manage response interference, while at the same time obviating any need to use a slower pathway for more selective stopping. Future work will be needed to test this more sophisticated hypothesis.

The central alternative we propose – that foreknowledge may have more general attentional effects while falling short of enabling strategic control over the precise subcortical pathway utilized for stopping – is also further testable. Specifically, it motivates three basic predictions. First, the frontal "Stop p3" component should be reduced when foreknowledge is

provided in adults but enhanced among children, as hinted at by previous developmental and neuropsychological work. Second, the rVLPFC should be more strongly recruited among children when foreknowledge is provided, but less strongly recruited among adults, given its putative role in the detection of behaviorally-relevant stimuli (Chatham et al., 2011). Effects consistent with this claim may have already been observed (Coxon, Stinear & Byblow, 2010). Third, adult subjects should still show some interference on concurrent actions even when no motoric stopping is required; this interference should be expressed relatively later when behaviorally-relevant foreknowledge is provided in advance. More concretely, subjects might perform a dual 2AFC task in which a rare "Go Again" signal indicates that both of the two possible responses should be committed by a particular hand. Foreknowledge would be provided regarding the hand to which this "Go Again" signal would pertain, if it appeared. The duration and time of onset of interference can then be interrogated through nonparametric distributional analyses of reaction time (Chatham et al., 2011).

If this alternative hypothesis is correct, it will provoke the question of why the effects of predictability might be reversed in children. Such questions can only be speculatively raised, much less answered, at this time. Nonetheless, one possibility is that children simply find it more behaviorally significant to have correctly predicted something than to have failed to predict something. Indeed, children may (at least initially) feel surrounded by a panoply of highly unpredictable sights and sounds, making those few that they can successfully predict all the more significant for their future behavior. This kind of bias towards "positive prediction" might encourage the updating and maintenance of environmental information that is predictive of future experiences, and thereby encourage the development of increasingly adult-like proactive control (Chatham, Frank & Munakata, 2009).

Conclusions

*W*e have presented a double dissociation of the effects previously taken to be diagnostic of strategic control over selective stopping. Foreknowledge reduced the latency to stop a response among children, but had no such effect on adults; conversely, foreknowledge reduced the interference exerted by stopping on concurrent actions in adults, but had no such effect in children. These results add to the literature by importantly circumscribing recent claims that such measures can index strategic control over stopping. These accounts cannot explain the present pattern, at least without making significant new assumptions. These results also add to the literature by demonstrating a behavioral pattern induced by foreknowledge that is similar to that observed in the developmental and neuropsychological literatures on target predictability. Finally, the present work adds to the current debate regarding the influence of attention and cognitive control on response inhibition by demarcating a specific set of testable predictions for future work in this rapidly-advancing domain.

CHAPTER 5: SELECTIVE STOPPING AND COMPUTATIONAL CONUNDRUMS

Abstract

It has been argued that subjects can selectively stop only one of multiple concurrent actions by strategically recruiting a distinct pathway for the stopping of motor actions when this is advantageous for the task at hand. However, close inspection of this hypothesis and the supporting data reveal a number of computational conundrums that pose a significant challenge to the plausibility of this account as well as alternative hypotheses. Simply put, it is unclear how such a system could actually work in the brain. After first introducing our mode of inquiry – whereby the failures of computational models can be used to advance theorizing – we then describe a problem posed by existing accounts of selective stopping, and demonstrate the insufficiency of alternative theoretical assumptions for explaining observed empirical patterns.

Introduction

While computational modeling has been widely acknowledged as a useful tool for theory testing (Marcovitch & Zelazo, 2009; McClelland, 2009; Thomas et al 2009; Munakata, Snyder & Chatham, in press; Elman, 2005; Weng et al., 2001; Tenenbaum, 2011), its utility has been most widely acknowledged as confirmatory. That is, models are widely used to confirm that a particular set of theoretical assumptions are sufficient for giving rise to a set of observed data, and thereby motivate more targeted future assessment of those hypotheses.

These demonstrations are sometimes criticized as theoretically uninteresting, on the intuitive basis that models can be made to "do anything." However, computational models are often less flexible than typically acknowledged, and their failures can be highly informative for future theorizing (Chatham, Yerys & Munakata, under review). For example, computational models that successfully simulate phenomena from semantic memory (Rumelhart & Todd, 1993) categorically fail to match phenomena from the domain of episodic memory. Motivated by these failures, a drastic reconceptualization of the computational demands of episodic and semantic memory led to formalized account of the minimal conditions under which these opposing computational demands could be met (McClelland, McNaughton & O'Reilly, 1996); this reformulated account has now become a widely-accepted theoretical framework for understanding the distinction between these memory systems. This example clearly demonstrates that models cannot simply be made to "do anything;" instead, cases in which models fail are informative because they identify the precise boundary conditions of existing hypotheses.

Identifying these boundary conditions is instrumental for motivating the development of alternative hypotheses – without a formal model, it is often not clear that existing hypothesis cannot, at any unambiguous level of detail, explain existing data. This may characterize the current domain of selective response inhibition, in which both dominant and alternative hypotheses – at least, at their current specified level of detail – appear insufficient for explaining phenomena from this domain. This is here demonstrated with a series of computational models. The insufficiency of these models reveals a minimal set of conditions that may be required for human-like selective stopping to occur, and thereby illuminates a promising direction for future research in this domain.

The problem of interference in selective stopping

It is now known that the stopping of unwanted motor actions occurs more slowly when subjects are provided with foreknowledge about which of multiple concurrent actions might need to be stopped (Aron & Verbruggen, 2008; Claffey et al., 2010). This foreknowledge is not entirely without benefit: foreknowledge actually reduces the interference that is exerted by the stopping of one action on the performance of other concurrent actions. These patterns are argued to reflect the strategic recruitment of a slower but more selective pathway for motoric stopping – the "indirect" pathway through the basal ganglia – as opposed to the more rapid and more global "hyperdirect" subthalamic pathway for stopping (Aron & Verbruggen, 2008).

At its currently specified level of detail, this hypothesis cannot explain the fact that motoric stopping appears to be an uncontrolled behavior. That is, motoric stopping is observed even when subjects have perfectly reliable foreknowledge that no response will ever need to be stopped (Chatham et al., 2011). This hypothesis also cannot explain the fact that increased recruitment of the subthalamic nucleus is observed when response force must be increased (Patil, Carmena Nicolelis, & Turner 2004; as opposed to only when response force must be decreased, as must occur in the case of stopping). For these reasons, we have taken an alternative hypothesis as a starting point (in the discussion, we also consider reformulations of the original hypothesis).

This alternative hypothesis states that motoric stopping is engaged globally and rather automatically, and thus falls outside the scope of cognitive control (consistent with ERP, fMRI, pupillometric, and behavioral data from Chatham, Claus, Kim, Curran, Banich & Munakata,

2011). By this account, as formalized in a computational model in Chapter 3, the role of the subthalamic pathway is better understood as modulating a response threshold on motor actions. This threshold must be increased when motor actions must be committed more energetically, thereby explaining why subthalamic activation may increase with increased response force.

By this alternative account, foreknowledge increases the latency of stopping because foreknowledge changes the information content of the stop signals. In effect, foreknowledge reduces the behaviorally-relevant information that is provided by stop signals, because their behavioral relevance is then known in advance. Behavioral-relevance is key for determining the recruitment of an area that drives the subthalamic nucleus in this task (the right ventrolateral prefrontal cortex; rVLPFC). Reductions in behavioral relevance may reduce the recruitment of the rVLPFC and of the interconnected subthalamic nucleus. The resulting behavioral expression of motoric stopping is then delayed, which in turn leads to the appearance of reduced interference merely because a global form of stopping happens later in time – only after a larger proportion of responses have already been emitted.

We assessed this hypothesis within a computational model (Figure 22). However, we find that relatively general learning dynamics prevent this hypothesis from being a viable account on its own. As we describe below, the information content of foreknowledge in this paradigm is most reliably that the alternative not-to-be-stopped response will certainly need to be committed. Similarly, the information content of the stop signals themselves includes the fact that the alternative response can be committed immediately.

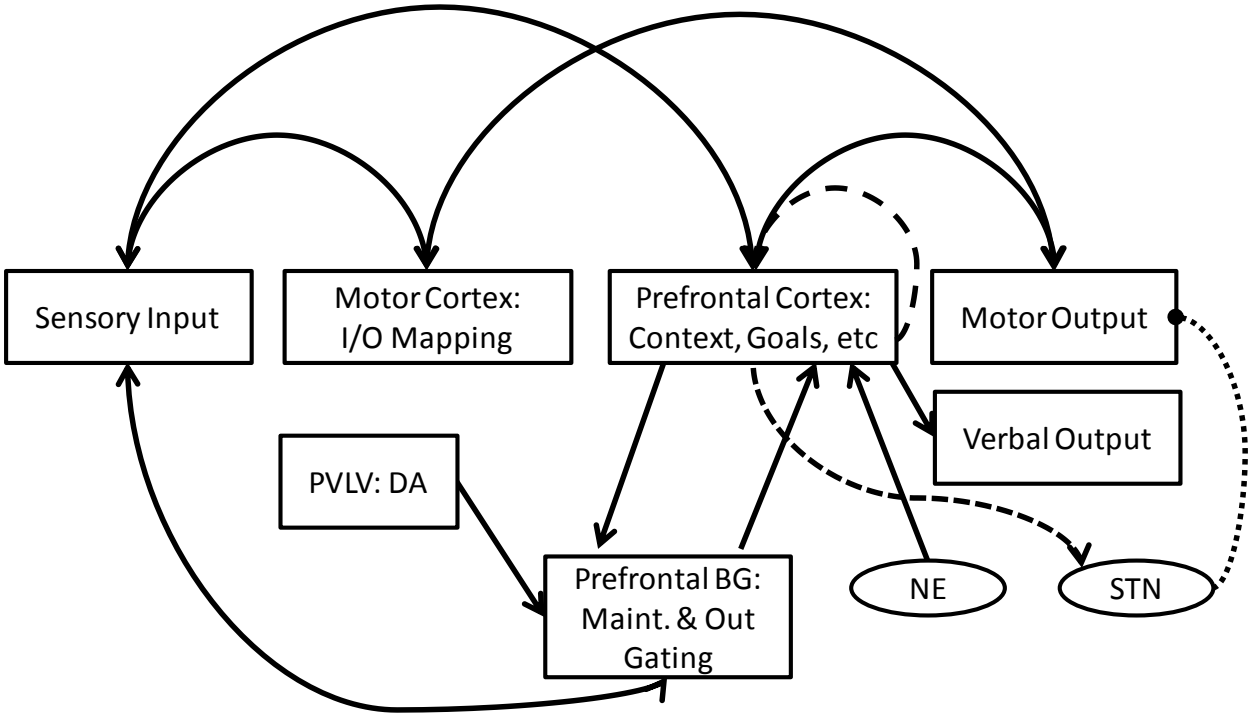


Figure 22. Architecture of the current model. The current model is a systematic extension of PBWM, but largely identical to the model presented in Chapter 3. Specific changes pertain mostly to the precise input units that were used in the current model (see also Figure 23).

The models unavoidably learn to take advantage of this characteristic of the task, and thereby show a pattern of facilitation, rather than interference, when only one of multiple actions must be stopped. In no case do human subjects appear to show this form of learning – which is a challenge for all extant theoretical accounts of selective stopping because they all, directly or indirectly, posit the involvement of the basal-ganglia mediated probabilistic learning mechanisms that should support precisely this kind of learning. Systematic attempts to eliminate the expression of this learning in the models, e.g., by decreasing the latency or increasing the strength of automatic stopping mechanisms, uniformly fail for interesting reasons. We conclude with a discussion of the minimal theoretical revisions necessary to accommodate this unexpected

competency of the models, and the apparently complete failure of human subjects to learn this characteristic of the task.

Methods

Implementation.

The current model is implemented using Leabra (O'Reilly et al., 2001) and the Prefrontal Basal Ganglia Working Memory architecture (Hazy, Frank & O'Reilly, 2006, 2007, 2010; O'Reilly & Frank, 2006). Leabra has been utilized in over 40 models and gives rise to a number of cognitive phenomena (O'Reilly & Munakata 2001). PBWM has likewise been used in dozens of models and represents a unified computational account of working memory and executive functions; it is sufficient for capturing phenomena from canonical tests of task switching (O'Reilly & Frank, 2006) and working memory updating (O'Reilly & Frank, 2006; Chatham, Herd, Brant, Hazy, Miyake, O'Reilly & Friedman, 2011).

The current PBWM model is also capable of capturing behavioral, neuroimaging, and pharmacological phenomena from the canonical test of response inhibition, the Stop task (Chapter 3). Here we utilized various extensions to this model in attempts to match empirical data, to assess its ability to capture phenomena from the selective stop tasks of Aron & Verbruggen (2008).

The inputs and outputs of the model are presented in Figure 23. As in the model presented in Chapter 3, the model contains a sensory input layer corresponding to the various stimuli of the task, a simulated norepinephrine layer for simulating the pronounced adrenergic modulation of the prefrontal areas recruited during the Stop task, and a manual output layer with units corresponding to the various behavioral responses required in the task. There is only one principle differences in these patterns between the models: there are several new sensory inputs

that are necessary for simulating these selective stop tasks. This includes sensory inputs corresponding to effector-specific stop signals (i.e., inputs corresponding to "stop the left response" or "stop the right response") as well as effector specific cues (inputs corresponding to "maybe stop the left response" or "maybe stop the right response"). These additional units were present in the prior model, but simply left unused.

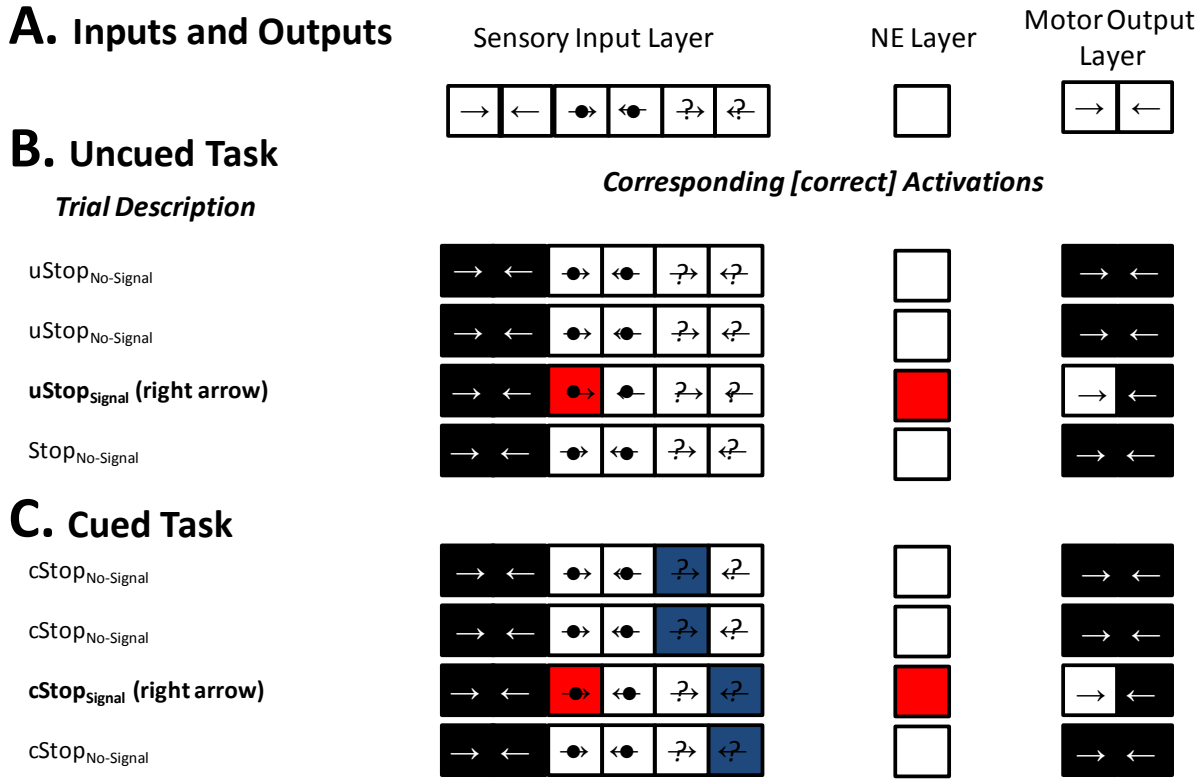


Figure 23. Inputs and Outputs of the model (A) and as activated in the Uncued (B) and Cued Stop Tasks (C). A. The sensory input layer of the model contains units for a right and left response, for a right and left stop signal, and for right and left cues. As in the model presented in Chapter 3, a simulated NE layer provides an excitatory signal to prefrontal layers upon presentation of the stop signal. A motor output layer contains units corresponding to right and left responses. B. In the Uncued stop task, every trial involves presentation of both the right and left sensory input stimuli. On a subset of trials, an effector-specific stop signal is presented with a variable delay (as indicated by the red shading) following the onset of the other sensory input. In response to this stimulation, only one of the two motor output units must be activated. C. The cued task is identical to the uncued task, except that a separate cue trial is administered prior to each performance trial, which consists only of the units shaded in blue. The information

conveyed by this signal is twofold: it indicates which of two responses might need to be stopped, but conversely also indicates which of the two responses will definitely need to be committed.

Training & Testing.

All models were run in batches of 25 networks; each network was initialized with random patterns of connection weights, and trained on one task alone (i.e., either the cued selective stop task, or the uncued stop task). This choice was made to match the fact that these tasks are behaviorally assessed in distinct blocks. Had the models been trained to perform both tasks, task-switching effects in the model would contaminate the measures we report here.

All aspects of the training, testing and analysis were identical to those performed in Chapter 3, with two exceptions. Both of these exceptions are parallel to the differences in how the selective stop tasks are assessed behaviorally, relative to the global stop task, and were thus adopted here for validity. First, SSRT was calculated separately for the two possible "hands" of the model – that is, the separate output units that corresponded to left and right responses – and averaged across them, as in behavioral work. In no case were substantial differences between these responses observed once models had reached criterial 15-85% accuracy on the Signal trials of either task. Second, interference was calculated as the difference between a given response when provided in the context of a successfully stopped Signal trial, relative to when provided in the context of Signal trials. This again is parallel to behavioral analyses conducted on this task. Interference should be positive to the extent that stopping one response leads to slowing in the commission of the other response.

Results

Consistent with the assumptions of the race model that underlie the calculation of SSRT, the use of an iterative algorithm for adjusting stop signal delay lead to approximately 50% accuracy across both tasks (Figure 24A). Accuracy on No Signal trials was at ceiling – 99% for the Uncued task and 98% for the Cued task – consistent with empirical data, which tends to show excellent performance on these trials. Finally, and further consistent with both the race model and empirical data, failed stop trials were reliably faster than No Signal trials for both the Uncued ($t(99)=6.3$, $p<.001$) and Cued tasks ($t(99)=2.7$, $p<.01$) (Figure 24B).

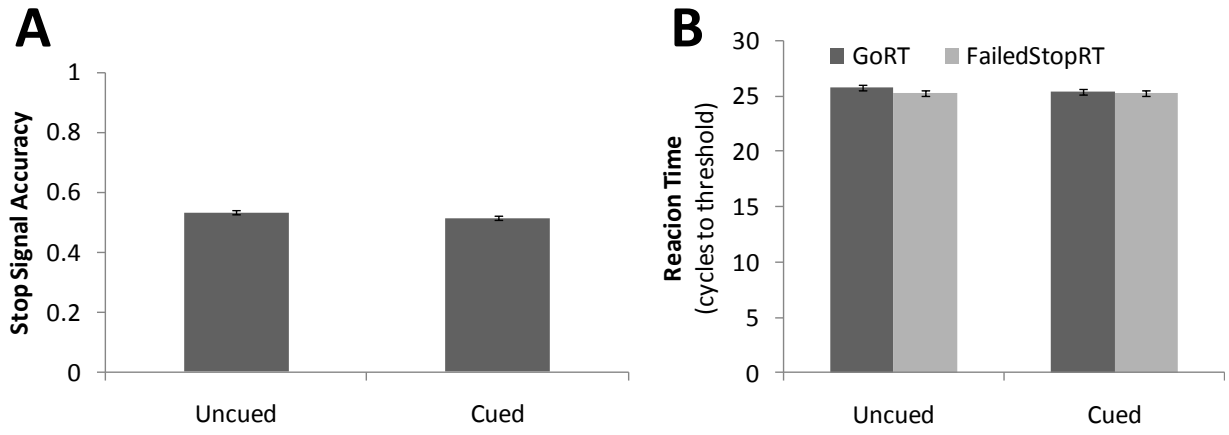


Figure 24. Results of training the model on asymptotic stop signal accuracy (A) and reductions in failed stop reaction times (B). A. Consistent with the assumption of the race model, and with empirical data on the selective stop tasks, the use of an adaptive algorithm for titrating the timing of the Stop Signal yielded approximately 50% correct performance on both tasks. B. Consistent with empirical observation, failed stop reaction times were significantly faster than "Go" reaction times on No Signal trials.

However, and contrary to the assumption of stochastic independence in the race model, SSRT and Go RT were strongly negatively correlated for both tasks (Uncued $R=-.70$, $p<.001$, Cued $R=-.77$, $p<.001$). The functional form of this relationship is clearly nonlinear for both the

Uncued (Figure 25A) and Cued tasks (Figure 25B). As described below, this nonindependence reflects substantial overlap in the generative processes yielding both stopping reactions to the stop signals and overt No Signal reaction times.

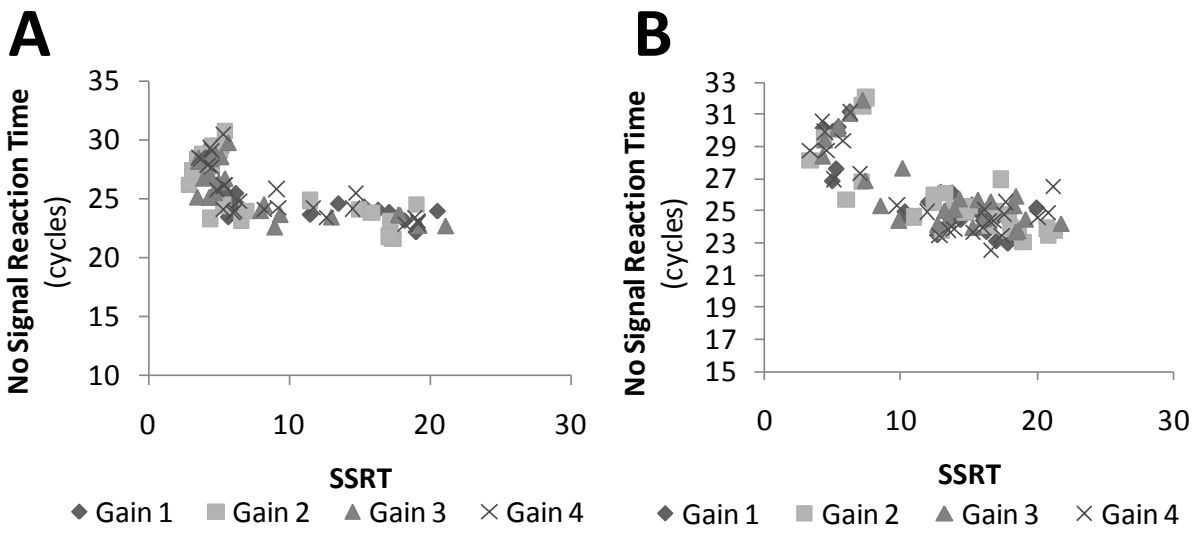


Figure 25. Nonindependence was observed between No Signal Reaction Times and SSRT on both the Uncued (A) and Cued (B) Stop Tasks, indicating a violation of the stochastic independence assumption underlying the race model. A. In the Uncued Stop task, SSRT was negatively correlated with No Signal reaction times, although the relationship was clearly nonlinear. This pattern actually reflects differential learning of the often-unacknowledged information value of stop signals across models, which is the origin of the nonindependence observed here. B. Similar patterns were observed in the Cued task, although both the nonindependence and the nonlinearity of that nonindependence were even more pronounced. This is consistent with additional learning about the often-unacknowledged information value of foreknowledge, which is the source of the more pronounced nonindependence observed in this task.

The model did give rise to the expected difference in SSRT, as previously observed in adults, as a function of foreknowledge: SSRT was reliably lengthened by foreknowledge ($t(99)=5.7$, $p<.001$; Figure 26A). However, the model showed a negative interference effect –

that is, pronounced facilitation due to presentation of the Stop signals, both with and without foreknowledge (Uncued $t(99)=3.52$, $p<.001$, Cued $t(99)=2.4$, $p<.02$; Figure 26B). Foreknowledge had no significant effect on interference ($t(99)=.8$, $p>.4$).

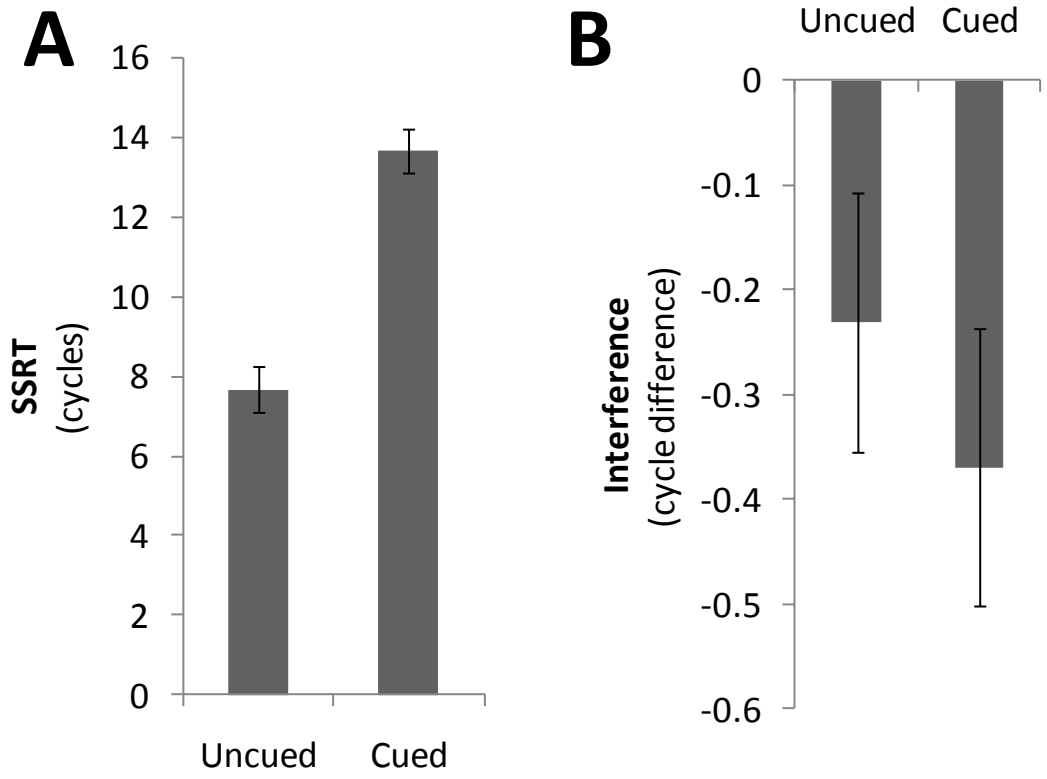


Figure 26. The effects of foreknowledge on SSRT (A) were consistent with some empirical observations, but the effects of stopping on the execution of concurrent actions led to negative interference (i.e., facilitation) that was pronounced with foreknowledge (B). A. As expected, SSRT was reliably increased by the provision of foreknowledge. B. Contrary to expectation, no models showed a significant interference of stopping on concurrent actions. Foreknowledge was without significant effect on these patterns.

We examined the reaction time distributions that give rise to interference effects in humans (Figure 27A&B for the Uncued and Cued tasks, respectively), and the corresponding facilitation effects in the model (Figure 27C & D for the Uncued and Cued tasks, respectively).

Several features of these data are noteworthy. First, in humans the reaction time distributions for correct Signal trials are shifted far to the right of the reaction time distributions for correct No Signal trials. This shift reflects the fact that incorrect Signal trials are those where responses were provided relatively earlier, and where response inhibition thus was incapable of interrupting these responses. Second, the correct Signal trial distribution exhibits a bimodality, where the left mode occurs largely within the reaction time distribution of No Signal trials. This left mode is the reduction in interference owing to foreknowledge; it occurs because the stopping initiated by the stop signal actually leads to a larger number of responses that can be emitted without being slowed by that stopping process. The local minimum between modes reflects the instantiation of that slowing.

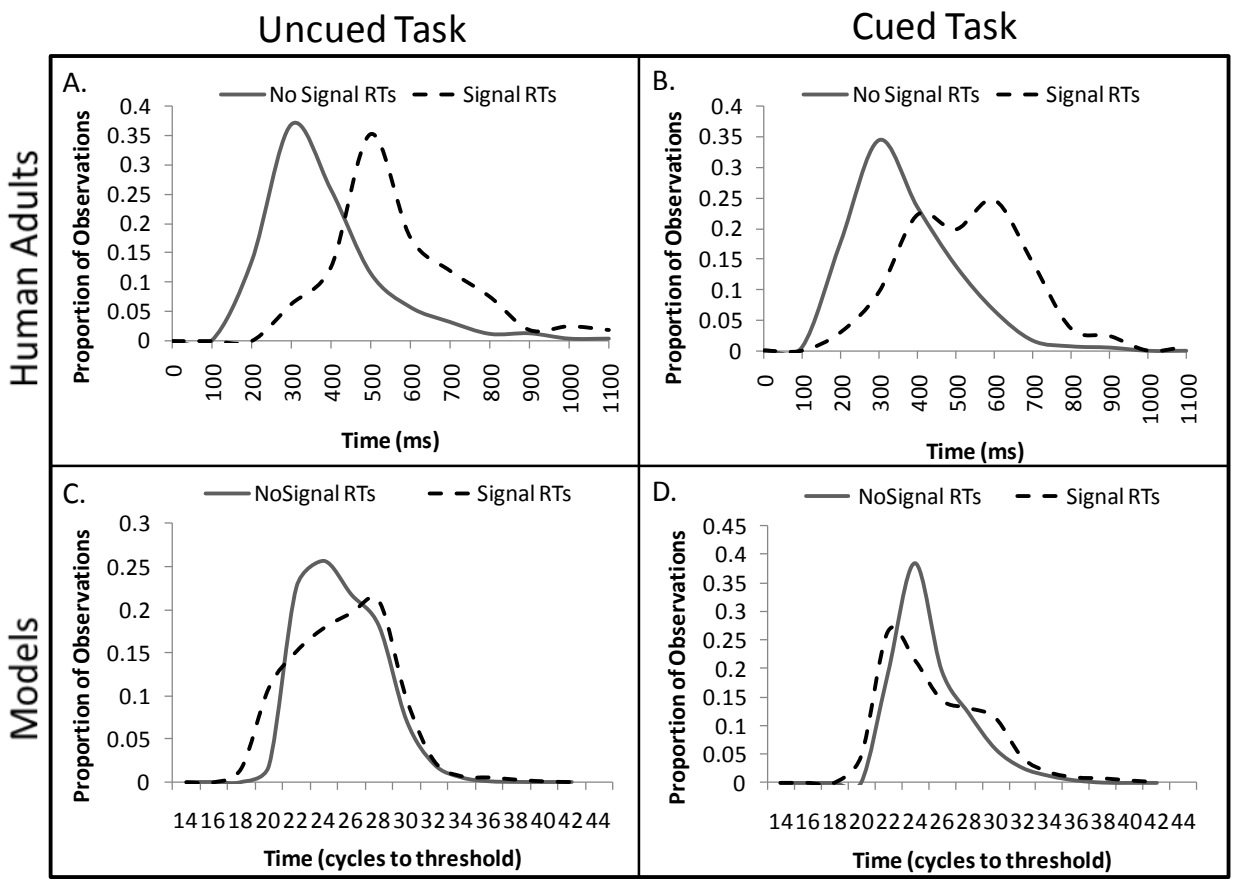


Figure 27. Reaction time distributions for both correct Signal trials (dotted lines) and correct No Signal trials (solid lines), across the Uncued (leftmost column) and Cued tasks (rightmost column) for human adults (upper row) and models (lower row). A. Human adults showed a pronounced rightward shift of their reaction times on correct Signal trials, reflecting in part the fact that the left half of the reaction time distribution is led to inhibitory failure (i.e., responses are emitted too quickly to be stopped) but also that concurrently-executed responses are significantly delayed. B. This delay is somewhat ameliorated by the provision of foreknowledge, which introduces a bimodality into the reaction time distribution on Signal trials. This bimodality reflects the fact that a larger proportion of responses continue unabated, as in the No signal trial distribution, before stopping occurs. The effect of stopping is directly reflected by the width between the modes. C. The model showed a significantly different pattern whereby the primarily effect of stopping was in fact to facilitate the concurrently-executed actions (as can be seen in the left side of the Signal RT distribution). D. This effect was largely similar to that observed when foreknowledge was provided, but with a pronounced leftward shift of the mode. A hint of bimodality visible in humans (c.f. B) is also present in the model.

Similar patterns of uni- and bimodality for the Uncued and Cued tasks, respectively, were also observed in the model, but critically without the rightward shift of the overall distributions (Figure 27C and D). This was a learned effect; a much more valid pattern was present in the first epoch of model training. (Figure 28 A & B reproduce the human data for convenience; Figure 28 C & D show the model's reaction time distributions in this first epoch.) However, other aspects of the model's behavior failed to match human data on this first epoch: they showed an inappropriately low rate of successful stopping. Subsequent training yielded improvements in stopping rate, but eliminated the desired reaction time pattern far before the models reached criterial performance.

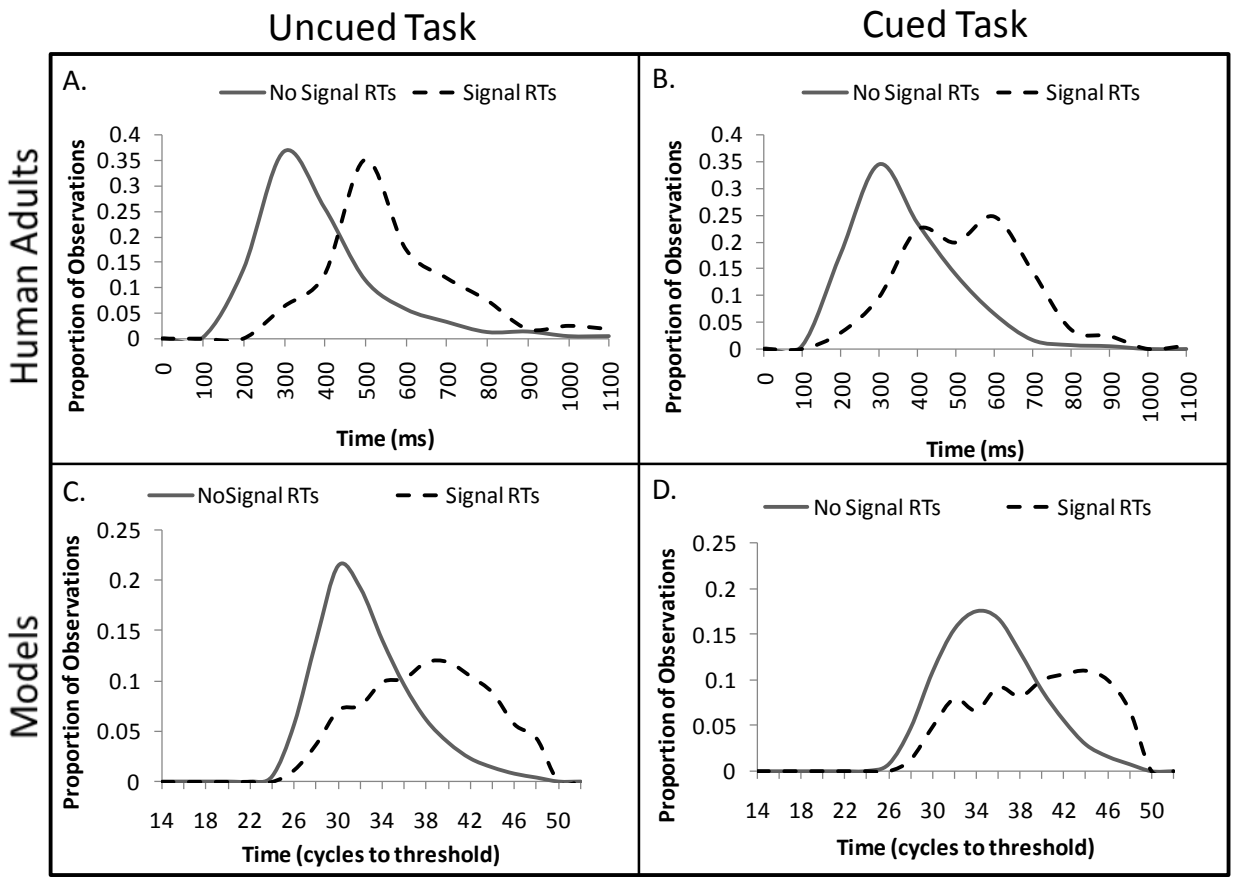


Figure 28. Learning effects were responsible for the pronounced facilitation observed in the model, as demonstrated by a comparison of the same adult data presented in Figure 6 (here, A and B) with corresponding data from the models during the first epoch of training on the Uncued task (C) and Cued Stop tasks (D). Similar to adult data from the Uncued (A) and Cued Tasks (B), early in training models showed a relatively greater proportion of responses on correct Signal trials that occurring at intervals longer than reaction times on correct No Signal trials (C), although the discrepancy between these proportions was somewhat reduced as a function of foreknowledge (D). In no case was facilitation observed within the first epoch.

Consistent with the robustness of this pattern, it arises from a very simple learning dynamic. Units throughout the model can learn to become conjunctive for the presentation of one imperative sensory stimulus (e.g., "respond left") and the stop signal pertaining to the opposite sensory stimulus (e.g., "stop right"). These units are able to rapidly respond to the

presentation of the stop signal, and then more strongly drive the imperative response ("respond left") relative to those units which detect the imperative stimulus alone (which, as described above, is by itself indeterminate; weights are accordingly weaker). This situation is slightly exacerbated when foreknowledge is provided, because this foreknowledge provides determinate information regarding the opposite sensory stimulus (e.g., if the cue was "maybe stop left" the determinate imperative stimulus is with 100% certainty "respond right").

Several alternative architectures were explored to rule out the possibility that these unwanted effects arose from peripheral sources of the model (a subset of which are schematically depicted in Figure 29). For example, the unwanted leftward shift in the Signal trial distribution could be instead viewed as an unwanted rightward shift in the No Signal distribution, if the two manual responses were interfering with one another in a way that they might not in humans. Indeed, facilitation is sometimes observed in the wider literature when two bimanually isomorphic effectors must commit symmetric actions. These facilitation effects were implemented in the model as a positive weight between congruent responses which were assigned to different layer groups, such that there was no inhibition between these groups (Figure 29A). This change actually yielded an increase in the unwanted facilitation effects on Signal trials, principally because they lengthened SSRT (and thus allowed more responses to continue unabated) without leading to a leftward shift of the No Signal distribution. The simulation of complete bimanual independence also failed to reproduce the desired patterns (Figure 29B).

It was also considered that the effect could in principle arise from an insufficiently quick form of motoric stopping, given that the shortest path from input to output was identical (two synapses) for both the posterior and prefrontal/subthalamic pathways. However, inclusion of a distinct intermediate processing step with the most basic biologically-based connectivity (Figure

29C; putatively analogous to the ventral and dorsal "what" and "how" streams; Goodale & Milner, 1995; O'Reilly, 2010) also failed to remediate the unwanted patterns of interference. Likewise, increasing and decreasing the potency of the subthalamic pathway (as simulated by changes to the gain of the subthalamic layer's activation, to the equation governing response threshold as a function of that activation, and changes to the relative weights projecting to this layer) also failed to yield the desired pattern (Figure 29D & E).

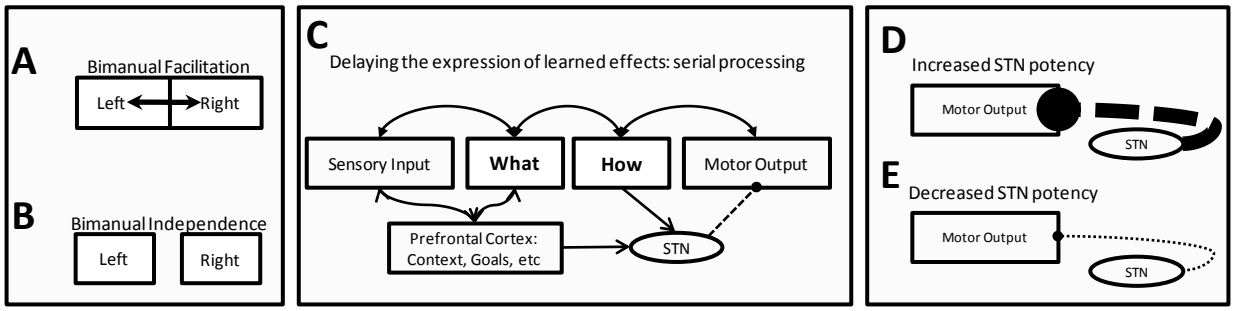


Figure 29. Alternative architectures explored in attempts to eliminate the facilitation effects observed in selective stopping, including changes to the motor output layer (A & B), changes to the posterior cortical pathway of the model (C) and changes to the potency of subthalamic mechanisms (D & E). A. Facilitation effects were not due to the absence of a bimanual facilitation process, as demonstrated by an architectural change to the output layer so that the various responses were not competitive, but instead cooperative. B. Complete bimanual independence also failed to reproduce the observed empirical patterns. C. Ultimately futile changes to the posterior cortical pathway were implemented to try to give motoric stopping processes a greater "advantage" in the race, and thus to intercept any facilitation effects. D. Neither possible change to the subthalamic layer of the model - increasing or decreasing its effect on response thresholds – were sufficient for eliminating interference.

Given the pervasiveness of this pattern within the model, and indeed its sensible origin in very domain-general learning dynamics, it seemed plausible that humans might demonstrate the same or a similar learning pattern at an effector-specific level. That is, such learning effects

might have been relatively subtle, and simply masked by differences in the reaction time distributions across hands. However, examination of the effector-specific reaction time distributions from adult subjects performing these tasks (as presented in Chapter 4) revealed no expression of any learning of this kind. In particular, on both the Uncued (Figure 30A) and Cued Tasks (Figure 30B) the reaction time distributions on Signal and No Signal trials alike were precisely matched across hands. Within the cued task, subjects also showed no tendency to even delay the cued hand (black lines), or to accelerate the uncued hand (gray lines), as would be expected if subjects proactively engaged some proactive form of selective response control. These patterns were unaffected by the duration of the task, such that there was no noticeable difference between the first and second half of each task. Apparently, no learning of this type is expressed in humans, although it appears unavoidable in the model.

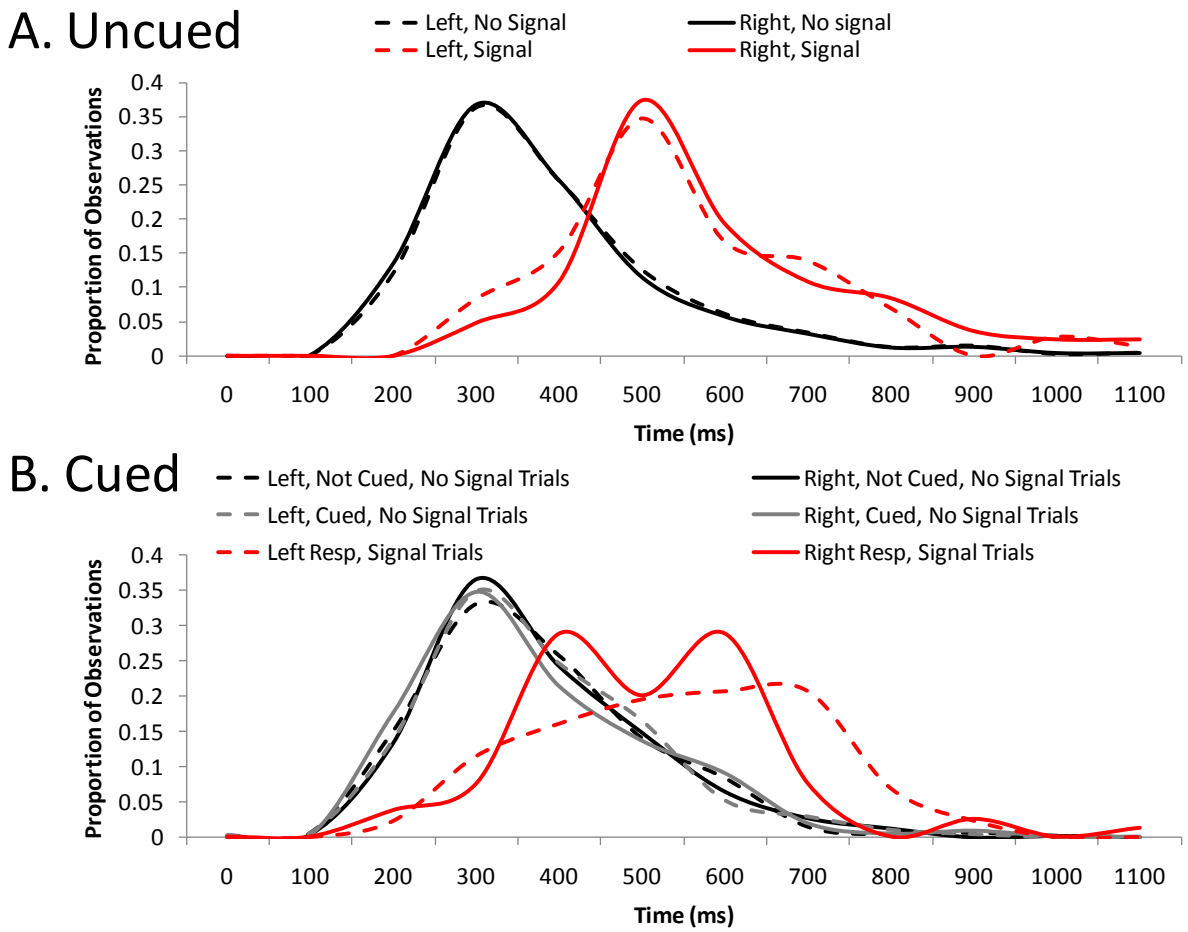


Figure 30. Reaction time distributions for Signal (red lines) and No Signal trials (black/gray lines) were examined separately for the left (dotted lines) and right (solid lines) hands in the adult subjects of Chapter 4. A. On the Uncued task, responses were strongly coupled across effectors, both on No Signal and Signal trials alike. For neither hand did subjects ever show a facilitation. B. Similar patterns were observed in the cued task, such that subjects failed to show any facilitation effects. Surprisingly, subjects also failed to delay their responses from the hands that might have needed to be stopped (grey lines), and critically also failed to accelerate their responses from the hands that would certainly not have to be stopped (black lines). The absence of these effects confirms that a significant challenge is posed by selective stopping to implementational-level theories.

Discussion

Although we found some evidence that a model capable of performing the Stop task was also capable of performing selective stop tasks, it showed a persistent nonindependence of overt reaction times and covert stopping processes. This nonindependence reflects facilitation effects arising from the most reliable information content of the stimuli presented in this task.

Specifically, the two sensory input stimuli provided to the model, one for "respond left" and one for "respond right", are not actually determinate of the correct response: a subsequent stop signal will often countermand either one of the two. This ambiguity is resolved upon presentation of the stop signal, which indicates with 100% certainty that the response to which the stop signal does not pertain will need to be committed. Foreknowledge is likewise largely indeterminate of the correct stopping response: a cue that one response may need to be stopped does not mean that response will need to be stopped. Instead, the determinate information provided by foreknowledge is that the alternative response will need to be committed (again, with 100% certainty). Both more determinate sources of information are robustly utilized by the model to produce a facilitation effect, rather than interference, when only one of multiple actions must be stopped.

The most surprising feature of this phenomenon is not that it occurs across numerous models. In retrospect, this effect is almost certain to arise in the absence of specific precautions taken to prevent it (discussed in more detail below). The more surprising fact is that this dynamic does not appear to be present in humans at any detectable level. It is of course possible that human subjects require additional training on this task before they will express these effects, in which case one might regard the current findings as prediction of the current model for a future training study. However, striatally-based learning mechanisms should be sufficient for generating these patterns even within the timeframe of the task's normal duration; striatal systems may show extremely rapid learning effects (within 60-80 trials) even when given only probabilistic input. Thus, the essential conundrum here is not that the model fails to perform at human levels; it is that humans apparently fail to learn something that the model inevitably takes strong advantage of.

These problematic learning dynamics are not intrinsic to the context-monitoring or controlled stopping accounts of response inhibition that have been contrasted in this dissertation, but both accounts must ultimately be compatible with such basic learning dynamics if they are to be considered implementationally-viable theories. There are two promising possibilities for resolving this dilemma: first, the empirical test of a unique prediction from the strategic stopping account which could falsify the context-monitoring hypothesis, and second, the reassessment of context-monitoring's viability in the domain of selective stopping with a more ecologically-valid computational model. Both will be discussed in turn.

Controlled stopping accounts could be seen to make a unique prediction regarding response-related learning effects in the context of a selective stop task. These accounts posit that prefrontal cognitive control mechanisms are biasing the use of the indirect pathway through the basal ganglia, comprised of so-called "NoGo" cells. These cells are thought to be subject to lateral competitive inhibition with the "Go" cells of the direct pathway through the basal ganglia. If Go cells are indeed suppressed via top-down biasing of the NoGo cells with which they compete, and such top-down biasing is not in fact targeted at a particular response, then a variety of learning processes may simply not occur during selective stop tasks. For example, subjects should be impaired at learning to choose a more rewarding stimulus, and relatively benefitted in learning to avoid less rewarding stimuli, if this kind of probabilistic learning paradigm were administered simultaneously with the selective stop task (i.e., a dual-task design). If correct, this phenomenon would provide extraordinarily compelling proof that selective stopping does indeed reflect a bias to use the indirect pathway, albeit in a more global fashion than typically realized.

In contrast, context-monitoring accounts make a much simpler prediction. Specifically, the failure of the current model to match human behavioral data may be a primarily technical

one, reflecting the model's lack of ecological validity, rather than a conceptual failure. The facilitation effects observed in the model arise because the representations for "respond left only" are perfectly correlated with those for "stop right," and conversely the representations for "respond right only" are perfectly correlated with those for "stop left." This leads to a facilitative activation of "respond left" representations in the presence of "stop right" representations, even though such 1:1 correspondence clearly does not characterize the learned experience of human subjects in the real world. By including a small proportion of trials in which models must "respond left only" but in which there is no demand to "stop right" (and vice versa), the model should develop uncorrelated representations for these behaviors that more closely matches those of humans.

This study has clear limitations. First, it is difficult to reason on the basis of a computational model that does not match the behavior of interest, although there are certainly compelling counterexamples (see Introduction). Second, models are useful precisely because they are simplifications of the system they simulate, but some of these simplifications can contribute to the mismatch between a model's behavior and that of humans. Future computational work will need to examine these simplifications, particularly with respect to the internal representations of responses in the model, to determine whether the current simplifications have had a deleterious effect here. Third, it is always possible that human subjects will show the pattern expressed by the model, but only with extended training. Future empirical work that assesses probabilistic learning during selective stopping may be useful in this regard: such experiments will necessitate the use of many trials over which learning effects, of the kind observed here, might manifest in people.

Nonetheless, the current work has substantial implications and does informatively shape future theorizing. In particular, it illuminates a previously unrecognized feature of selective stop tasks, and highlights important boundary conditions on current conceptual and computational theories of selective stopping. It also quantitatively demonstrates that the behavioral effects of selective stopping could be far less specific than commonly theorized, with substantial implications for the current understanding of response inhibition. Specifically, if the control mechanisms supporting response inhibition can operate only in a relatively coarse fashion – that is, only switching between top-down biasing of the global hyperdirect pathway or top-down biasing of the indirect pathway in general – it implies that our ability to control responses is far less flexible than implied by current accounts of selective stopping, but perhaps more flexible than predicted by context-monitoring accounts.

Conclusions

Subjects can selectively stop only one of multiple concurrent actions, and suffer less interference on concurrent actions when provided with foreknowledge. Strategic accounts of motoric stopping processes seem inconsistent with this hypothesis, by failing to explain the large amounts of interference that persist despite foreknowledge, by failing to explain the activation dynamics of the subthalamic nucleus, and by failing to explain the apparently uncontrolled nature of stopping in other domains. However, an alternative hypothesis is also apparently inconsistent with the absence of response commission-related learning effects in humans. This discrepancy could reflect a technical shortcoming related to the lack of ecological validity of the model. Alternatively, this discrepancy may reflect that these learning dynamics genuinely do not manifest behaviorally in selective stop tasks. This raises two possibilities: the mechanisms supporting selective stopping may not be directed at particular responses, and instead have a

simultaneously suppressive effect on learning about response commission from alternative responses; or that subjects may eventually express these learning dynamics in a way that eliminates and in fact reverses the interference commonly observed in selective stopping. In either case, the current results offer a clear computational challenge to extant theories of selective stopping and highlight a concrete direction for future work.

CHAPTER 6: SUMMARY AND CONCLUSIONS

The domain of response inhibition is exceedingly complex. This complexity arises partially from theoretical confusions regarding what processes actually support response inhibition, and theoretical ambiguities as to how widely these processes might apply. At the same time, this complexity is also intrinsic to cognitive neuroscience: any cognitive phenomenon will likely reflect the interactive contributions of many neural mechanisms, relatively few of which can be uniquely ascribed to any specific function at the cognitive level of analysis.

This complexity can be tamed with an integrative approach, of the kind taken here. Chapter 1 presented an overview of the computational underpinnings of response inhibition as they are currently understood, highlighting ambiguities in what processes comprise the "stopping" function implemented in abstract race models, diffusion models and Bayesian decision making models of the canonical Stop task. A context-monitoring hypothesis was proposed, in which the processes supporting the detection of behaviorally-relevant but infrequent stimuli might underlie the effortful and controlled ability of the Stop task. It was argued that context-monitoring might explain the similarities observed across the Stop task and other tasks (principally so-called "oddball" tasks) that do not appear to require motoric stopping per se, not only in terms of hemodynamic and electroencephalographic phenomena, but also those arising from frontal lobe pathology as well as pharmacological manipulations.

Chapter 2 presented an assessment of this hypothesis. Neuroimaging data revealed that the same region of the brain implicated in response inhibition (the rVLPFC), and often assumed to be specific to motoric stopping demands, was also recruited during a task that did not demand

any motoric stopping. This region showed prominent sustained activity across tasks, as though involved in a tonic monitoring process. Multivariate patterns in this area were more consistent across tasks, and across individuals, when demands on motoric stopping differed but demands on context-monitoring were most pronounced. An event-related potential previously thought to be characteristic of motoric stopping demands was likewise similarly expressed across tasks. Patterns of mental effort, as assessed through pupillometry, demonstrated that stopping an unwanted behavior was associated with less effort than monitoring for rare but behaviorally-relevant stimuli that failed to appear, and far less than that required for overtly responding to such stimuli. Behavioral analyses indicated that some form of motoric stopping was engaged even in this latter case – even though such stopping ran contrary to subject's goals in the task – suggesting that stopping might occur relatively automatically. A reanalysis of all data from this Chapter indicated that the presence of this stopping across tasks could not explain their hemodynamic or electroencephalographic commonalities.

Chapter 3 presented a computational implementation of the context-monitoring hypothesis, using a domain-general connectionist framework that has successfully captured numerous phenomena from other domains of cognitive control. The model also successfully captured behavioral, neuroimaging and pharmacological phenomena from the Stop task. Analysis of the model's behavior indicated that this success was driven by the development of a monitoring-like function from a more basic prefrontally-based mechanism for active maintenance. In no case did the models develop a privileged set of prefrontal units that were dedicated for motoric stopping, consistent with the hypothesis developed in Chapter 1 and assessed in Chapter 2.

Chapter 4 turned to another domain in which motoric stopping processes are thought to be under strategic cognitive control: when only one of multiple concurrent actions must be stopped. By these "strategic control" accounts, cognitive control processes can support the use of a slower but more response-specific pathway for motoric stopping when that is advantageous. However, a developmental approach yielded a double dissociation in the speed and specificity of stopping, whereby foreknowledge affected children only in terms of SSRT (and did so in the opposite way expected from published work) but only affected interference in adults. These data challenge strategic control accounts of selective response inhibition by indicating that phenomena from the selective stop task cannot be unambiguously taken as an indication of the controlled use of this slower but more response-specific pathway.

A computational model was used to explore the dynamics of selective stop task in Chapter 5, and revealed an unexpected pattern that was not previously predicted by either account. In particular, the design of the selective stop task is such that the stopping of one response (or merely the provision of probabilistic foreknowledge to that effect) can lead to a facilitation of the alternative response. While the model rapidly learned to take advantage of this characteristic of the task, adult subjects failed to do so. This discrepancy challenges both strategic control and context-monitoring accounts of response inhibition. It motivates specific future modeling work (in particular, the use of more ecologically-valid training schemes) but also the focused empirical assessments of a prediction from an elaborated version of the strategic control account of selective stopping. According to this elaborated account, subjects may fail to learn from approach-related information during the performance of selective stop tasks, owing to a rather global inhibition of the direct pathway through the basal ganglia.

Through a convergent use of computational, behavioral, electroencephalographic, pupillometric, and hemodynamic methods, the current work reduces much of the complexity in the cognitive neuroscience of response inhibition. Specifically, it suggests that the role of prefrontal cortex in response inhibition is to monitor the environmental context in support of contextually-infrequent behaviorally-relevant stimuli, and that a rather global and automatic form of motoric stopping may be engaged when the prefrontal cortex is thereby activated. This perspective accords with revisions to recent taxonomies of executive function (Stuss & Alexander, 2007) and can be computationally understood as a specific instantiation of the capacities for goal maintenance and top-down biasing enabled by the prefrontal cortex at large. Context-monitoring may also explain much of the shared variance across tasks regardless of their demands on motoric stopping. The tension between this context-monitoring perspective and those invoking strategically-controlled stopping with foreknowledge is ameliorated by our demonstration that the evidence taken to support such strategic control can be developmentally dissociated, contrary to strategic-control accounts. This developmental dissociation in fact has parallels in developmental, aging, and neuropsychological work with tasks involving rare behaviorally-relevant stimuli that do not demand motoric stopping, suggesting that these dissociations may not be specific to strategic demands on stopping, but reflect some more general dynamics in context-monitoring.

Of course much complexity remains, owing in part to clear limitations of the current work. First, a viable computational account of the dynamics of selective stopping is still outstanding. Second, it is unclear how context-monitoring processes may develop over time, given the counterintuitive effects from the developmental literature on tasks requiring the detection of rare but behaviorally-relevant stimuli (whereby early frontal event-related potentials

show a trend towards reversal as a function of stimulus predictability). Third, it remains to be demonstrated that deficits in the processes supporting detection of rare and behaviorally-relevant stimuli can consistently and parametrically explain deficits in response inhibition; such a demonstration requires the use of transcranial magnetic stimulation, psychopathologically disordered populations, or the assessment of those with frontal lobe injury, none of which was undertaken here. Indeed, such work would allow for more causal inferences which, owing to the largely correlational nature of the enclosed work, are currently speculative. Finally, the current work is firmly based in the laboratory; it is always possible that more ecologically valid studies (and models) will show patterns that contravene the current account, and reintroduce more complexity than was tamed here.

Despite these limitations, these findings have a number of larger implications. A large set of implications pertains to the basic cognitive neuroscience of frontal lobe function. A review of this wider literature, and the specific insights offered by the context-monitoring hypothesis are provided below. The second major class of implications pertains to the results of frontal lobe damage and psychopathologies like stuttering behavior, which can often be better characterized as failures of monitoring than of controlled stopping. The final set of implications pertain to the targeted intervention and remediation of inhibitory deficits, which can be fruitfully informed by context-monitoring accounts.

Implications of Context Monitoring: Broader Domains

The context-monitoring hypothesis of rVLPFC function offers an alternative and plausible account for data that might otherwise be attributed to response inhibition. However,

the rVLPFC is involved in many tasks. Below, we speculate on the potential roles of this brain region in monitoring the environment, in monitoring language processing, and in monitoring of memory.

Monitoring of the Environment

Monitoring the environment is not only important for detecting task-relevant and infrequent stimuli as in oddball and stop signal tasks. As reviewed below, the rVLPFC may play a larger role in determining our awareness of events in the environment, as they unfold across time (as revealed by task block effects and studies of the attentional blink), as they inform future behavior (as in task-switching paradigms), and as they pertain to affective states (as in studies of affective labeling and of pain).

Task Block effects. In a meta-analysis of blocked fMRI experiments, the rVLPFC was found to activate transiently in block transitions (Konishi, Donaldson, & Buckner, 2001). rVLPFC regions are recruited at these points of task-relevant contextual change. Critically, the rVLPFC was activated both during transitions from task to fixation blocks and for the reverse transition, whereas previous controlled stopping accounts of the rVLPFC might have predicted response inhibition for only the former.

Attentional Blink. In rapid serial visual presentation, the detection of the first of two briefly presented targets will often preclude the detection of the second, if distractors intervene and the inter-target interval is between 200 and 600ms (the so-called "attentional blink"). This paradigm is associated with activation of lateral prefrontal regions including the right inferior frontal gyrus (Marois, Chun, & Gore, 2000), and damage to the right prefrontal cortex causes a prolonged attentional blink (Masud Husain, Shapiro, Martin, & Kennard, 1997). The duration of this attentional blink can be predicted based on ERPs recorded over the right prefrontal cortex:

Those with attenuated attentional blinks demonstrate an earlier posterior P3 to the second target, as well as a reduced right prefrontal, P3-like component to the distractors relative to the first target (Martens, Munneke, Smid, & Johnson, 2006), both effects possibly indicating more efficient extraction of the targets by those with smaller attentional blinks.

These data are inconsistent with that expected from a controlled stopping account: If rVLPFC were required for inhibiting distractors, one would expect more involvement of the rVLPFC to distractors than targets, and that the magnitude of this difference should be larger among those with an attenuated attentional blink. Instead, this evidence is consistent with a role of the rVLPFC in monitoring for and detecting the targets, made contextually infrequent by the distractors, such that stronger engagement of the rVLPFC reflects more effective detection and interpretation of the targets when they appear.

Task-switching. The rVLPFC is one of only two prefrontal regions that are more active for unpredictable than predictable switches between tasks, particularly when the onset of trials is also unpredictable (Dreher, Koechlin, Ali, & Grafman, 2002). From an effortful and controlled stopping account, one might assume that the rVLPFC was inhibiting the undesired task set, and that this capacity must be recruited more strongly when one is uncertain of both which task set must be inhibited and when that should occur. However, even when subjects are aware both of which task set must be inhibited and when that needs to occur, the rVLPFC is more active when subjects must infer the next task's identity, relative to when they are explicitly informed of it (Forstmann, Brass, Koch & Von Cramon, 2004). Instead, the role of the rVLPFC in task-switching is more consistent with monitoring for and detecting contextual change – a capacity that is less needed when those contextual changes are explicitly-cued.

Pain. For most of us, pain is an infrequent experience; depending on its intensity, pain may also be behaviorally-relevant. Acute pain stimulation leads to rVLPFC activity (Brooks, Nurmikko, Bimson, Singh, & Roberts, 2002; Peyron et al., 1999), consistent with "a right-lateralized attentional system to alert an organism to an infrequent, but behaviorally relevant, stimulus such as pain" (Symonds, Gordon, Bixby, & Mande, 2006). Other work indicates this region may be more strongly recruited under conditions of high pain, or when the intensity of pain is explicitly task-relevant (Kong et al., 2006). Furthermore, scalp-recorded "pain evoked potentials" elicit a component with extensive similarities to the p300 oddball response (Zaslansky, Sprecher, Tenke, Hemli, & Yarnitsky, 1996). In all cases, it seems that monitoring functions may underlie the recruitment of the rVLPFC during the largely infrequent but behaviorally-relevant state of pain.

Affective Labeling. Some studies have reported rVLPFC activity when labeling the affective valence of pictures (Hariri, Bookheimer, & Mazziotta, 2000), and one reported a negative correlation with amygdala response (Lieberman et al., 2007). An interpretation based on the association of the rVLPFC with inhibition might suggest that this region actively inhibits the amygdala, notwithstanding previous evidence showing positive correlations between amygdala and rVLPFC activity (Greenberg et al., 2005). Nonetheless, a familiar problem confronts the inhibitory account: the study demonstrating a negative correlation between rVLPFC and amygdala activity failed to match positively- and negatively-valenced images in stimulus probability, unlike the relevant dimensions for other conditions (which were matched in stimulus probability). According to the context monitoring hypothesis, then, rVLPFC activity should uniquely differentiate this condition from others – due to its task-relevant and contextually-infrequent stimuli, as was indeed observed. Similar confounds appear in other work

on affective labeling, such that the rVLPFC shows greater responsiveness for those stimuli that are less frequent (AA faces were infrequent in the verbal encoding condition of (Lieberman, Hariri, Jarcho, Eisenberger, & Bookheimer, 2005) or for tasks that are contextually infrequent by virtue of block order (Hariri et al., 2000).

Monitoring of Language Processes

Although left-hemispheric regions of the prefrontal cortex are more commonly associated with language processing, the rVLPFC appears to have some role in the processing of language. For example, and as reviewed below, the role of the rVLPFC in stem-completion paradigms may be due to the processing of contextually-infrequent words. A second linguistic role for the rVLPFC is revealed by current research on discourse processing.

Stem Completion. Typical word-stem completion paradigms contrast those stems with many permissible completions (e.g., STA___) and those with few permissible completions (e.g., PSA__). In general, left inferior prefrontal regions are more strongly recruited to stems with many completions, as though these regions are important for selecting among these competitors (Thompson-Schill et al., 1999). On the other hand, rVLPFC has shown the opposite pattern – of increased recruitment to stems with few possible completions – when those completions are lower in average frequency (Desmond, Gabrieli, & Glover, 1998). This inverse pattern of sensitivity in the rVLPFC may indicate that it has less to do with selection among or suppression of competitors (for which more activity would be expected in the many completions condition) than with monitoring the process of stem retrieval for a permissible answer, which is particularly important when those permissible answers are low-frequency items.

Discourse Processing. A number of studies have focused on the role of the rVLPFC in the processing of complex semantic relationships; the consistent thread in these studies is that

contextual frequency is a major determinant of rVLPFC involvement in language processing. For example, while both the left and right VLPFC were sensitive to the presence of semantic anomalies, only the rVLPFC shows an attenuation of that response when the local context made those semantic anomalies more acceptable (Menenti, Petersson, Scheeringa, & Hagoort, 2008). The rVLPFC response to semantic anomalies may be increased following metaphorical statements (Stringaris et al., 2006), consistent with early work in neuropsychology emphasizing the role of the right hemisphere in the processing of figurative or abstract language.

More recent work indicates that the involvement of the rVLPFC in metaphor processing may be limited to novel or unusual metaphors (Lee & Dapretto, 2006; Mashal, Faust, Hendler, & Jung-Beeman, 2007). While the linguistic functions of the rVLPFC are not well understood, contextual modulations and frequency are clearly important features in determining the degree of rVLPFC involvement, consistent with a role for this region in the monitoring of context and contextual change.

Monitoring of Memory

As reviewed above, the rVLPFC may be important for the processing of contextually-infrequent but task-relevant stimuli in the environment, and even in some linguistic contexts. However, its role may not be limited to exogenous attention, as previously supposed (Corbetta & Shulman, 2002). To the contrary, evidence from memory paradigms reveal that the rVLPFC may be involved in monitoring memory processes. Here we will focus on tip-of-the-tongue and prospective memory phenomena as example domains.

Tip of the tongue phenomena. A number of studies indicate the importance of the rVLPFC in item (Nyberg et al., 1996), source (Kim et al., 2009), and autobiographical memory (Greenberg et al., 2005). While "retrieval monitoring" is sometimes offered as an explanation of

these effects, and while such accounts are not necessarily inconsistent with the context-monitoring hypothesis, other work casts doubt on the retrieval monitoring hypothesis with respect to the rVLPFC: it is not always involved in successful remembering (Henson et al., 2000; Henson et al., 1999; Kikyo, Ohki, & Miyashita, 2002). Instead, it may be more strongly recruited in feeling-of-knowing or tip-of-the-tongue (TOT) states, in which subjects are not able to retrieve a memory but have the subjective impression that they will (Kikyo et al., 2002; Maril, Wagner, & Schacter, 2001). TOT states are difficult to induce, typically occurring only on a small subset of trials (10-20%) even with experimental designs optimized to produce them and subjects trained to report them. Thus, TOT states constitute a contextually-infrequent but task-relevant experience in any memory task, and should therefore evoke rVLPFC activity independent of any retrieval-specific role the rVLPFC may have in monitoring. .

Prospective Memory. Prospective memory – "remembering to remember" – shares many similarities with task-switching; accordingly, fMRI studies reveal that prospective memory recruits a neural network very similar to that observed in studies of task-switching (Simons, Schlövinck, Gilbert, Frith, & Burgess, 2006). Dominant theoretical accounts of prospective memory posit that subjects engage in a strategic monitoring of the environment for prospective memory cues (Burgess, Quayle, & Frith, 2001; Smith, 2003). Indeed, activity in the rVLPFC is observed during the identification of prospective memory cues, and also shows elevated activity during blocks where cue identification is particularly difficult (Simons et al., 2006), consistent with a role for this region in monitoring the environment. Furthermore, no differential activity in the rVLPFC is observed when prospective memory processes are contrasted with an oddball task (Jeremy R. Reynolds, West, & Braver, 2008), suggesting that the functional role of the rVLPFC

in prospective memory may be limited to processing the infrequent but task-relevant stimuli present in both paradigms.

Implications of Context-monitoring: Frontal Lobe Damage and Psychopathology

The second major set of implications arising from the context-monitoring hypothesis pertain to not only to the putatively-inhibitory deficits arising from damage to the rVLPFC, but also to cases where inhibition does not appear to explain the effects of rVLPFC damage. Context-monitoring also strongly informs the characterization of psychopathologies that are sometimes thought to be associated with deficits in response inhibition, including Attention Deficit Hyperactivity Disorder (ADHD) and stuttering. The insights provided by the context-monitoring hypothesis in each of these cases is discussed in turn, below.

Inhibitory Deficits due to rVLPFC Damage. Evidence from brain damaged patients indicates that lesions to the rVLPFC may be associated with deficits in task switching (Aron, Monsell et al., 2004; Stuss & Alexander, 2007). The precise pattern of these deficits is strongly indicative of failures in monitoring for contextual change. Right lateral prefrontal patients fail to benefit not only from larger preparation time for task switches (Stuss & Alexander, 2007). but may more generally fail to check for the occurrence of stimuli, and thus fail to increase response readiness over time (Stuss et al., 2005; Vallesi, unpublished PhD Dissertation). The role of rVLPFC in context-monitoring strongly converges with these findings and offers a unified way of characterizing these effects of damage to right lateral prefrontal cortex.

This framework also provides a new perspective on detailed patterns of performance that have been previously interpreted to reflect response inhibition deficits. For example, focal

rVLPFC damage can lead to poor target detection, such that even when the location of an upcoming target is cued before trial onset, this location is not effectively monitored following the onset of any stimulus (Michael et al., 2006). The context-monitoring hypothesis suggests this patient's difficulty reflects a deficit in monitoring contextually-appropriate locations in the service of behaviorally-relevant stimulus detection.

Spatial Neglect. While spatial neglect is traditionally associated with damage to the right parietal lobe, spatial neglect also occurs following damage to the rVLPFC (Husain & Kennard, 1996; Husain et al., 1997). It appears that deficits in phasic alerting underlie some forms of spatial neglect: in these cases, damage-induced spatial bias is abolished when patients are warned in advance about impending trials (Robertson, Mattingley, Rorden, & Driver, 1998). Frontal spatial neglect is also associated with abnormal P3 oddball effects (for a review, see Deouell, Hämäläinen, & Bentin, 2000). This effect can be understood if spatial neglect arising from rVLPFC damage reflects a deficit in monitoring the environment for behaviorally-relevant information.

Stuttering. Stuttering is often associated with over-activation of the rVLPFC (Fox et al., 1996), which has been interpreted in terms of "an overactive stopping process which may inappropriately brake speech output" (Xue, Aron, & Poldrack, 2008). However, stutterers also show abnormal behavioral and ERP profiles in classic oddball paradigms (Hampton & Weberfox, 2008; Morgan, Cranford, & Burk, 1997), indicating the deficit may not be unique to stopping demands. rVLPFC activation is also negatively correlated with speech dysfluencies (Preibisch et al., 2003), and it remains more highly activated in those who stutter than controls following unaided recovery (Kell, Kriegstein, Neumann, & Giraud, 2007), contrary to the

proposal that an overactive "stopping" process is responsible for stuttering and reflected in rVLPFC activity.

An alternative account posits that an abnormal monitoring process is chronically engaged in checking phonology for errors among those who stutter (Vasic & Wijnen, 2005). In addition to correctly predicting a number of behavioral findings, this monitoring theory suggests that therapies like delayed auditory feedback and frequency altered feedback may actually work as "alerting" cues that serve to orient attention away from phonology, similar to noise-induced fluency (Postma & Kolk, 1992). Those with less severe stuttering may even adopt a compensatory strategy of self-orienting away from phonology in linguistic contexts, leading to greater activity in the rVLPFC (Preibisch et al., 2003).

Attention deficit hyperactivity disorder (ADHD). While ADHD is often understood as a deficit in response inhibition – such that hyperactivity and inappropriate shifts of attention both reflect problems with the act of stopping those unwanted behaviors – the data do not support a deficit that is stopping-specific (Vaurio, Simmonds & Mostofsky, 2009). In particular, meta-analysis reveals that ADHD is more strongly associated with reaction time variability overall than with deficits in SSRT, as might arise from deficits in the ability to effectively monitor the environment for behaviorally-relevant information. It is possible that ADHD is actually most directly associated with an uncoupling of behavioral-relevance from monitoring processes, such that many items in the environment – regardless of their behavioral-relevance – are deemed worthy of attention.

Implications of Context-Monitoring: Targeted Interventions

The context-monitoring hypothesis not only provides insights regarding the basic cognitive neuroscience of frontal lobe function and dysfunction, but also motivates more targeted interventions. Interventions for frontal lobe insult or pathology often focus on the training of response inhibition when they might more fruitfully target context monitoring.

One method by which this might occur is alerting-based training. In this domain, subjects are trained to more closely monitor their environment for behaviorally-relevant stimuli, which require response commission. Such training has lasting effects in the rVLPFC (Thimm, Fink, Küst, Karbe, & Sturm, 2006), and the prognosis for recovery in frontal spatial neglect can be predicted on the basis of preserved alerting function (Robertson et al., 1998). These data thus suggest that, to the extent the rVLPFC is subserving a common context-monitoring function across tests of response inhibition and indeed many domains of cognitive functioning, alerting-based training might be a particularly fruitful direction for applied work.

Meditation training may be another method by which effective interventions could occur, because it may influence monitoring processes and rVLPFC in a way that generalizes beyond the meditative experience itself (Lutz, Slagter, Dunne & Davidson, 2008). For example, meditation training reduces the attentional blink (Slagter et al., 2007), a phenomenon which may rely crucially on rVLPFC monitoring functions in particular (as described above). Meditation training also improves subjects' ability to maintain vigilance in the absence of a warning about an upcoming target stimulus (Jha, Krompinger & Baime, 2007). Meditation experience producing changes in rVLPFC cortical thickness (Lazar, et al., 2005). The act of meditating occurs alongside increased recruitment of the rVLPFC, and subjective reports on the depth of the achieved meditative state are positively correlated with rVLPFC activation (Lutz, Brefczynski-Lewis, Johnston & Davidson, 2008), perhaps suggesting that the act of meditating exercises the

cognitive functions subserved by this region. Together, this evidence implies that meditation training might be an important dimension for interventions in populations with deficits in response inhibition as a function of meditation's effects on monitoring processes; indeed, meditation training has already shown some promise as an intervention technique for ADHD (Baijal & Gupta, 2008). More traditional cognitive interventions have so far been met with limited success (Alderson et al., 2008).

Final Remarks

A reframing of the cognitive control demands of response inhibition as context monitoring requires an overhaul of the interpretation and understanding of many associated cognitive phenomena, deficits from frontal injury, and psychopathology. There are several limitations to the current work, both with respect to the substrates that might support more selective forms of stopping, and the developmental trajectory of these abilities. Nonetheless, context-monitoring usefully informs basic theorizing about the functions of the frontal lobes, and the rVLPFC in particular, across a wide number of domains, including those that involve the monitoring of the environment, of language, and of memory. The context-monitoring hypothesis may further motivate specific cognitive interventions for frontal lobe dysfunctions and pathologies characterized by deficits in response inhibition, including alerting and meditation training.

REFERENCES

Aisa B, Mingus B, O'Reilly RC (2008) The emergent neural modeling system. Neural Networks, 21(8):1146-1152.

Altmann, E. M., & Gray, W. D. (2008). An integrated model of cognitive control in task switching. *Psychological review, 115*(3), 602-639.

Anderson, M. C. (2005). The role of inhibitory control in forgetting unwanted memories: A consideration of three methods. In C. M. Macleod & B. Uttl (Eds.), *Dynamic Cognitive Processes* (pp. 159-190). Tokyo: Springer-Verlag.

Aron AR (2007) The neural basis of inhibition in cognitive control. Neuroscientist 13:214-28

Aron AR, Monsell S, Sahakian BJ, Robbins TW (2004) A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. *Brain* 127:1561-1573.

Aron AR, Poldrack RA (2006) Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. *J Neurosci* 26:2424-2433.

Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. *Trends Cogn Sci* 8:170-7.

Aron, A. R. (2003). Methylphenidate improves response inhibition in adults with attention-deficit/hyperactivity disorder. *Biological Psychiatry, 54*(12), 1465-1468.

Aron, A. R. (2007). The Neural Basis of Inhibition in Cognitive Control. *Neuroscientist, 13*(3), 214-228.

Aron, A. R., & Poldrack, R. (2005). The Cognitive Neuroscience of Response Inhibition: Relevance for Genetic Research in Attention-Deficit/Hyperactivity Disorder. *Biological Psychiatry, 57*(11), 1285-1292.

Aron, A. R., & Poldrack, R. (2006). Cortical and Subcortical Contributions to Stop Signal Response Inhibition: Role of the Subthalamic Nucleus. *J. Neurosci., 26*(9), 2424-2433.

Aron, A. R., & Verbruggen, F. (2009). Stop the Presses: Dissociating a Selective From a Global Mechanism for Stopping. *Psychological Science, 19*(11), 1146-1153.

Aron, A. R., Monsell, S., Sahakian, B. J., & Robbins, T. W. (2004). A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. *Brain : a journal of neurology, 127*(Pt 7), 1561-1573.

Aron, A. R., Robbins, T., & Poldrack, R. (2004). Inhibition and the right inferior frontal cortex. *Trends in Cognitive Sciences, 8*(4), 170-177.

Aron, AR (2010) From Reactive to proactive and Selective Control: Developing a Rcicher Model for Stopping Inappropriate Responses. *Biol Psychiat,* 69(12):e55-68.

Azizian A, Freitas AL, Parvaz MA & Squires NK (2006) Beware misleading cues: Perceptual similarity modulates the N2/P3 complex. *Psychophysiol*, 43:253-260.

Band, G. P., van der Molen, M. W., & Logan, G. D. (2003). Horse-race model simulations of the stop-signal procedure. *Acta Psychol (Amst), 112*(2), 105-142.

Band, GPH., Ridderinkhof, KR. & Van der Molen, MW. (2003). Speed-Accuracy Modulation in Case of Conflict: The Roles of Activation and Inhibition. *Psychol. Res.,* 67, 266-279.

Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. *Neuropsychologia*, 35: 1373-1380.

Barry, R. J., Clarke, A. R., & Johnstone, S. J. (2003). A review of electrophysiology in attention-deficit/hyperactivity disorder: I. Qualitative and quantitative electroencephalography. *Clinical Neurophysiology, 114*(2), 171-183.

Beatty J, Lucero-Wagoner, B. (2000) The Pupillary System. In: Caccioppo J, Tassinary LG, Berntson G, editors. *The Handbook of Psychophysiology*. Hillsdale: Cambridge University Press, pp 142-62

Bekker EM, Kenemans JL & Verbaten MN (2005). Source analysis of the N2 in a cued Go/No Go Task. *Brain Res Cogn Brain Res*. 22:221-31.

Bekker, E. M., Bocker, K. B., Van Hunsel, F., van den Berg, M. C., & Kenemans, J. L. (2005). Acute effects of nicotine on attention and response inhibition. *Pharmacology, Biochemistry, and Behavior, 82*(3), 539-548.

Boehler, C. N., Mante, T. F., Krebs, R. M., Heinze, H. J., Schoenfeld, M. A., & Hopf, J. M. (2009). Sensory MEG responses predict successful and failed inhibition in a stop-signal task. *Cerebral Cortex 19*(1), 134-145.

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*(2), 376-397.

Brazdil M, Dobsik M, Mikl M, Hlustik P, Daniel P, Pazourkova M, Krupe P, Rektor P (2005) Combined event related fMRI and intracerebral ERP study of an auditory oddball task.

Neuroimage 26:285-393.

Brooks, J. C., Nurmikko, T. J., Bimson, W. E., Singh, K. D., & Roberts, N. (2002). fMRI of thermal pain: effects of stimulus laterality and attention. *NeuroImage, 15*(2), 293-301.

Bubic, A., von Cramon, D. Y., Jacobsen, T., Schroger, E., & Schubotz, R. I. (2008). Violation of Expectation: Neural Correlates Reflect Bases of Prediction. *Journal of Cognitive Neuroscience*.

Burgess, P. W., Quayle, A., & Frith, C. D. (2001). Brain regions involved in prospective memory as determined by positron emission tomography. *Neuropsychologia, 39*(6), 545-555.

Cabeza, R., Locantore, J. K., & Anderson, N. D. (2003). Lateralization of prefrontal activity during episodic memory retrieval: evidence for the production-monitoring hypothesis. *Journal of Cognitive Neuroscience, 15*(2), 249-259.

Cai W, Leung HC (2011) Rule-Guided Executive Control of Response Inhibition: Functional Topography of the Inferior Frontal Cortex. *PLoS One* 6(6): e20840. doi:10.1371/journal.pone.0020840

Cappelletti, M., Fregni, F., Shapiro, K., Pascual-Leone, A., & Caramazza, A. (2007). Processing Nouns and Verbs in the Left Frontal Cortex: A Transcranial Magnetic Stimulation Study. *J Cogn Neurosci*. 20(4):707-20.

Casada, J. H., & Roache, J. D. (2005). Behavioral inhibition and activation in posttraumatic stress disorder. *J Nerv Ment Dis, 193*(2), 102-109.

Casey, B. J., Trainor, R., Orendi, J., Schubert, A., Nystrom, L., Giedd, J., et al. (1997). A developmental functional mri study of prefrontal activation during performance of a go-no-go task. *J. Cognitive Neuroscience, 9*(6), 835-847.

Castellanos FX, Sonuga-Barke EJS, Milham MP, Tannock R (2006) Characterizing cognition in ADHD: beyond executive dysfunction. *Trends Cog Sci* 10: 117-123.

Chamberlain, S., Muller, U., Blackwell, A., Clark, L., Robbins, T., & Sahakian, B. (2006). Neurochemical Modulation of Response Inhibition and Probabilistic Learning in Humans. *Science, 311*(5762), 861-863.

Chambers, C., Bellgrove, M., Stokes, M., Henderson, T., Garavan, H., Robertson, I., et al. (2006). Executive "Brake Failure" following Deactivation of Human Frontal Lobe. *Journal of Cognitive Neuroscience, 18*(3), 444-455.

Chevrier, A. D., Noseworthy, M. D., & Schachar, R. (2007). Dissociation of response inhibition and performance monitoring in the stop signal task using event-related fMRI. *Hum Brain Mapp*.

Chikazoe J, Jimura K, Hirose S, Yamashita K, Miyashita Y, Konishi S (2009) Preparation to inhibit a response complements response inhibition during performance of a stop-signal task. *J Neurosci* 29:15870–15877.

Chikazoe J, Konishi S, Asari T, Jimura K, Miyashita Y (2007) Activation of right inferior frontal gyrus during response inhibition across response modalities. *J Cogn Neurosci* 19:69-80.

Chikazoe, J., Jimura, K., Asari, T., Yamashita, K.-I., Morimoto, H., Hirose, S., et al. (2009). Functional Dissociation in Right Inferior Frontal Cortex during Performance of Go/No-Go Task. *Cereb. Cortex, 19*(1), 146-152.

Chikazoe, J., Konishi, S., Asari, T., Jimura, K., & Miyashita, Y. (2007). Activation of Right Inferior Frontal Gyrus during Response Inhibition across Response Modalities. *J. Cogn. Neurosci., 19*(1), 69-80.

Civardi C, Cantello R, Asselman P, Rothwell JC (2001) Transcranial magnetic stimulation can be used to test connections to primary motor areas from frontal and medial cortex in humans. *Neuroimage* 14:1444–1453.

Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nat Neurosci*. 3:292-7.

Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. *Neuron* 58:306-324.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci, 3*(3), 201-215.

Corbetta, M., Patel, G., & Shulman, G. L. (2008). The Reorienting System of the Human Brain: From Environment to Theory of Mind. *Neuron, 58*(3), 306-324.

Coxon, J., Stinear, C., & Byblow, W. (2007). Selective Inhibition of Movement. *J Neurophysiol, 97*(3), 2480-2489.

Craig AB (2009) How do you feel – now? The anterior insula and human awareness. *Nat Rev Neurosci* 10:59-70.

De Jong, R., Berendsen, E., & Cools, R. (1999). Goal neglect and inhibitory limitations: dissociable causes of interference effects in conflict situations. *Acta Psychologica, 101*(2-3), 379-394.

de Wit, H., Crean, J., & Richards, J. B. (2000). Effects of d-amphetamine and ethanol on a measure of behavioral inhibition in humans. *Behavioral neuroscience, 114*(4), 830-837.

Dempster, F. N. (1992). The rise and fall of the inhibitory mechanism: Toward a unified theory of cognitive development and aging. *Developmental Review, 12*(1), 45-75.

Deouell, L. Y., Hämäläinen H., & Bentin, S. (2000). Unilateral neglect after right-hemisphere damage: contributions from event-related potentials. *Audiol Neuroot, 5*(3-4), 225-234.

Depue, B. E., Curran, T., & Banich, M. T. (2007). Prefrontal regions orchestrate suppression of emotional memories via a two-phase process. *Science (New York, N.Y.), 317*(5835), 215-219.

Desmond, J., Gabrieli, J., & Glover, G. (1998). Dissociation of Frontal and Cerebellar Activity in a Cognitive Task: Evidence for a Distinction between Selection and Search. *NeuroImage, 7*(4), 368-376.

Diamond, A. (1990). Developmental time course in human infants and infant monkeys, and the neural bases of, inhibitory control in reaching. *Annals of the New York Academy of Sciences, 608*.

Diamond, A. (2009). All or None Hypothesis: A Global-Default Mode That Characterizes the Brain and Mind. *Developmental Psychology, 45*(1), 130-138.

Dillon, D. G. & Pizzagalli, D. A. (2007). Inhibition of action, thought, and emotion: A selective neurobiological review. *Applied and Preventive Psychology, 12*(3), 99-114.

Dodds CM, Morein-Zamir S, Robbins TW (2011) Dissociating Inhibition, Attention, and Response Control in the Frontoparietal Network Using Functional Magnetic Resonance Imaging. *Cerebral Cortex* 21:1155-1165.

Doeller, C., Opitz, B., Mecklinger, A., Krick, C., Reith, W., & Schrager, E. (2003). Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. *NeuroImage, 20*(2), 1270-1282.

Donchin, E., & Coles, M. (1998). Context updating and the P3. *Behavioral and Brain Sciences, 21*(01), 152-154.

Donkers FCL & van Boxtel GJM (2004) The N2 in Go/No-Go Tasks reflects conflict monitoring not response inhibition. *Brain and Cognit*, 56:165-176.

Downar, J., Crawley, A., Mikulis, D., & Davis, K. (2002). A Cortical Network Sensitive to Stimulus Salience in a Neutral Behavioral Context Across Multiple Sensory Modalities. *J Neurophysiol, 87*(1), 615-620.

Drake, M. E., Phillips, B. B., & Pakalnis, A. (1991). Auditory evoked potentials in borderline personality disorder. *Clinical EEG, 22*(3), 188-192.

Dreher, J. C., Koechlin, E., Ali, S. O., & Grafman, J. (2002). The roles of timing and task order during task switching. *NeuroImage, 17*(1), 95-109.

Eagle, Dawn, Tufft, Miles, Goodchild, Hannah, et al. (2007). Differential effects of modafinil and methylphenidate on stop-signal reaction time task performance in the rat, and interactions with the dopamine receptor antagonist cis-flupenthixol. *Psychopharmacology, 192*(2), 193-206.

Egner, T., & Hirsch, J. (2005). Where Memory Meets Attention: Neural Substrates of Negative Priming. *Journal of Cognitive Neuroscience, 17*(11), 1774-1784.

Ettinger, U., Ffytche, D., Kumari, V., Kathmann, N., Reuter, B., Zelaya, F., et al. (2008). Decomposing the Neural Correlates of Antisaccade Eye Movements Using Event-Related fMRI. *Cereb. Cortex, 18*(5), 1148-1159.

Falkenstein M, Hoormann J & Hohnsbein J (1999) ERP components in Go/No Go Tasks and their relation to inhibition. *Acta Psychol*. 101:267-91.

Fallgatter AJ, Mueller TJ & Strik WK (1999) Age-related changes in the brain electrical correlates of response control. *Clin Neurophysiol*. 110:833-8.

Fan, J., Byrne, J., Worden, M. S., Guise, K. G., McCandliss, B. D., Fossella, J., et al. (2007). The relation of brain oscillations to attentional networks. *J Neurosci, 27*(23), 6197-6206.

Ferrier, D (1876) The functions of the brain. London: Elder.

Folstein, J., & Van Petten, C. (2008). Influence of cognitive control and mismatch on the N2 component of the ERP: A review. *Psychophysiology, 45*(1), 152-170.

Forstmann, B., Jahfari, S., Scholte, S., Wolfensteller, U., van den Wildenberg, W., & Ridderinkhof, R. (2008). Function and Structure of the Right Inferior Frontal Cortex Predict Individual Differences in Response Inhibition: A Model-Based Approach. *J. Neurosci., 28*(39), 9790-9796.

Fox, M. D., Snyder, A. Z., Barch, D. M., Gusnard, D. A., Raichle, M. E. (2005). Transient BOLD Responses at Block Transitions. *Neuroimage, 28*(4), 956-966.

Fox, P. T., Ingham, R. J., Ingham, J. C., Hirsch, T. B., Downs, J. H., Martin, C., et al. (1996). A PET study of the neural systems of stuttering. *Nature, 382*(6587), 158-161.

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw*.

Friedman NP & Miyake, A (2004) The relations among inhibition and interference control

functions: A latent variable analysis. *J Exp Psychol Gen*, 133, 101-135.

Friedman, N. P., Miyake, A., Young, S. E., Defries, J. C., Corley, R. P., & Hewitt, J. K. (2008). Individual differences in executive functions are almost entirely genetic in origin. *Journal of experimental psychology. General, 137*(2), 201-225.

Gallinat, J., Bajbouj, M., Sander, T., Schlattmann, P., Xu, K. & Ferro, E. F. (2003). Association of the G1947A COMT (Val 108/158Met) gene polymorphism with prefrontal P3 duing information processing. *Biological Psychiatry* 54, 40–48.

Gehring, W. J., Himle, J., & Nisenson, L. G. (2000). Action-monitoring dysfunction in obsessive-compulsive disorder. *Psychological Science, 11*(1), 1-6.

Gilzenrat, MS, Nieuwenhuis S, Jepma M, Cohen JD (2010). Pupil diameter tracks changes in control state predicted by the adaptive gain theory of locus coeruleus function. *Cogn Affect Behav Neurosci* 10:252-69 .

Greenberg, D. L., Rice, H. J., Cooper, J. J., Cabeza, R., Rubin, D. C., & Labar, K. S. (2005). Co-activation of the amygdala, hippocampus and inferior frontal gyrus during autobiographical memory retrieval. *Neuropsychologia, 43*(5), 659-674.

Hämmerer D, Li SC, Müller V & Lindenberger U (2010) An electrophysiological study of response conflict processing across the lifespan: assessing the roles of conflict monitoring, cue utilization, response anticipation, and response suppression. *Neuropsychologia*. 48:3305-16.

Hampshire A, Chamberlain SR, Monti MM, Duncan J, Owen AM (2010) The role of the right inferior frontal gyrus: inhibition and attentional control. *Neuroimage* 50: 1313-9.

Hampton, A., & Weberfox, C. (2008). Non-linguistic auditory processing in stuttering: Evidence from behavior and event-related brain potentials. *Journal of Fluency Disorders*. 33(4): 253–273.

Hariri, A. R., Bookheimer, S. Y., & Mazziotta, J. C. (2000). Modulating emotional responses: effects of a neocortical network on the limbic system. *Neuroreport, 11*(1), 43-48.

Hazeltine, E., Bunge, S. A., Scanlon, M. D., & Gabrieli, J. D. (2003). Material-dependent and material-independent selection processes in the frontal and parietal lobes: an event-related fMRI investigation of response competition. *Neuropsychologia, 41*(9), 1208-1217.

Henson, R. N. A., Burgess, N., & Frith, C. D. (2000). Recoding, storage, rehearsal and grouping in verbal short-term memory: an fMRI study. *Neuropsychologia, 38*(4), 426-440.

Henson, R. N. A., Shallice, T., & Dolan, R. J. (1999). Right prefrontal cortex and episodic memory retrieval: a functional MRI test of the monitoring hypothesis. *Brain, 122*(7), 1367-1381.

Herd, S., Banich, M., & O'Reilly, R. (2006). Neural Mechanisms of Cognitive Control: An Integrative Model of Stroop Task Performance and fMRI Data. *Journal of Cognitive Neuroscience, 18*(1), 22-32.

Hodgson, T., Chamberlain, M., Parris, B., James, M., Gutowski, N., Husain, M., et al. (2007). The role of the ventrolateral frontal cortex in inhibitory oculomotor control. *Brain* (Pt 6), 1525-1537.

Holmes, E. A., Moulds, M. L., & Kavanagh, D. (2007). Memory suppression in PTSD treatment? *Science, 318*(5857).

Husain, M., & Kennard, C. (1996). Visual neglect associated with frontal lobe infarction. *Journal of Neurology, 243*(9), 652-657.

Husain, M., Shapiro, K., Martin, J., & Kennard, C. (1997). Abnormal temporal dynamics of visual attention in spatial neglect patients. *Nature, 385*(6612), 154-156.

Jahfari S, Stinear C, Claffey M, Verbruggen F, & Aron AR. (2010) Responding with restraint: What are the neurocognitive mechanisms? *J Cog Neuro*, 22: 1479–1492.

Jennings JR, van der Molen MW & Tanase C (2009) Preparing hearts and minds: cardiac slowing and a cortical inhibitory network. *Psychophys*. 46:1170-8.

Jiang Y, Saxe R & Kanwisher N. (2004) Functional Magnetic Resonance Imaging Provides New Constraints on Theories of the Psychological Refractory Period. *Psychol Sci*,15:390-396.

Jonkman LM, Sniedt FL, Kemner C (2007) Source localization of the Nogo-N2: a developmental study. *Clin Neurophysiol* 118:1069-77.

Junghoefer M, Elber T, Tucker DM, Braun C (1999) The polar average reference effect: A bias in estimating the head surface integral in EEG recording. *Clinical Neurophysiology*, 110, 1149–1155.

Kahneman D, Beatty J (1966) Pupillary changes in two memory tasks. *Science* 154:1583-5.

Kaladjian, A., Jeanningros, R., Azorin, J. M., Grimault, S., Anton, J. L., & Mazzola-Pomietto, P. (2007). Blunted activation in right ventrolateral prefrontal cortex during motor response inhibition in schizophrenia. *Schizophrenia Research, 97*(1-3), 184-193.

Kane, M. J., & Engle, R. W. (2003). Working-memory capacity and the control of attention: the contributions of goal neglect, response competition, and task set to Stroop interference. *Journal of experimental psychology. General, 132*(1), 47-70.

Karl, A., Malta, L. S., & Maercker, A. (2006). Meta-analytic review of event-related potential

studies in post-traumatic stress disorder. *Biological Psychology, 71*(2), 123-147.

Kell, C., Kriegstein, K., Neumann, K., & Giraud, A. (2007). Imaging of the recovery from stuttering reveals spontaneous neuroplasticity. *Clinical Neurophysiology, 118*(4), e57-e57.

Kiehl, K. A., Laurens, K. R., Duty, T. L., Forster, B. B., & Liddle, P. F. (2001). An event-related fMRI study of visual and auditory oddball tasks. *15*(4), 221-240.

Kikyo, H., Ohki, K., & Miyashita, Y. (2002). Neural correlates for feeling-of-knowing: an fMRI parametric analysis. *Neuron, 36*(1), 177-186.

Kim, Y. Y., Roh, A. Y., Namgoong, Y., Jo, H. J., Lee, J. M., & Kwon, J. S. (2009). Cortical network dynamics during source memory retrieval: current density imaging with individual MRI. *Human Brain Mapping, 30*(1), 78-91.

Kok, A., Ramautar, J. R., De Ruiter, M. B., Band, G. P., & Ridderinkhof, K. R. (2004). ERP components associated with successful and unsuccessful stopping in a stop-signal task. *Psychophysiology, 41*(1), 9-20.

Kong, J., Gollub, R., Rosman, I., Webb, M., Vangel, M., Kirsch, I., et al. (2006). Brain Activity Associated with Expectancy-Enhanced Placebo Analgesia as Measured by Functional Magnetic Resonance Imaging. *J. Neurosci., 26*(2), 381-388.

Konishi S, Nakajima K, Uchida I, Sekihara K, Miyashita Y (1998) No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *Eur J Neurosci* 10:1209-1213.

Konishi, S., Donaldson, D. I., & Buckner, R. L. (2001). Transient activation during block transition. *NeuroImage, 13*(2), 364-374.

Konishi, S., Nakajima, K., Uchida, I., Sekihara, K., & Miyashita, Y. (1998). No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. *The European Journal of Neuroscience, 10*(3), 1209-1213.

Kramer, U. M., Cunillera, T., Camara, E., Marco-Pallares, J., Cucurell, D., Nager, W., Bauer, P., Schule, R., Schols, L., Rodriguez-Fornells, A., & Nunte, T. F. (2007). The impact of catechol-0-methyltransferase and dopamine D4 receptor genotypes on neurophysiological markers of performance monitoring. *The Journal of Neuroscience*, 27, 14190–14198.

Krikorian, R., Zimmerman, M. E., & Fleck, D. E. (2004). Inhibitory control in Obsessive-Compulsive Disorder. *Brain and cognition, 54*(3), 257-259.

Kumaran, D., & Maguire, E. A. (2007). Which computational mechanisms operate in the

hippocampus during novelty detection? *Hippocampus*. 17(9):735-48.

Lansbergen, M., Bocker, K., Bekker, E., & Kenemans, J. (2007). Neural correlates of stopping and self-reported impulsivity. *Clinical Neurophysiology, 118*(9), 2089-2103.

Lee, S., & Dapretto, M. (2006). Metaphorical vs. literal word meanings: fMRI evidence against a selective role of the right hemisphere. *NeuroImage, 29*(2), 536-544.

Leotti LA, Wager TD. (2009) Motivational influences on response inhibition measures. *JEP:HPP*, 36: 430-47.

Li, C.-S. R., Yan, P., Bergquist, K. L., & Sinha, R. (2007). Greater activation of the "default" brain regions predicts stop signal errors. *Neuroimage*.

Liddle EB, Scerif G, Hollis CP, Batty MJ, Groom MJ, Liotti M, Liddle PF. (2009) Looking before you leap: a theory of motivated control of action. *Cognition* 112:141–158.

Liddle, P. F., Kiehl, K. A., & Smith, A. M. (2001). Event-related fMRI study of response inhibition. *Human Brain Mapping, 12*(2), 100-109.

Lieberman, M. D., Eisenberger, N. I., Crockett, M. J., Tom, S. M., Pfeifer, J. H., & Way, B. M. (2007). Putting feelings into words: affect labeling disrupts amygdala activity in response to affective stimuli. *Psychological Science, 18*(5), 421-428.

Lieberman, M. D., Hariri, A., Jarcho, J., Eisenberger, N., & Bookheimer, S. (2005). An fMRI investigation of race-related amygdala activity in African-American and Caucasian-American individuals. *Nature Neuroscience, 8*(6), 720-722.

Lim SL, Padmala S, Pessoa L (2009) Segregating the significant from the mundane on a moment-to-moment basis via direct and indirect amygdala contributions *Proc Natl Acad Sci U S A* 106:16841-16846.

Linden, D. E. (2005). The P3: where in the brain is it produced and what does it tell us? *The Neuroscientist, 11*(6), 563-576.

Linden, D. E., Prvulovic, D., Formisano, E., Vallinger, M., Zanella, F. E., Goebel, R., et al. (1999). The functional neuroanatomy of target detection: an fMRI study of visual and auditory oddball tasks. *Cereb Cortex, 9*(8), 815-823.

Logan GD, Cowan, WB (1984) On the ability to inhibit thought and action: A theory of an act of control. *Psych Review* 91:295-327.

Lovero KL, Simmons AN, Aron JL, Paulus MP (2009) Anterior insular cortex anticipates

impending stimulus significance. *Neuroimage* 45:976-83.

MacLeod CM, Dodd MD, Sheard ED, Wilson DE, Bibi, U (2003) In Opposition to Inhibition. In: B H Ross, editor. *The Psychology of Learning and Motivation*, San Diego: Academic Press, pp. 163-214.

Maril, A., Wagner, A. D., & Schacter, D. L. (2001). On the tip of the tongue: an event-related fMRI study of semantic retrieval failure and cognitive conflict. *Neuron, 31*(4), 653-660.

Marois, R., Chun, M. M., & Gore, J. C. (2000). Neural correlates of the attentional blink. *Neuron, 28*(1), 299-308.

Mars, R. B., Piekema, C., Coles, M. G., Hulstijn, W., & Toni, I. (2007). On the programming and reprogramming of actions. *Cerebral cortex (New York, N.Y. : 1991), 17*(12), 2972-2979.

Marsh, A. A., Blair, K. S., Jones, M. M., Soliman, N., & Blair, R. J. R. (2008). Dominance and Submission: The Ventrolateral Prefrontal Cortex and Responses to Status Cues. *Journal of Cognitive Neuroscience*.

Martens, S., Munneke, J., Smid, H., & Johnson, A. (2006). Quick Minds Don't Blink: Electrophysiological Correlates of Individual Differences in Attentional Selection. *J. Cogn. Neurosci., 18*(9), 1423-1438.

Mashal, N., Faust, M., Hendler, T., & Jung-Beeman, M. (2007). An fMRI investigation of the neural correlates underlying the processing of novel metaphoric expressions. *Brain and Language, 100*(2), 115-126.

Mayr, U., & Kliegl, R. (2000). Task-set switching and long-term memory retrieval. *JEP:LMC, 26*(5), 1124-1140.

Mazzola-Pomietto, P., Kaladjian, A., Azorin, J. M., Anton, J. L., & Jeanningros, R. (2009). Bilateral decrease in ventrolateral prefrontal cortex activation during motor response inhibition in mania. *Journal of Psychiatric Research, 43*(4), 432-441.

Melcher, T. & Gruber, O. (2006). Oddball and incongruity effects during Stroop task performance: A comparative fMRI study on selective attention. *Brain Research*, *1121*(1), 136-149.

Menenti, L., Petersson, K. M., Scheeringa, R., & Hagoort, P. (2008). When Elephants Fly: Differential Sensitivity of Right and Left Inferior Frontal Gyri to Discourse and World Knowledge. *Journal of Cognitive Neuroscience*. 21(12):2358-68.

Michael GA, Garcia S, Fernandez D, Sellal F, Boucart M (2006) The ventral premotor cortex

(vPM) and resistance to interference. *Behav Neurosci* 120:447–62.

Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. *Ann Rev Neurosci*, 24:167-202.

Miyake, A. *et al*. (2000) The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. *Cognit Psychol*. 41:49-100

Morault, P. M., Bourgeois, M., Laville, J., Bensch, C., & Paty, J. (1997). Psychophysiological and clinical value of event-related potentials in obsessive-compulsive disorder. *Biological Psychiatry, 42*(1), 46-56.

Morein-Zamir, S., Chua, R., Franks, I., Nagelkerke, P., & Kingstone, A. (2007). Predictability influences stopping and response control. *J Exp Psychol Hum Percept Perform, 33*(1), 149-162.

Morgan, M. D., Cranford, J. L., & Burk, K. (1997). P300 event-related potentials in stutterers and nonstutterers. *JSLHR, 40*(6), 1334-1340.

Morimoto, H. M., Hirose, S., Chikazoe, J., Jimura, K., Asari, T., Yamashita, K.-I., et al. (2008). On Verbal/Nonverbal Modality Dependence of Left and Right Inferior Prefrontal Activation during Performance of Flanker Interference Task. *Journal of Cognitive Neuroscience*. 20(11):2006-14.

Mostofsky, S. H., Schafer, J. G., Abrams, M. T., Goldberg, M. C., Flower, A. A., Boyce, A., et al. (2003). fMRI evidence that the neural basis of response inhibition is task-dependent. *Brain Res Cogn Brain Res, 17*(2), 419-430.

Mulert, C., JÃ¤ger, L., Schmitt, R., Bussfeld, P., Pogarell, O., MÃ¶ller, H.-J. r., et al. (2004). Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. *NeuroImage, 22*(1), 83-94.

Munakata, Y. (2001). Graded representations in behavioral dissociations. *Trends in Cognitive Sciences, 5*(7), 309-315.

Naeser, M., Martin, P., Nicholas, M., Baker, E., Seekins, H., Kobayashi, M., et al. (2005). Improved picture naming in chronic aphasia after TMS to part of right Broca's area: An open-protocol study. *Brain and Language, 93*(1), 95-105.

Nambu A, Tokuno H, Inase M, Takada M (1997) Corticosubthalamic input zones from forelimb representations of the dorsal and ventral divisions of the premotor cortex in the macaque monkey: comparison with the input zones from the primary motor cortex and the supplementary motor area. *Neurosci Lett* 239:13-16.

Nelder JA, Mead RA (1965) Simplex method for function minimization. *Comp J* 7 :308–313.

Neubert FX, Mars RB, Buch ER, Olivier E, Rushworth MF (2010) Cortical and subcortical interactions during action reprogramming and their related white matter pathways. *Proc Natl Acad Sci U S A* 107:13240-5.

Nieuwenhuis S, De Geus EJ, Aston-Jones, G (2011). The anatomical and functional relationship between the P3 and autonomic components of the orienting response. *Psychophysiology* 48: 162-175.

Nieuwenhuis S, Yeung N, van den Wildenberg W & Ridderinkhof KR (2003) Electrophysiological correlates of anterior cingulate function in a go/no-Go Task: effects of response conflict and trial type frequency. *Cognit, Affective, and Behav Neurosci* 3:17–26.

Nigg, J. T. (2000). On Inhibition/Disinhibition in Developmental Psychopathology: Views From Cognitive and Personality Psychology and a Working Inhibition Taxonomy. *Psychological Bulletin*, *126*(2), 220-246.

Nigg, J. T., Silk, K. R., Stavro, G., & Miller, T. (2005). Disinhibition and borderline personality disorder. *Development and psychopathology, 17*(4), 1129-1149.

Norman, K. A., Newman, E. L., & Detre, G. (2007). A neural network model of retrieval-induced forgetting. *Psychol Rev, 114*(4), 887-953.

Nyberg, L., McIntosh, A. R., Cabeza, R., Habib, R., Houle, S., & Tulving, E. (1996). General and specific brain regions involved in encoding and retrieval of events: what, where, and when. *Proc Natl Acad Sci U S A, 93*(20), 11280-11285.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D. Y., & Schroger, E. (2002). Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. *NeuroImage, 15*(1), 167-174.

O'Reilly RC (2010) The What and How of prefrontal cortical organization. *Trends in Neurosciences*, 33:355-361.

O'Reilly, R.C., & Frank, M. (2006). Making Working Memory Work: A Computational Model of Learning in the Prefrontal Cortex and Basal Ganglia. *Neural Computation, 18*(2), 283-328.

Patil PG, Carmena JM, Nicolelis, MA & Turner DA. (2004) Ensemble recordings of human subcortical neurons as a source of motor control signals for a brain-machine interface. *Neurosurgery* 55:27-35.

Petrides, M. (2005). Lateral prefrontal cortex: architectonic and functional organization.

Philosophical Transactions of the Royal Society B, 360(1456), 781-795.

Peyron, R., Garcia-Larrea, L., Gragoire, M. C., Costes, N., Convers, P., Lavenne, F., et al. (1999). Haemodynamic brain responses to acute pain in humans: sensory and attentional networks. *Brain, 122 (Pt 9)*, 1765-1780.

Pfefferbaum A, Ford JM, Weller BJ & Kopell BS (1985) ERPs to response production and inhibition. *Electroenceph & Clin Neurophysiol*, 60:423-434.

Polich, J., & Criado, J. (2006). Neuropsychology and neuropharmacology of P3a and P3b. *International Journal of Psychophysiology, 60*(2), 172-185.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P3: an integrative review. *Biol Psychol, 41*(2), 103-146.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biol Psychol, 41*(2), 103-146.

Postma, A., & Kolk, H. (1992). The effects of noise masking and required accuracy on speech errors, disfluencies, and self-repairs. *Journal of Speech and Hearing Research, 35*(3), 537-544.

Potter, A. S., & Newhouse, P. A. (2008). Acute nicotine improves cognitive deficits in young adults with attention-deficit/hyperactivity disorder. *Pharmacol Biochem Behav, 88*(4), 407-417.

Preibisch, C., Neumann, K., Raab, P., Euler, H. A., von Gudenberg, A. W., Lanfermann, H., et al. (2003). Evidence for compensation for stuttering by the right frontal operculum. *NeuroImage, 20*(2), 1356-1364.

Ramautar JR, Kok A & Ridderinkhof K (2006). Effects of stop-signal modality on the N2/P3 complex elicited in the stop-signal paradigm. *Biol Psychol*. 72:96-109

Ramautar, J. R., Kok, A., & Ridderinkhof, K. R. (2004). Effects of stop-signal probability in the stop-signal paradigm: the N2/P3 complex further validated. *Brain Cogn, 56*(2), 234-252.

Ramautar, J. R., Slagter, H. A., Kok, A., & Ridderinkhof, K. R. (2006). Probability effects in the stop-signal paradigm: the insula and the significance of failed inhibition. *Brain Res, 1105*(1), 143-154.

Ramos, B.P., Arnsten, A.F., 2007. Adrenergic pharmacology and cognition: focus on the prefrontal cortex. *Pharmacology & Therapeutics*, *113*(3), 523–536.

Ranganath, C., & Rainer, G. (2003). Neural mechanisms for detecting and remembering novel events. *Nat Rev Neurosci, 4*(3), 193-202.

Ratcliff R, Starns JJ (2009) Modeling confidence and response time in recognition memory. *Psychol Rev* 116:59-83.

Ray, Huang, C., Constable, T., & Sinha, R. (2006). Imaging Response Inhibition in a Stop-Signal Task: Neural Correlates Independent of Signal Monitoring and Post-Response Processing. *J. Neurosci., 26*(1), 186-192.

Reynolds, J. R., Braver, T., Brown, J., & Van der Stigchel, S. (2006). Computational and neural mechanisms of task switching. *Neurocomputing, 69*(10-12), 1332-1336.

Reynolds, J. R., West, R., & Braver, T. (2008). Distinct Neural Circuits Support Transient and Sustained Processes in Prospective Memory and Working Memory. *Cerebral Cortex*. 19(5): 1208–1221.

Ridderinkhof, K. R., Scheres, A., Oosterlaan, J., & Sergeant, J. A. (2005). Delta plots in the study of individual differences: new tools reveal response inhibition deficits in AD/Hd that are eliminated by methylphenidate treatment. *Journal of Abnormal Psychology, 114*(2), 197-215.

Roberts, A. C., & Wallis, J. D. (2000). Inhibitory control and affective processing in the prefrontal cortex: neuropsychological studies in the common marmoset. *Cereb Cortex, 10*(3), 252-262.

Robertson, I., Mattingley, J., Rorden, C., & Driver, J. (1998). Phasic alerting of neglect patients overcomes their spatial deficit in visual awareness. *Nature, 395*(6698), 169-172.

Rubia, K., Russell, T., Overmeyer, S., Brammer, M. J., Bullmore, E. T., Sharma, T., et al. (2001). Mapping motor inhibition: conjunctive brain activations across different versions of go/no-go and stop tasks. *NeuroImage, 13*(2), 250-261.

Rubia, K., Smith, A. B., Brammer, M. J., & Taylor, E. (2003). Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. *Neuroimage, 20*(1), 351-358.

Sakagami, M., Tsutsui, K., Lauwereyns, J., Koizumi, M., Kobayashi, S., & Hikosaka, O. (2001). A code for behavioral inhibition on the basis of color, but not motion, in ventrolateral prefrontal cortex of macaque monkey. *The Journal of Neuroscience, 21*(13), 4801-4808.

Salisbury DF, Griggs CB, Shenton ME & McCarley RW. (2004) The NoGo P300 'anteriorization' effect and response inhibition. *Clin Neurophysiology* 115:1550-58.

Salomons, T. V., Johnstone, T., Backonja, M. M., Shackman, A. J., & Davidson, R. J. (2007). Individual differences in the effects of perceived controllability on pain perception: critical role of the prefrontal cortex. *Journal of Cognitive Neuroscience, 19*(6), 993-1003.

Sangal, R. B., & Sangal, J. M. (2006). Attention-deficit/hyperactivity disorder: use of cognitive evoked potential (P300) to predict treatment response. *Clinical neurophysiology : official journal of the International Federation of Clinical Neurophysiology, 117*(9), 1996-2006.

Sangal, R. B., Sangal, J. M., & Belisle, C. (1999). Longer auditory and visual P300 latencies in patients with narcolepsy. *Clinical EEG (electroencephalography), 30*(1), 28-32.

Sauleau P *et al*. (2009) Involvement of the subthalamic nucleus in engagement with behaviourally relevant stimuli. *Eur J Neurosci*. 29:931-42 .

Shallice, T., Stuss, D., Alexander, M., Picton, T., & Derkzen, D. (2008). The multiple dimensions of sustained attention. *Cortex, 44*(7), 794-805.

Sharp DJ *et al*. (2010) Distinct frontal systems for response inhibition, attentional capture, and error processing. PNAS, 107:6106-11.

Simmonds, D. J., Pekar, J. J., & Mostofsky, S. H. (2008). Meta-analysis of Go/No-go tasks demonstrating that fMRI activation associated with response inhibition is task-dependent. *Neuropsychologia, 46*(1), 224-232.

Simons, J. S., Schlövinck, M. L., Gilbert, S. J., Frith, C. D., & Burgess, P. W. (2006). Differential components of prospective memory? Evidence from fMRI. *Neuropsychologia, 44*(8), 1388-1397.

Smith JL, Johnstone SJ, Barry RJ (2008) Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. *Clin Neurophysiol* 119:704-14.

Smith JL, Smith EA, Provost AL & Heathcote, A (2010) Sequence effects support the conflict theory of N2 and P3 in the Go/No Go Tak. *Int J Psychophysiol*. 75:217-26.

Smith, R. E. (2003). The cost of remembering to remember in event-based prospective memory: investigating the capacity demands of delayed intention performance. *JEP:LMC, 29*(3), 347-361.

Spencer, K. M., Dien, J., & Donchin, E. (2001). Spatiotemporal analysis of the late ERP responses to deviant stimuli. *Psychophysiology, 38*(2), 343-358.

Stevens, A. (2000). Event-related fMRI of auditory and visual oddball tasks. *Magnetic Resonance Imaging, 18*(5), 495-502.

Stevens, D., Hasher, L., Chiew, K., & Grady, C. (2008). A Neural Mechanism Underlying Memory Failure in Older Adults. *J. Neurosci., 28*(48), 12820-12824.

Strik WK, Fallgatter AJ, Brandeis D & Pascual-Marqui RD (1998) Three-dimensional tomography of event-related potentials during response inhibition: evidence for phasic frontal

lobe activation. *Electroencephalogr Clin Neurophysiol*. 108:406-13.

Stringaris, A. K., Medford, N., Giora, R., Giampietro, V. C., Brammer, M. J., & David, A. S. (2006). How metaphors influence semantic relatedness judgments: the role of the right frontal cortex. *NeuroImage, 33*(2), 784-793.

Sturm, W., Willmes, K., Orgass, B., & Hartje, W. (1997). Do Specific Attention Deficits Need Specific Training? *Neuropsychological Rehabilitation*, 81-103.

Stuss DT, Alexander MP (2007) Is there a dysexective syndrome? *Philos Trans R Soc Lond B Biol Sci* 362:901-915.

Stuss, D. T., & Alexander, M. P. (2007). Is there a dysexecutive syndrome? *Philos Trans R Soc Lond B Biol Sci, 362*(1481), 901-915.

Symonds, L. L., Gordon, N. S., Bixby, J. C., & Mande, M. M. (2006). Right-lateralized pain processing in the human cortex: an FMRI study. *Journal of Neurophysiology, 95*(6), 3823-3830.

Thimm, M., Fink, G. R., Küst, J., Karbe, H., & Sturm, W. (2006). Impact of alertness training on spatial neglect: a behavioural and fMRI study. *Neuropsychologia, 44*(7), 1230-1246.

Thorell LB, Lindqvist S, Nutley SB, Bohlin G, Klingberg T (2009) Training and transfer effects of executive functions in preschool children. *Dev Sci* 12:106-113.

Tieges, Z., Snel, J., Kok, A., & Richard Ridderinkhof, K. (2009). Caffeine does not modulate inhibitory control. *Brain and Cognition, 69*(2), 316-327.

Towse, J., Lewis, C., & Knowles, M. (2007). When knowledge is not enough: The phenomenon of goal neglect in preschool children. *Journal of Experimental Child Psychology, 96*(4), 320-332.

Tucker DM (1993) Spatial sampling of head electrical fields: The geodesic sensor net. *Electroenceph & Clinical Neurophys* 87:154–16.

Turner, D. C., Clark, L., Dowson, J., Robbins, T. W., Sahakian, B. J. (2004). Modafinil improves cognition and response inhibition in adult attention-deficit/hyperactivity disorder. *Biological Psychiatry, 55*(10), 1031-1040.

van Gaal S, Ridderinkhof KR, Fahrenfort JJ, Scholte HS. & Lamme VA (2008) Frontal cortex mediates unconsciously triggered inhibitory control. *J Neurosci* 28:8053-8062.

van Gaal S, Ridderinkhof KR, Scholte HS, Lamme VA (2010) Unconscious activation of the prefrontal no-go network. *J Neurosci* 30:4143-4150..

Vasic, N., & Wijnen, F. (2005). Stuttering as a monitoring deficit. In R. J. Hartsuiker, R. Bastiaanse & A. Postma (Eds.), *Phonological encoding and monitoring in normal and pathological speech.* Hove (East Sussex): Psychology Press.

Verbruggen F, & Logan, GD (2009). Proactive adjustments of response strategies in the stop-signal paradigm. *JEP:HPP*, 35, 835-854.

Verbruggen F, Aron AR, Stevens MA, Chambers CD (2010) Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. *Proc Natl Acad Sci U S A* 107:13966-71.

Verbruggen F, Liefooghe, B; Szmalec, A; Vandierendonck,, A (2005) Inhibiting Responses When Switching: Does it Matter? *JEP:G* 52:125-130.

Verbruggen V, Logan, G (2009). Automatic and Controlled Response Inhibition: Associative Learning in the Go/No-Go and Stop-Signal Paradigms. *JEP:G* 137:649-672.

Verbruggen, F., & Logan, G. (2008). Models of response inhibition in the stop-signal and stop-change paradigms. *Neuroscience & Biobehavioral Reviews*. 33, 647-661.

Vossel, S., Weidner, R., Thiel, C. M., & Fink, G. R. (2008). What is "Odd" in Posner's Location-cueing Paradigm? Neural Responses to Unexpected Location and Feature Changes Compared. *Journal of cognitive neuroscience*. 21(1):30-41.

Whitehead, R. (1991). Right hemisphere processing superiority during sustained visual attention. *J. Cognitive Neuroscience, 3*(4), 329-334.

Winsberg, B. G., Javitt, D. C., Silipo, G. S., & Doneshka, P. (1993). Mismatch negativity in hyperactive children: effects of methylphenidate. *Psychopharmacology bulletin, 29*(2), 229-233.

Xue, G., Aron, A., & Poldrack, R. (2008). Common Neural Substrates for Inhibition of Spoken and Manual Responses. *Cereb. Cortex, 18*(8), 1923-1932.

Zaslansky, R., Sprecher, E., Tenke, C. E., Hemli, J. A., & Yarnitsky, D. (1996). The P3 in pain evoked potentials. *Pain, 66*(1), 39-49.

APPENDIX I. SUPPORTING INFORMATION FOR CHAPTER 2

This appendix includes:

Supporting Figures 1-6

Supporting Tables 1-5

Supporting Methods and Results

Supporting Discussion

Supporting References

Supporting Methods: Instructions to Subjects

Instructions to subjects were essentially identical across all experiments. For the Double Go Task in Experiment 1, subjects were told, "In this task you will respond TWICE to arrows that are subsequently covered up by a square. AS QUICKLY AS POSSIBLE, press the z button to left arrows, and the m button to right arrows. If a square appears, press the same button again. Always anticipate the appearance of the square by PREPARING to press the button twice. You MUST respond to squares very quickly - if you're too slow, the square will turn red, and the trial will be marked as incorrect. Keep your fingers on the keys throughout all the tasks. If you have any questions please ask them now." The instructions for the Double Go Task differed in Experiment 2 in that subject were told to expect frequent "blink breaks," but should attempt to blink as little as possible throughout the rest of the task. The instructions for the Double Go Task also differed in Experiment 3, in that a MR-compatible button box was used for responding; thus "z button" was replaced with "leftmost button" and "m button" was replaced with "rightmost button" in the instructions. Instructions for Experiment 2 and 3 were otherwise identical to those in Experiment 1. The stop task instructions were as follows in Experiment 1, with changes throughout subsequent experiments that were analogous to those just described for the Double Go Task: "Now you will perform the opposite task. Your task is to respond to the arrows UNLESS the square appears, but you should not wait for the square. We are interested both in how fast you can respond and how well you can stop - both are equally important. Keep your fingers on the buttons so that you can respond to the arrows as soon as they appear. If you see the square, try to STOP yourself from responding. This will sometimes be physically

impossible, but try your best. If you fail to stop, the square will turn red. If you have any questions please ask them now."

We note that while subjects were not told to prepare to stop on all trials within the Stop task (consistent with the standard instructions for that task), subjects *were* told to prepare to respond twice on all trials within the Double Go task. This particular procedure was adopted based on results from extensive piloting of the Double Go task, which indicated that subjects were unlikely to proactively prepare their responses without such instructions (as reflected in unacceptably long RTs on Signal trials, and a large number of omission errors). This stands in stark contrast to results from the Stop task, which indicate that subjects engage proactive control even in the absence of instructions to do so [1-5]. This difference in instructions could thus be reasonably expected to bring the strategies of the two tasks into alignment. Nonetheless, the role of instructions and strategic responding are important directions for future work, given that departures from standard instructions for the Stop task (although such departures did not occur here) are known to influence performance [6].

Supporting Methods: Behavioral Task Design

As mentioned in the main text, a fixed task order was adopted for four reasons. First, had tasks been administered in the opposite order, subjects might have associated the infrequent stimulus with the act of stopping in the Stop Task and then potentially recruited stopping processes during the Double Go Task to overcome this prepotent association. Second, the current task order may maximize our ability to detect effortful response inhibition processes, because the act of responding should be more prepotent following the Double Go Task. Third, previous work using

a counterbalanced design demonstrated that the univariate hemodynamics of the rVLPFC are similar during stopping and monitoring tasks [7]; because we observed this similarity as well as similarities in multivariate hemodynamics, electrophysiology, pupillometry, and behavior, use of a fixed task order is unlikely to have influenced the results reported here. Fourth, although not as standard in purely experimental work, the use of fixed task orders is a central tenet of individual differences studies [8-10], because counterbalancing can introduce substantial noise into individual difference correlations. Thus the use of a fixed task order allows us to substantially expand upon previous work by providing increased sensitivity to individual differences influencing both stopping and monitoring processes.

Concerns about the impact of task order on our effects are largely addressed by a very recent study which has used intermixed or counterbalanced designs to examine the similarity between a Double Go or Single Go task and the Stop task. Owing to their intermixed designs, these studies are clearly not subject to order effects, and yet reveal results that are conceptually identical to our own [11-12]. In particular, the latter of these studies utilizes the same Double Go task we use here, and finds increased activation to Double Go_{Signal} trials than to $Stop_{Signal}$ trials, which is entirely consistent with our group-level fMRI results. Thus it seems that any order effects would not change our conceptual conclusions or bear on these group analyses of our manuscript.

Pupillometry and event-related potentials have not been previously investigated on Double Go trials intermixed with Stop Signal trials; our experiments are the first to use these methods within this task. Thus any arguments about order effects on pupillometry or event-related potentials are necessarily speculative. However, one might expect intermixed designs to also yield: 1) an increased "Stop P3" on Double Go Signal trials, to the extent that the associated increased hemodynamic response in rVLPFC is linked to the P3 [13], and in turn, 2) an increased pupillary

response on Double Go Signal trials, to the extent that the P3 and pupillary response are both driven by the LC/norepi system [14-15].

Also as mentioned in the main text, there were minor variations across experiments in the precise design of stimuli, inter-trial intervals, and other characteristics; the visual differences are illustrated in Supporting Fig. 1, and other differences are described in Supporting Table 1.

Supporting Methods: fMRI Acquisition and Regression Model

Functional data were collected in a two runs of 216 EPI volumes, each consisting of 32 4 mm thick slices (gap=0 mm, field-of-view (FOV)=220 mm, in-plane matrix= 64 x 64, in-plane resolution= 3.44 x 3.44 mm2), angled parallel to the AC-PC line. Prior to the functional runs, high-resolution T1-weighted 3D IR-SPGR full head anatomical images were acquired along the coronal plane (TR=9 ms, TE=2 ms, flip angle=10º, inversion time=500 ms; 220 mm FOV, 256 x 256 matrix, 0.87 mm x 0.87 mm in-plane resolution, 124 slices, 1.7-mm slice thickness). The scanner was equipped with a standard head coil and participant's heads were secured with moldable pillows to minimize head motion. Stimuli were displayed through fiber-optic goggles and participants responding by pressing one of two buttons on a fiber-optic button box. Customized square waveforms were generated for each participant and run. Separate waveforms were generated for No Signal trials with left and right pointing arrows, for Signal trials in which the subject had been correct and incorrect, and for task blocks. In addition, we generated several control waveforms to control for nuisance variables; this included parametric waveforms generated for the number of TRs since the last signal trial and for the interstimulus delay with which each signal was presented, as well square waveforms for cues indicating the onset and

offset of task blocks. Finally, a waveform was generated specifying the onset and duration of task blocks. These waveforms were convolved with a double gamma hemodynamic response function (HRF). For each participant, we used FILM (FMRIB's Improved Linear Model) to estimate the hemodynamic parameters for the different explanatory variables (EVs; e.g. one for each of the separate waveforms) and generate statistical contrast maps of interest (e.g. a contrast between Signal and No Signal trials), with separate models fit to each task. As with the ERP and pupillometry analyses, this procedure ensures the task-evoked hemodynamic response is calculated relative to within-task baselines, thereby controlling for non-specific task differences such as scanner drift or fatigue.

Supporting Results: univariate fMRI analyses

In addition to the contrasts reported in the main text, we also contrasted Double $Go_{No\text{-}Signal}$ and $Stop_{No\text{-}Signal}$ directly, to confirm that any differences between those trial types would not contaminate the contrast of (Double Go_{Signal} -Double $Go_{No\text{-}Signal}$) vs. ($Stop_{Signal}$ - $Stop_{No\text{-}Signal}$). No significant differences were observed between Double $Go_{No\text{-}Signal}$ and $Stop_{No\text{-}Signal}$ trials.

Supporting Results: fMRI Pattern Classification analyses

Z-transformed beta-weights from the estimation routines implemented by FSL were provided to the network as inputs, with one input unit per voxel. Each unit in this input layer projected to a distinct set of 30 units designed to encode real valued inputs as a distributed pattern (a ScalarVal layer in Emergent). Thus, unit #1 would code for the value -4.5, unit #30 would code for the

value 4.5, and other units interpolate between those values. These patterns were smoothed across adjacent units with a Gaussian smoothing kernel of σ = .105. For example, this smoothing allows unit #1 to code for the value -4.5 but also -4.4 and -4.3, with less activation resulting from values that diverged more from the unit's preferred value. Next, each of these ScalarVal units were fully connected with random weights to all output units, of which there were 18 for classifying individuals (1 per subject; Supporting Fig. 2A) or 2 for classifying trial types (Signal trials & No-Signal trials; Supporting Fig.2B).

As mentioned in the main text, separate networks were then trained for each ROI (and therefore differed in terms of the number of input units), and for identifying which individuals generated the data vs. what trial type the data was estimated from (and therefore differed in terms of the number of output units) but all other aspects of the network architecture were the same. Specifically, all ScalarVal layers contained 30 times the number of input units for that network, and all network connection weights were adjusted via Hebbian and contrastive Hebbian learning rules, with the same mixtures (k_{hebb} = .01) and learning rates (0.05). The equation for the Hebbian weight change is:

$$\Delta_{hebb} w_{ij} = x_i^+ y_j^+ - y_j^+ w_{ij} = y_j^+ (x_i^+ - w_{ij}) \quad (2)$$

and for contrastive Hebbian learning:

$$\Delta_{err} w_{ij} = \left(x_i^+ y_j^+\right) - \left(x_i^- y_j^-\right) \quad (3)$$

which is subject to a soft-weight bounding to keep within the 0 − 1 range:

$$\Delta_{sberr} w_{ij} = [\Delta_{err}] + \left(1 - w_{ij}\right) + [\Delta_{err}] - w_{ij} \quad (4)$$

The two terms are then combined additively with a normalized mixing constant k_{hebb}:

$$\Delta w_{ij} = \in [k_{hebb}(\Delta_{hebb}) + (1 - k_{hebb})(\Delta_{sberr})] \quad (5)$$

For classifying individuals, we trained a set of 40 networks (10 each for BA44, 45, 47 and 4) to activate one of 18 output units corresponding directly to which of our 18 subjects generated the z-transformed beta weight input values from the contrast Signal > Null in the Double Go Task. We trained a separate set of 40 networks to do the same for the contrast No-Signal > Null in the Double Go Task. For classifying trial types, we trained a set of 180 networks (10 for each subject) to classify which trial type the z-transformed beta weight input values came from in the Double Go Task: either Signal>Null or No-Signal>Null, using two output units corresponding directly to these two contrasts. For classifying trial types, a separate set of 180 networks was trained for each ROI (44, 45, 47 and 4). Once all networks had performed correctly on 4 successive epochs of testing, the learning rate was turned to zero and networks were presented with data from the corresponding individuals, ROIs, and contrasts in the Stop task. Networks were scored as performing at chance on any trial where they activated all output units equally, or as performing correctly if the correct output unit was the most active. These performance scores were then averaged to yield the data presented in main text Fig. 4.

These analyses require that networks generalize not only across tasks but also across runs, because the tasks were collected in separate runs. Although we excluded subjects for

excessive motion, some motion artifacts likely contribute to the below-perfect generalization that we observed across tasks.

It is also possible to train the same networks to discriminate the tasks given a particular ROI and contrast, but that discrimination could be explained as a function of the noise that is not specific to our tasks, but rather specific to the separate runs in which the tasks were collected. To demonstrate this, we trained separate sets of 10 networks to discriminate tasks based on activation in each ROI for each individual on odd-numbered trials, and tested them on even-numbered trials, for three separate contrasts: Signal>Null, No-Signal>Null, and Nuisance>Null, where nuisance trials were those where subjects saw the word "RELAX" for 2 seconds. We found that discrimination of tasks/runs (since these are confounded in our data) across these 2160 networks (10 runs of each network x 18 subjects x 4 ROIs x 3 contrasts) was reliably above chance even for the contrast Nuisance>Null (with mean accuracy of 65%), indicating that run-specific rather than task-specific variance is contributing to the classifier's accuracy. A separate batch of networks was trained to distinguish the tasks based on activity in bilateral primary visual cortex (BA17), which proved to be the best ROI for distinguishing the tasks on average ($F(1,17)=8.462$, $p=.01$) and on the Signal > Null contrast in particular Supporting($F(1,17)=17.42$, $p=.001$; Supporting Figure 3). Thus, the tasks/runs can be discriminated based on nuisance trials; furthermore, BA17 is the most successful at discriminating tasks in general and on the critical Signal > Null contrast in particular. These results indicate that unambiguous inferences about multivariate patterns that discriminate stopping and context monitoring processes cannot be made when infrequent Go and Stop trials are collected in separate runs. Intermixing those trial types into the same run is likely to introduce other problems, such as the strategic

prioritization of stopping demands[16], which is one reason we did not adopt this design ourselves.

Supporting Results: ERP Analyses

As described in the main text, frontal correlations increased more than occipital correlations following the onset of the signal, relative to changes in correlations that otherwise happen at the same time (i.e., during No Signal trials). Main text Fig. 3C depicts the most focused contrast, which is this three-way interaction of Trial Type (Signal vs. No-Signal) x Montage (frontal vs. occipital) x Time (before vs. after signal onset) ($F(1,98)=12.59$, $p=.001$). This three-way interaction is significant both with and without baseline correction of the ERPs, indicating that individual differences in baselines are not driving the effect, and that the influence of any non-specific task effect (e.g., fatigue) on these ERPs is minimal.

Also as described in the main text, stimuli that demand stopping typically elicit a positive-going potential (i.e., the Stop P3) that is frontally enhanced relative to the potentials on trials that require response commission[17-22]. This "anteriorization" effect is so robust across response inhibition paradigms, including the Stop and Go/NoGo Tasks, that some have argued such anteriorization directly indexes response suppression[23]. Thus, one strong prediction of stopping accounts is that the distribution of the P3 elicited by $Stop_{Signal}$ trials should be more anterior than the P3 elicited by Double Go_{Signal} trials. However, and in direct contradiction to this prediction, the P3 elicited on $Stop_{Signal}$ trials was enhanced relative to Double Go_{Signal} trials only at more posterior electrodes, a significantly different pattern than observed over more

anterior electrodes (Cz&Pz vs. Fz: $F(1,34)=17.81$, $p<.0005$, Supporting Fig. S4; see also Supporting Fig. S5). As such, our results actually demonstrate "centralization" of the P3 on $Stop_{Signal}$ trials – suggesting that the anteriorization effect typically observed in the electrophysiology of stopping should not be taken to directly index response inhibition, but rather a more general monitoring process shared across the "signal" trials of our tasks.

The centralized distribution more often characterizes another component typically elicited by stopping tasks: an ERP with a slightly earlier and negative-going central potential known as the N2. This ERP, unlike the Stop P3, has already been previously shown to be functionally non-specific to response inhibition, but is instead thought to reflect the detection of response conflict[24-27]. For this reason, as well as the fact that source localization demonstrates the N2 has a source in the anterior cingulate – not the rVLPFC [28-29] – the N2 is not of primary interest here. Nonetheless, we note that the N2 was marginally enhanced in the Stop task at central (Cz) electrodes ($F(1,34)=4.6$, $p<.05$). In this case, the enhanced N2 may reflect the additional response conflict in the Stop task (where planned responses may have to be unpredictably cancelled) relative to the Double Go Task (where planned responses are always committed).

As described in the main text, the scalp distributions of the two tasks' ERPs were strikingly similar in terms of individual differences (main text Fig. 4). This similarity is clearly visible even in the group average, such that even relatively subtle features of the group average ERPs are matched across both time and space (Supporting Fig. 5). The "centralization" of the P3 during the Stop task, relative to the Double Go Task, is also visible in these group averages – specifically at 300ms after signal onset (highlighted regions of Supporting Fig. 5).

Supporting Methods: Behavioral Analysis (identical across all experiments)

Double Go Task. Because response slowing was observed in the Double Go Task, this slowing could confound obvious measures of the efficiency of context-monitoring. For example, reaction times to the infrequent stimulus might be used as a proxy measure of context-monitoring, such that larger reaction times are interpreted to reflect less efficient detection of the signal. However, these reaction times could in fact be large for a subject who very efficiently detects the signal but is also unusually slowed by it[2]. As mentioned in the main text, we adopted a model-based approach to confront this confound. This model is first introduced conceptually with respect to its core predictions; the underlying mathematics are described next; analyses of the resulting parameter estimates and verifications of the model's core predictions are described last. Conceptually the model is analogous to the race model used to analyze data from the Stop task (Supporting Fig. 6A&B). According to these models, responses cannot be affected until the signal has been detected. Thus, the time of signal detection (TOSD) in the Double Go Task can be understood as the amount of time that must elapse once a signal is presented until responses are affected by slowing. The first parameter to be estimated, therefore, is whether any given trial is "slowed" or "unslowed"; once this has been determined for each trial, we then estimate for each subject the time that must elapse after signal presentation until responses are categorized as

[2] In fact, efficient detection of the signal may lead to more observed slowing precisely because signals will be detected in time to slow responses. This issue confounds model-free estimates of response slowing in tasks with infrequent stimuli that do not demand stopping .

"slowed." The first core prediction of the model is that this measure should positively correlate with SSRT because both measures contain variance related to the efficiency of signal detection. However, these measures may also share variance related to the efficiency of a putative inhibitory or motoric stopping process – a process that gives rise to slowing in the Double Go Task, and to stopping in the Stop task. The efficiency of this putative inhibitory process can be independently estimated in the Double Go Task in terms of the amount of slowing experienced by subjects – informally, we are asking "when subjects are slowed, how slowed are they?" This duration of slowing can be estimated as the difference between Double Go_{Signal} trials categorized as "slowed" and corresponding reaction times on Double $Go_{No\text{-}Signal}$ trials. We hypothesize that subjects who are more slowed by the signal will not tend to show higher SSRT (i.e., a positive correlation), because we predict that SSRT primarily reflects context-monitoring processes. It is possible that subjects that who are more slowed by the signal will in fact have a stronger inhibitory process (i.e., STN-mediated inhibition is more difficult to overcome), and therefore the duration of slowing may negatively correlate with SSRT.

The model's third core prediction is that the positive correlation between TOSD and SSRT should remain when controlling for DoS, because both TOSD and SSRT primarily measure context-monitoring, and do not substantially measure any putative inhibitory processes.

Verifying all three core predictions requires estimating the full model, but the model's basic assumptions can be preliminarily validated using a more straightforward prediction. Responses should be less slowed in the Double Go Task when signals are presented late (because the signals will not be detected in time to slow responses). This prediction was confirmed in our data (Supporting Figure 6C and main text Fig. 6B).

Having verified this preliminary prediction, we move to estimating the full Double Go Task model. As described above, this requires categorizing each trial as slowed or unslowed. To estimate which trials undergo slowing, we adopted a nonparametric technique based on rank order. In particular, we subtracted from each signal trial RT the observation with corresponding percent rank in the no signal trial RT distribution. To the extent that the two distributions are equivalent, these residuals should be centered on zero. As mentioned in the main text, and consistent with our model, many residuals were centered on zero but there was also a pronounced positive skew to the distribution of these residuals, indicative of slowing for some trials. In contrast to the positive skew typically observed in RTs, this skew was sufficiently strong to be essentially unaffected by logarithmic or square-root transformation.

Our mixture modeling approach decomposed the distribution into two underlying distributions: a Gaussian distribution with a mean of zero (corresponding to unslowed RTs), and a Gamma distribution (corresponding to the slowed RTs). Our choice of the Gamma distribution for slowed RTs was motivated by the fact that the two parameters determining the shape of the Gamma (scale and rate) can generate positively skewed pseudo-normal distributions (which would appear to match the observed distribution of skewed residuals) but can also generate exponential distributions, sometimes used to model queue and waiting times – seemingly a good candidate for the functional form of response slowing. The two free parameters to the Gamma and the one free parameter to the Gaussian yielded a better overall fit than a single Gaussian (Supporting Table 2).

Individual RTs were categorized as belonging to the slowed distribution if there was even weak evidence in favor of the RT belonging to that distribution (as quantified by a difference in BIC of ≥ 2.35); otherwise RTs were assigned to the unslowed distribution. We adopted this weak

standard of evidence for classification to ensure that any observation that might have undergone slowing would be classified as such. Other standards of evidence lead to similar results as those presented here, but do not as cleanly separate the slowed and unslowed trials (c.f. main text Fig. 6D, where unslowed Double Go_{Signal} trials show a nonsignificant difference from corresponding Double $Go_{No\text{-}Signal}$ trials.).

Consistent with the Double Go Task model's first core prediction, SSRT and TOSD were positively correlated when collapsing across all three experiments ($R=.418$, $p<.0005$) and in each experiment individually (Supporting Table 3). This positive correlation indicates that SSRT largely reflects the efficiency of context monitoring, as reflected in the time of signal detection. We also did not observe a positive correlation between SSRT and TOSD, consistent with the Double Go Task model's second core prediction. In fact, SSRT and slowing duration were negatively correlated when collapsing across all three experiments ($R=-.188$, $p<.05$), indicating that those with stronger inhibition (i.e., more slowing) show better performance in the Stop task (i.e., smaller SSRTs). This is also consistent with our model, although this relationship is fairly weak, and thus does not constitute particularly strong support for it.

Finally, we tested the model's third core prediction by performing a partial correlation of TOSD and SSRT, controlling for the duration of slowing. The robust positive correlation between SSRT and TOSD remained ($R=.41$, $p<.0005$), indicating that any variance related to inhibitory processes is not strongly measured by either SSRT or TOSD, relative to the variance in context-monitoring captured by these measures. In other words, the individual differences variance in motoric stopping/slowing processes captured by DoS does not overlap with the individual differences variance that is shared by SSRT and TOSD, indicating that the commonality of TOSD and SSRT does not reflect a common motoric stopping process. This conclusion is

concordant with a subsequent re-analysis of our data, presented below, which demonstrates that the overlapping task variance in univariate and multivariate hemodynamics, event-related potentials, and the relative patterns of mental effort do not substantially change when only those Double Go task trials categorized as "unslowed" are analyzed.

Supporting Methods: Analyses of only *un*slowed Double Go_{Signal} trials

To further test our hypothesis that the commonalities of the Double Go and Stop tasks do not reflect a common process of motoric stopping or motor plan replacement, we re-analyzed the fMRI, ERP, and pupillometric data including only those trials that were categorized as "unslowed."

For fMRI, this entailed the respecification of the design matrix for each individual subject. Trials categorized as "slowed" were separately modeled with a new boxcar regressor, and convolved with a double gamma hemodynamic response function (just like our other regressors). These same trials were then omitted from all other regressors (except for the sustained regressor) to avoid issues related to colinearity. This technique allows the hemodynamic response to slowed trials to be separately estimated, and therefore not contaminate the contrasts of transient hemodynamic activity across tasks, nor to contaminate the estimates of percent signal change for the sustained hemodynamic activity across task blocks. This re-analysis revealed similar patterns as we had observed across all Double Go_{Signal} trials: univariate transient hemodynamics were still observed throughout the rVLPFC on unslowed Double Go_{Signal} trials, and these effects were still significantly larger than those observed on $Stop_{Signal}$ trials (Supporting Table 4, row 1). This result indicates that the increased hemodynamic

response to Double Go_{Signal} trials is not driven by the motoric slowing that is captured by "slowed" trials within the Double Go task.

Likewise, this analysis also replicated our previous finding of sustained activity across all trials in the Double Go task (Supporting Table 4, row 2). This finding again indicates that context-monitoring processes, and not motoric stopping or slowing processes, contribute to the sustained activity that is recruited across all trials within the Double Go task.

For the ERP data, this re-analysis entailed re-segmenting the original timeseries of each recording session so that trials categorized as "slowed" could be given their own category, and then excluded from subsequent analysis steps. Otherwise, all ERP analysis procedures were then performed as in the primary analysis, including 40Hz filtering, bad channel replacement, average referencing and polar average reference correction, stimulus locking, baseline correction, montage averaging, and ERP correlations. This re-analysis once again replicated our primary analyses, such that the so-called "Stop P3" was in fact significantly *enhanced* on unslowed Double Go_{Signal} trials relative to $Stop_{Signal}$ trials (Supporting Table 4, row 5), indicating that this ERP does not reflect stopping-specific processes. Indeed, all recorded ERPs were still strongly correlated across tasks, indicating that this similarity at the group level was paralleled by similarities in ERPs at the level of individual subjects, even when all of the "slowed" Double Go_{Signal} trials were excluded from analysis (Supporting Table 4, row 6). Finally, these correlations were again disproportionately increased over frontal electrodes (relative to occipital ones) in the period following signal onset (Supporting Table 4, row 7). By fully replicating our original result when slowed Double Go_{Signal} trials were excluded, this pattern indicates that the increased cross-task similarity in frontal ERPs that is yielded by Signal presentation is not simply driven by the motoric slowing that occurred within the Double Go task. For the pupillometric

data, this re-analysis entailed again resegmenting the original timeseries recording, and omitting the "slowed" Double Go_{Signal} trials from all subsequent analysis steps (which were otherwise identical to those for the primary analysis). Once again, we were able to replicate our original findings after excluding those Double Go_{Signal} trials that had been categorized as slowed, such that pupil diameter was still largest on Double-Go_{Signal} trials than on any other trial type, including $Stop_{Signal}$ trials (Supporting Table 4, row 8). In fact, these patterns were slightly enhanced, indicating that the increased mental effort on Signal trials of the Double Go task is not merely driven by any additional effort required for motoric slowing, which itself appears to be negligible.

Supporting Discussion

Although the fixed task order used here might be expected to yield greater fatigue in the Stop task, fatigue cannot viably explain at least five prominent features of our results. First, fatigue would predict a reduction in pupil diameter during the Stop task, but instead a larger pupil diameter was observed on $Stop_{NoSignal}$trials than Double $Go_{NoSignal}$ trials. Second, fatigue would predict a global reduction in ERPs in the Stop task, but central and posterior ERPs were enhanced in that task. Third, fatigue would predict global reductions in activation during the Stop task, but areas thought to have domain-general attentional roles showed highly-similar profiles across tasks (e.g., TPJ). Fourth, within-task baselines are used in pupillometry, ERP, and fMRI analysis precisely to control for run-specific effects like fatigue – thus any global differences between tasks (owing to fatigue or more innocuous factors, like scanner drift) are subtractively eliminated in the analyses reported here. Fifth, fatigue would predict that SSRT

should be increased in our sample relative to experiments that administer a stop task first, but the range of SSRT observed here was well within normal.We have argued that our ERP and fMRI analyses demonstrating the shared prefrontal substrates of the tasks does not reflect the slowing or stopping process engaged by the Double Go Task. Instead, we suggest that the slowing effect may instead be better understood as an indirect and peripheral consequence of more general attentional processes. For example, cardiac slowing is also observed during tasks involving preparation of a speeded response[30]. Our observation of response slowing and this previous observation of cardiac slowing alike are highly unlikely to reflect a controlled, effortful, or prefrontally-based stopping process. Although subcortical nuclei such as the subthalamic nucleus might be considered more likely to perform these stopping-specific functions, we caution that the subthalamic nucleus is operative in many tasks, not just those requiring an act of stopping. In fact, its activity positively correlates with response force[31], suggesting it may also not have a stopping-specific function. Indeed, others have recently shown a role for the STN in detecting behaviorally relevant stimuli, findings that support a re-evaluation of STN function that is analogous to the current re-evaluation of rVLPFC function[32].

While we have argued that context-monitoring, not stopping, is the cognitively-controlled component to response inhibition tasks, we do not have a position on whether one or both of these processes may be accessible to consciousness. The relationship between consciousness and cognitive control is controversial, with some recent work indicating that rVLPFC can be activated by subthreshold infrequent stimuli that require stopping[33-34]. These results might be taken to contradict accounts that either context-monitoring or stopping is cognitively controlled – at least, under the untested assumption that controlled processes are always consciously

accessible. We suggest this assumption warrants empirical test, and is therefore a promising direction for future work.

We also note that cognitive control is not defined by temporal order (if monitoring must occur prior to stopping, this does not imply that monitoring is cognitively controlled) nor solely by prefrontal recruitment (striatum may have cognitive control functions, and the rVLPFC in particular has been previously argued to *not* subserve cognitive control[35]). Instead, assessing cognitive control requires more comprehensive analysis, of the kind we provide here.

Supporting Table 1

	Experiment		
	Exp. 1	**Exp. 2**	**Exp. 3**
# of trials (trials per block)	400 (100)	970 (46.19)	486 (60.75)
Luminance-matched signal and no signal trials	No	Yes	Yes
Feedback	Sham for Double Go Task, veridical for Stop task	Veridical for both	Veridical for both

Supporting Table 2.

Model	Maximum Likelihood Estimates				BIC
	M	σ	*k*	Θ	
Gaussian/Gamma Mixture	- (fixed to 0)	22.99	4.65	16.45	39654.22
Gaussian	15.15	33.41	-	-	40421.96

Supporting Table 3

Measure	Exp. 1	Exp. 2	Exp. 3
$SSRT_{AV}$(sd)	216ms (33)	206ms (45)	232ms (40)
TOSD(sd)	246ms (60)	203ms (66)	200ms (41)
Duration of Slowing (sd)	60ms (25)	72ms (21)	56ms (14)
Pearson R: TOSD vs $SSRT_{AV}$	.38 (p<.0005)	.46 (p=.002)	.61 (p=.006)
Pearson R: Slowing Duration vs $SSRT_{AV}$	-.18 (p>.1)	-.246 (p>.1)	.43 (p=.1)*
Double Go TaskRT_{Signal}(sd)	387ms (45)	363ms (54)	361ms (41)
Double Go Task$RT_{No\ Signal}$(sd)	371ms (35)	343ms (36)	349ms (34)
Stop TaskRT_{Signal}(sd)	379ms (36)	353ms (40)	366ms (40)
Stop Task $RT_{No\ Signal}$(sd)	426ms (74)	381ms (50)	383ms (59)

* - a single outlier contributed strongly to this positive trend. Exclusion of that subject led to a highly non-significant correlation ($p > .49$) while not substantially affecting the TOSD vs $SSRT_{av}$correlation (which actually became more significant, with the p value reduced to .001)

Supporting Table 4

Critical Tests Involving Double Go_{Signal} trials	**Analysis of All Correct Trials (As described in main text)**	**Re-analysis after Excluding Trials Categorized as "Slowed"**
Univariate fMRI; Transient Recruitment across ROIs: Contrasts of Percent Signal Change (Double Go Task > Stop Task)	STN: t(17)=5.49, p<.0001 BA 44: t(17)=5.08, p<.0001 BA 45: t(17)=2.83, p=.012 BA 47: t(17)=2.5, p=.023 Interaction with TPJ: F(1,17)=31.57, p<.0001	STN: t(17)=4.18, p=.001 BA 44: t(17)=4.11, p=.001 BA 45: t(17)=2.75, p=.014 BA 47: t(17)=3.38, p=.004 Interaction with TPJ: F(1,17)=14.82, p=.001
Univariate fMRI: Mean Percent Signal Change for Sustained rVLPFC Activity Within Double Go Task (and t-statistics)	BA 44: M= .139 t(17)=2.76, p=.01 BA 45: M=.284 t(17)=4.51, p<.001 BA 47: M=.211 t(17)=3.37, p<.005	BA 44: M=.144 t(17)=2.91, p=.01 BA 45: M= .319 t(17)=4.95, p<.001 BA 47: M=.234 t(17)=3.66, p=.002
fMRI MVPA; Classification of Individual Subjects: Contrasts of Performance on Signal vs. No Signal trials	BA44: t(9)=13.5, p<.0001; BA45: t(9)=11.39, p<.0001; BA47: t(9)=12.35, p<.001 Interaction with M1: F(1,9)=85.12, p<.0001	BA44: t(9)=11.84, p<.0001; BA45: t(9)=14.10, p<.0001; BA47: t(9)=8.29, p<.001 Interaction with M1: F(1,9)=200.132, p<.0001
fMRI MVPA; Classification of Trial Types: Interactions of D-Prime across ROIs	Interaction of rVLPFC's BA's (44, 45 and 47) with M1: F(1,17)=13.14, p<.005	Interaction of rVLPFC's BA's (44, 45 and 47) with M1: F(1,17)=9.17, p<.01
ERPs: Stop P3 Amplitude Comparison Across Tasks	t(35)=2.92, p<.03	t(35)=2.19 p<.04
ERPs: Correlation of Scalp Voltages Across Tasks Following Signal Onset	Pearson R: Median: .815 Range: .429-.890	Pearson R: Median: .805 Range: .462-.871

ERPs: Change in Correlation After Signal Onset – Interaction with Montage	Interaction of Frontal vs. Occipital Electrodes: $F(1,98)=12.59$, $p=.001$	Interaction of Frontal vs. Occipital Electrodes: $F(1,98)=46.79$, $p<.0005$
Pupillometry: Comparison of average pupil diameter across trial types	$Stop_{Signal}$<Double Go_{Signal} $t(85)=13.67$, $p<.001$	$Stop_{Signal}$<Double Go_{Signal} $t(85)=13.02$, $p<.001$

SUPPORTING FIGURE LEGENDS

Supporting Fig. 1. Stimuli used in the three experiments. (A) Experiment 1 included null trials consisting only of a fixation ring, constituting 33% of the total number of trials. Of the remaining trials, 75% were No-Signal trials – i.e., 2AFC trials in which either a left-pointing or right-pointing arrow was presented. 25% were Signal trials, in which a white box followed the onset of the 2AFC stimulus. (B) Experiments 2 & 3 used this slightly different set of stimuli, in which the arrows were replaced with left- or right-pointing triangles, and the number of illuminated pixels was matched between the triangles and squares.

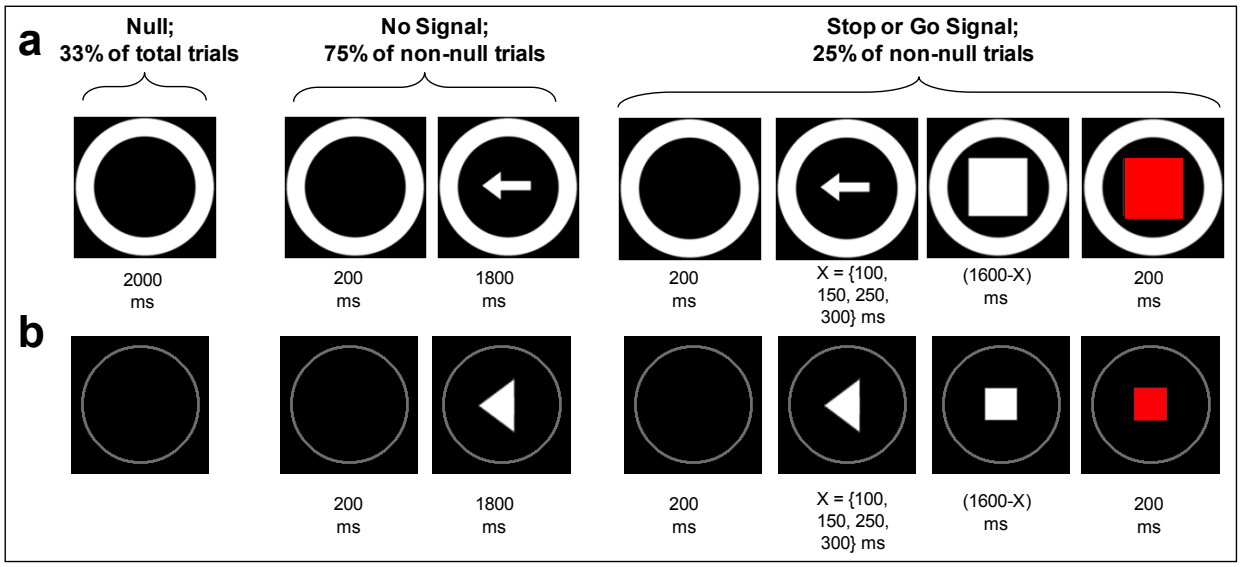
Supporting Fig. 2. MVPA methods. (A). For classifying subjects, neural networks received inputs consisting of 1 unit per voxel in a given ROI, where the activity of those units corresponds to the z-transformed and trimmed parameter estimates from the unsmoothed BOLD data. This input layer projects to a hidden "Scalar Val" layer, which transforms each input unit's activity into a distributed pattern across 30 dedicated units. Finally, this hidden layer is fully connected with an output layer consisting of 18 units, one corresponding directly to each of our subjects. (B). For classifying trial types, we used the same architecture as in A except that only 2 output units were used, corresponding directly to each of the trial type contrasts. In addition, separate networks were trained for each subject.

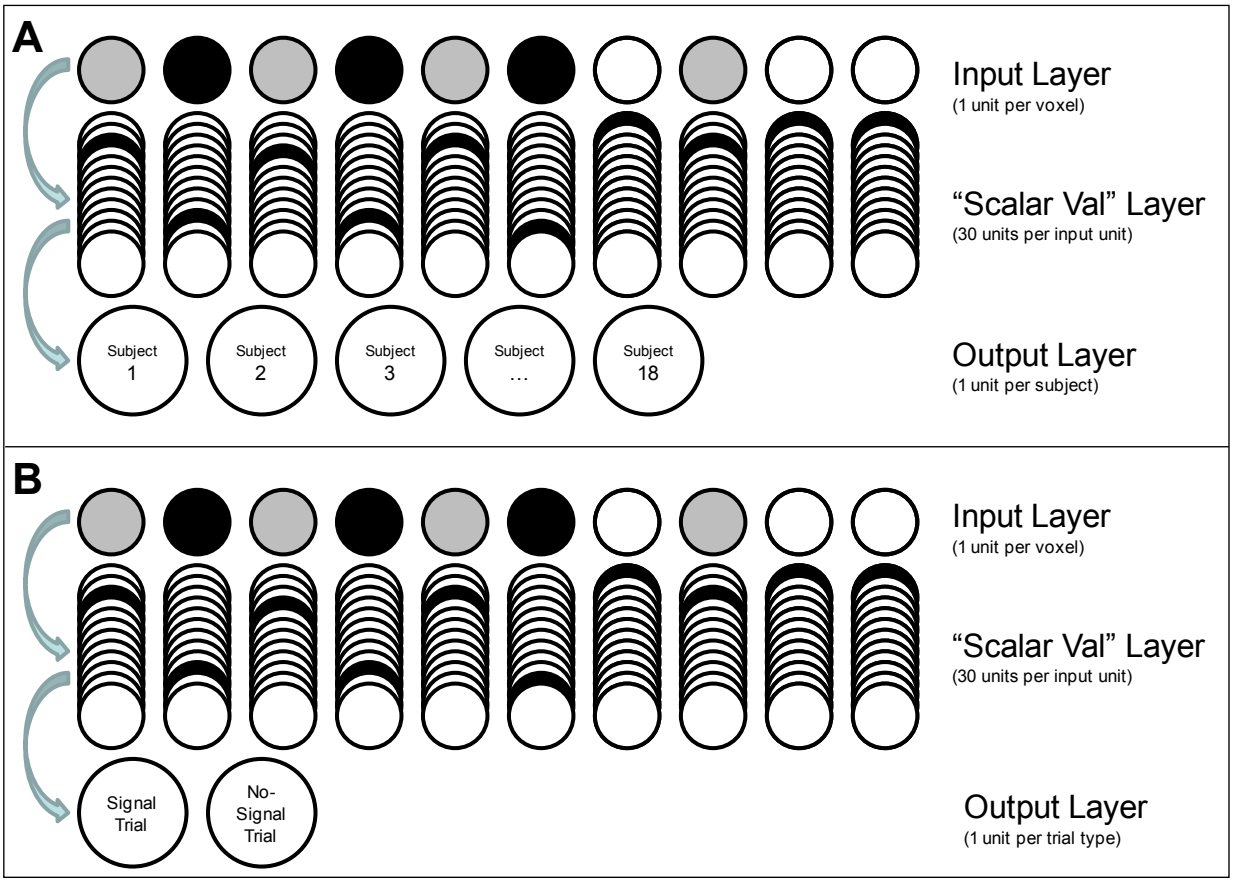
Supporting Fig. 3. Tasks can be discriminated in all ROIs, including V1. Although tasks were best classified on the basis of the Signal > Null contrast (white bars), this is unlikely to reflect stopping-specific processes, since activity patterns in V1 allowed the best classification on this contrast. Indeed, V1 showed the best classification of tasks across all ROIs, when averaging across contrasts. Because our tasks were collected in separate runs, this good classification performance is likely to reflect run-specific variance, rather than task-specific variance. This conclusion is further supported by above-chance discrimination of tasks on the basis of nuisance trials, during which both stimuli and responses were precisely matched across tasks/runs.

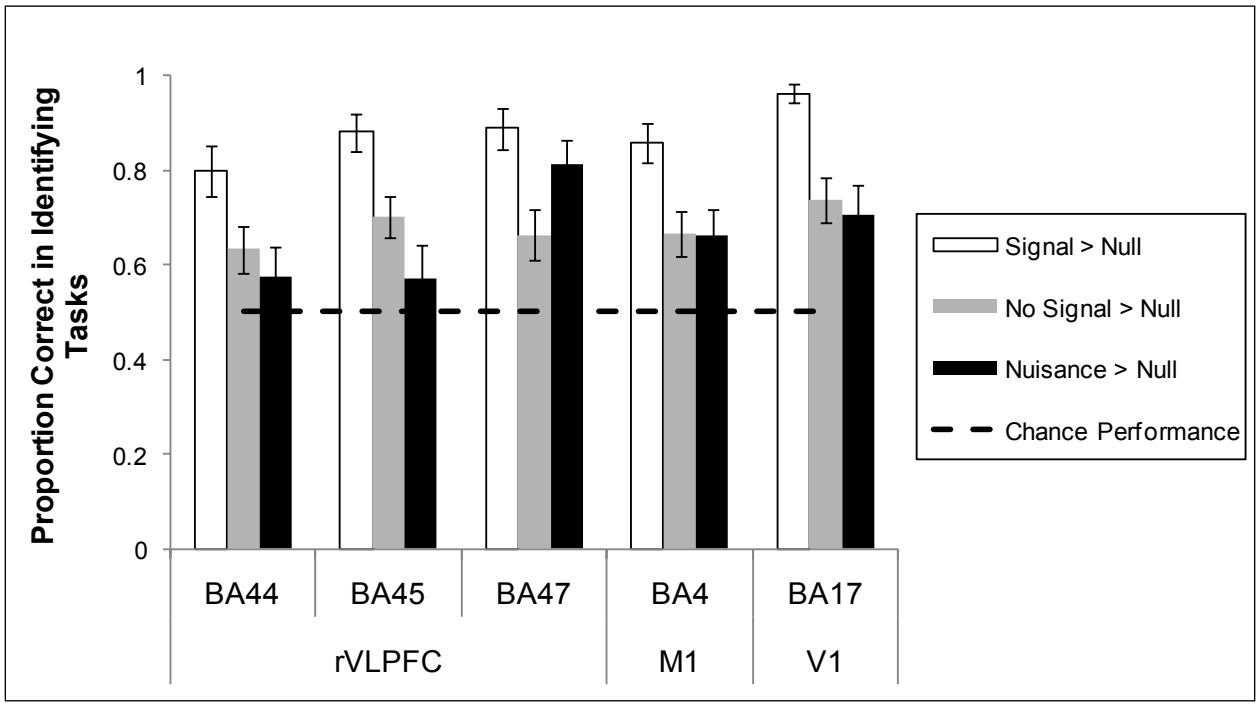
Supporting Fig. 4.The typical pattern of "P3 anteriorization" in tasks that demand stopping was reversed in our tasks, such that Double Go_{Signal} trials elicited a larger P3 than the $Stop_{Signal}$ trials at the site where the Stop P3 is typically maximal (A). In contrast, the opposite was true of more posterior electrodes (B & C), indicating that anteriorization effects cannot not be taken to index explicit motoric stopping demands.

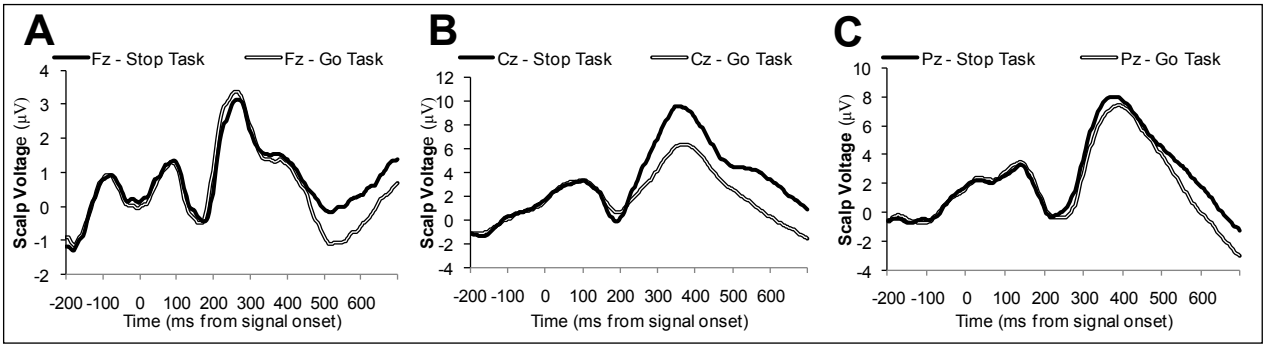
Supporting Fig. 5.The group-average scalp distribution of ERPs elicited by $Stop_{Signal}$ and Double Go_{Signal} trials were markedly similar, consistent with the strong relationship of these ERPs at the level of individual differences. In particular, the anteriorization of the P3 ERP elicited by Double Go_{Signal} trials, relative to that elicited by $Stop_{Signal}$ trials, is visible in the highlighted portion of each figure. Each contour represents a change of .79μV; red is positive.

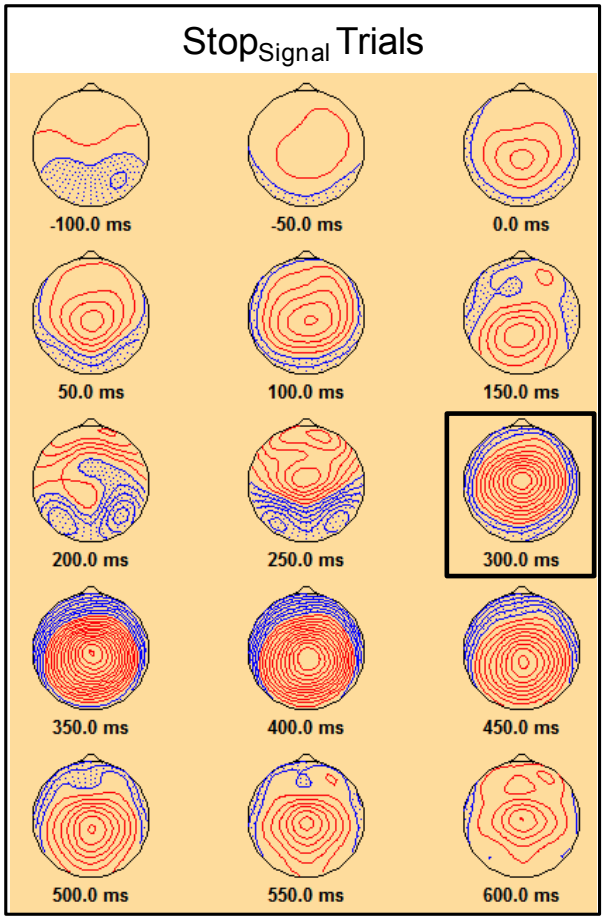
Supporting Fig. 6. Process-models of our tasks. (A). The race model is used to analyze behavior in the Stop task, such that the amount of warning necessary to stop (Stop Signal Reaction Time, or SSRT) can be extracted as the n^{th} percentile of the $Stop_{No\text{-}Signal}$ distribution, where n corresponds to the percent of unsuccessfully stopped responses at a particular signal delay. (B) A conceptually similar model is used to analyze behavior in the Double Go Task, but allows the extraction of two underlying parameters. The duration of slowing can be estimated as the difference between slowed 1^{st} responses on Double Go_{Signal}trials and responses of the same percent rank on Double $Go_{No\text{-}Signal}$ trials. The time of signal detection can be estimated as the amount of time that must elapse following a signal before responses are slowed. (C) The process model of the Double Go Task predicts that slowing should be larger when signals are presented earlier; this prediction was confirmed.

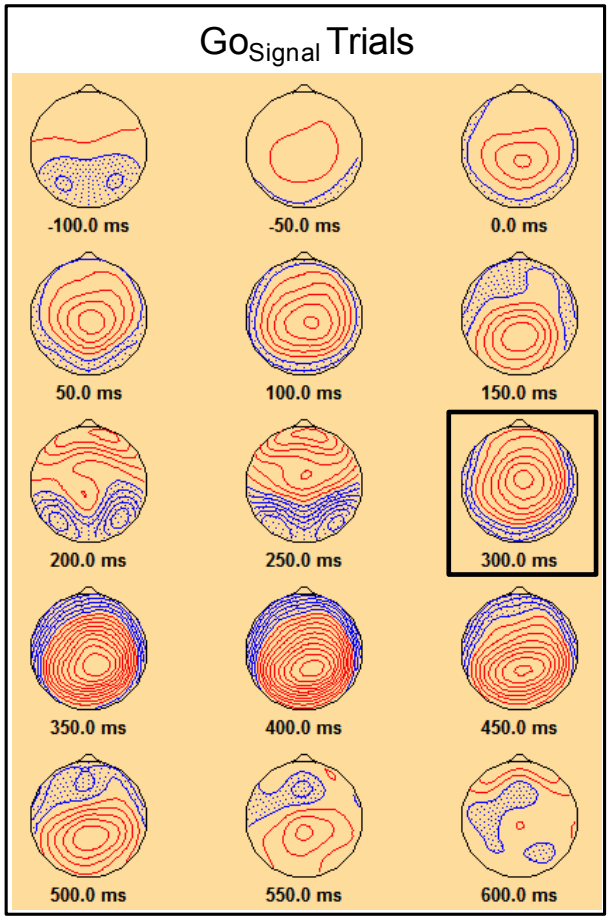

a
Null;
33% of total trials
No Signal;
75% of non-null trials
Stop or Go Signal;
25% of non-null trials
2000
ms
200
ms
1800
ms
200
ms
X = {100,
150, 250,
300} ms
(1600-X)
ms
200
ms
b
200
ms
1800
ms
200
ms
X = {100,
150, 250,
300} ms
(1600-X)
ms
200
ms

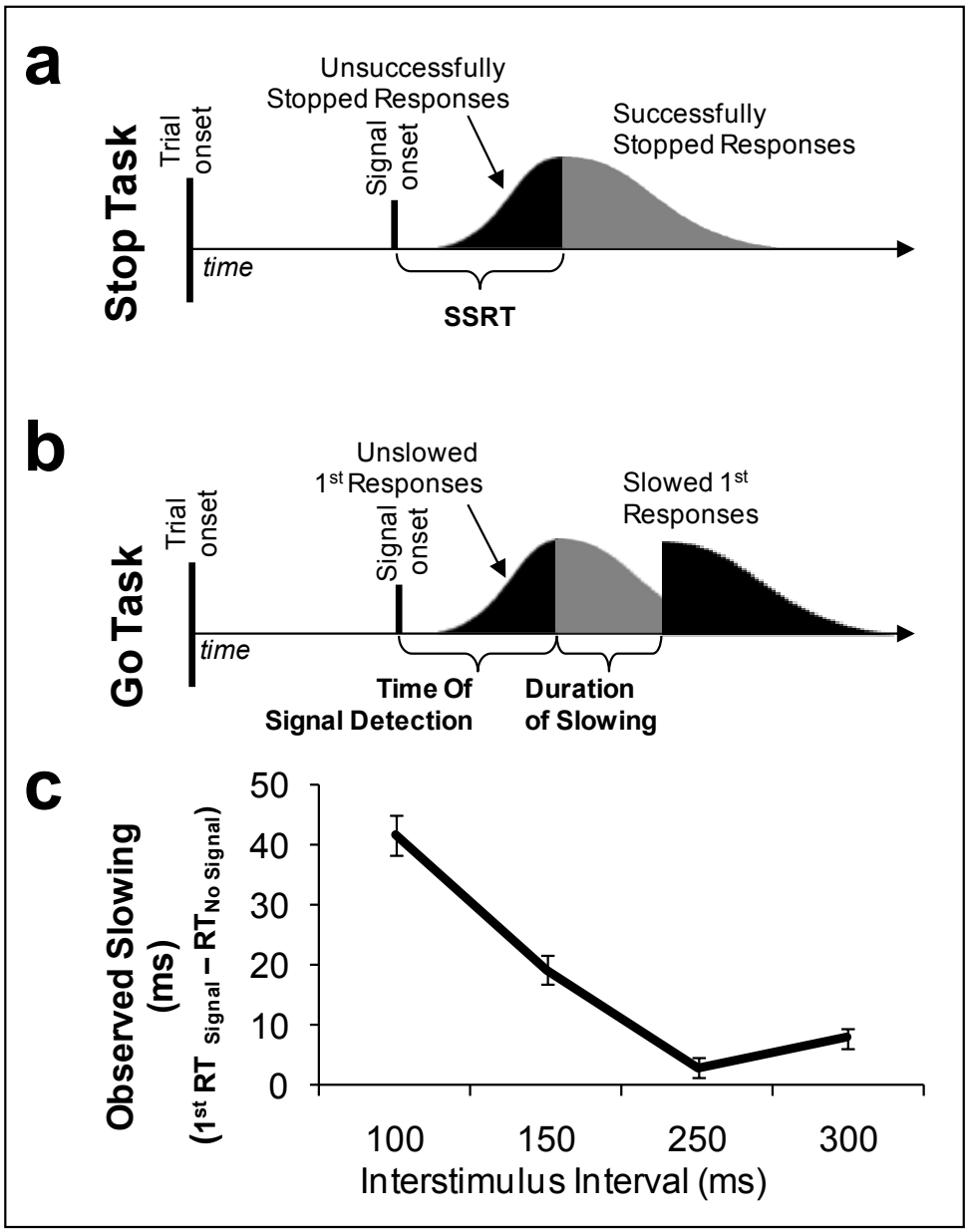

A
Input Layer
(1 unit per voxel)
"Scalar Val" Layer
(30 units per input unit)
Subject 1
Subject 2
Subject 3
Subject …
Subject 18
Output Layer
(1 unit per subject)
B
Input Layer
(1 unit per voxel)
"Scalar Val" Layer
(30 units per input unit)
Signal Trial
No-Signal Trial
Output Layer
(1 unit per trial type)


Proportion Correct in Identifying Tasks
1
0.8
0.6
0.4
0.2
0
BA44
BA45
BA47
BA4
BA17
rVLPFC
M1
V1
Signal > Null
No Signal > Null
Nuisance > Null
Chance Performance


A
Fz - Stop Task
Fz - Go Task
Scalp Voltage (μV)
4
3
2
1
0
-1
-2
-200 -100 0 100 200 300 400 500 600
Time (ms from signal onset)
B
Cz - Stop Task
Cz - Go Task
Scalp Voltage (μV)
12
10
8
6
4
2
0
-2
-4
-200 -100 0 100 200 300 400 500 600
Time (ms from signal onset)
C
Pz - Stop Task
Pz - Go Task
Scalp Voltage (μV)
10
8
6
4
2
0
-2
-4
-200 -100 0 100 200 300 400 500 600
Time (ms from signal onset)


Stop Signal Trials
-100.0 ms
-50.0 ms
0.0 ms
50.0 ms
100.0 ms
150.0 ms
200.0 ms
250.0 ms
300.0 ms
350.0 ms
400.0 ms
450.0 ms
500.0 ms
550.0 ms
600.0 ms


Go Signal Trials
-100.0 ms
-50.0 ms
0.0 ms
50.0 ms
100.0 ms
150.0 ms
200.0 ms
250.0 ms
300.0 ms
350.0 ms
400.0 ms
450.0 ms
500.0 ms
550.0 ms
600.0 ms


a
Stop Task
Trial onset
Unsuccessfully Stopped Responses
Signal onset
Successfully Stopped Responses
time
SSRT
b
Go Task
Trial onset
Unslowed 1st Responses
Signal onset
Slowed 1st Responses
time
Time Of Signal Detection
Duration of Slowing
c
Observed Slowing (ms)
(1st RT Signal – RTNo Signal)
50
40
30
20
10
0
100
150
250
300
Interstimulus Interval (ms)

SUPPORTING REFERENCES

1. Aron AR (2007) The neural basis of inhibition in cognitive control. Neuroscientist 13:214-28
2. Aron AR, Robbins TW, Poldrack RA (2004) Inhibition and the right inferior frontal cortex. Trends Cogn Sci 8:170-7.
3. Friedman NP, et al. (2008) Individual differences in executive functions are almost entirely genetic in origin. J Exp Psychol Gen 137:201-25.
4. MacLeod CM, Dodd MD, Sheard ED, Wilson DE, Bibi, U (2003) In Opposition to Inhibition. In: B H Ross, editor. The Psychology of Learning and Motivation, San Diego: Academic Press, pp. 163-214.
5. Ferrier, D (1876) The functions of the brain. London: Elder.
6. Chikazoe J, et al (2009) Functional dissociation in right inferior frontal cortex during performance of Go/No-Go Task. Cereb Cortex 19:146-52
7. Hampshire A, Chamberlain SR, Monti MM, Duncan J, Owen AM (2010) The role of the right inferior frontal gyrus: inhibition and attentional control. Neuroimage 50: 1313-9.
8. Sharp DJ, et al. (2010) Distinct frontal systems for response inhibition, attentional capture, and error processing Proc Natl Acad Sci 107:6106-11 .
9. Cai W, Leung HC (2011) Rule-Guided Executive Control of Response Inhibition: Functional Topography of the Inferior Frontal Cortex. *PLoS One* 6(6): e20840. doi:10.1371/journal.pone.0020840
10. Dodds CM, Morein-Zamir S, Robbins TW (in press) Dissociating Inhibition, Attention, and Response Control in the Frontoparietal Network Using Functional Magnetic Resonance Imaging. Cerebral Cortex.
11. Logan GD, Cowan, WB (1984) On the ability to inhibit thought and action: A theory of an act of control. Psych Review 91:295-327.
12. Aron, AR (2010) From Reactive to proactive and Selective Control: Developing a Rcicher Model for Stopping Inappropriate Responses. *Biol Psychiat*
13. Jahfari S, Stinear CM, Claffey M, Verbruggen F, Aron AR (2010) Responding with restraint: what are the neurocognitive mechanisms? J Cog Neurosci 22:1479-92.
14. Smith JL, Johnstone SJ, Barry RJ (2008) Movement-related potentials in the Go/NoGo task: the P3 reflects both cognitive and motor inhibition. Clin Neurophysiol 119:704-14.
15. Kahneman D, Beatty J (1966) Pupillary changes in two memory tasks. Science 154:1583-5.
16. Beatty J, Lucero-Wagoner, B. (2000) The Pupillary System. In: Caccioppo J, Tassinary LG, Berntson G, editors. The Handbook of Psychophysiology. Hillsdale: Cambridge University Press, pp 142-62
17. Chambers CD, et al (2006) Executive "brake failure" following deactivation of human frontal lobe. J Cogn Neurosci 18:444-55.
18. Verbruggen F, Aron AR, Stevens MA, Chambers CD (2010) Theta burst stimulation dissociates attention and action updating in human inferior frontal cortex. Proc Natl Acad Sci U S A 107:13966-71.
19. Chikazoe J, Konishi S, Asari T, Jimura K, Miyashita Y (2007) Activation of right inferior frontal gyrus during response inhibition across response modalities. J Cogn Neurosci 19:69-80.

20. Konishi S, Nakajima K, Uchida I, Sekihara K, Miyashita Y (1998) No-go dominant brain activity in human inferior prefrontal cortex revealed by functional magnetic resonance imaging. Eur J Neurosci 10:1209-1213.
21. Rubia K, Smith AB, Brammer MJ, Taylor E (2003) Right inferior prefrontal cortex mediates response inhibition while mesial prefrontal cortex is responsible for error detection. Neuroimage 20:351-8.
22. Corbetta M, Patel G, Shulman GL (2008) The reorienting system of the human brain: from environment to theory of mind. Neuron 58:306-324.
23. Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. Nat Rev Neurosci 3:201-215.
24. Barch, D. M., Braver, T. S., Nystrom, L. E., Forman, S. D., Noll, D. C., & Cohen, J. D. (1997). Dissociating working memory from task difficulty in human prefrontal cortex. Neuropsychologia, 35: 1373-1380.
25. Neubert FX, Mars RB, Buch ER, Olivier E, Rushworth MF (2010) Cortical and subcortical interactions during action reprogramming and their related white matter pathways. Proc Natl Acad Sci U S A 107:13240-5.
26. Civardi C, Cantello R, Asselman P, Rothwell JC (2001) Transcranial magnetic stimulation can be used to test connections to primary motor areas from frontal and medial cortex in humans. Neuroimage 14:1444–1453.
27. Verbruggen V, Logan, G (2009). Automatic and Controlled Response Inhibition: Associative Learning in the Go/No-Go and Stop-Signal Paradigms. Journal of Experimental Psychology 137:649-672.
28. Verbruggen F, Liefooghe, B; Szmalec, A; Vandierendonck,, A (2005) Inhibiting Responses When Switching: Does it Matter? Journal of Experimental Psychology 52:125-130.
29. Aron AR, Monsell S, Sahakian BJ, Robbins TW (2004) A componential analysis of task-switching deficits associated with lesions of left and right frontal cortex. 127:1561-1573.
30. Stuss DT, Alexander MP (2007) Is there a dysexective syndrome? Philos Trans R Soc Lond B Biol Sci 362:901-915.
31. Miller EK, Cohen JD (2001) An integrative theory of prefrontal cortex function. Ann Rev Neurosci, 24:167-202.
32. O'Reilly RC (2010) The What and How of prefrontal cortical organization Trends in Neurosciences, 33:355-361.
33. Nambu A, Tokuno H, Inase M, Takada M (1997) Corticosubthalamic input zones from forelimb representations of the dorsal and ventral divisions of the premotor cortex in the macaque monkey: comparison with the input zones from the primary motor cortex and the supplementary motor area. Neurosci Lett 239:13-16.
34. Craig AB (2009) How do you feel – now? The anterior insula and human awareness. Nat Rev Neurosci 10:59-70.
35. Lovero KL, Simmons AN, Aron JL, Paulus MP (2009) Anterior insular cortex anticipates impending stimulus significance. Neuroimage 45:976-83.
36. Lim SL, Padmala S, Pessoa L (2009) Segregating the significant from the mundane on a moment-to-moment basis via direct and indirect amygdala contributions Proc Natl Acad Sci U S A 106:16841-16846.

37. Michael GA, Garcia S, Fernandez D, Sellal F, Boucart M (2006) The ventral premotor cortex (vPM) and resistance to interference Behav Neurosci 120:447–62.
38. Castellanos FX, Sonuga-Barke EJS, Milham MP, Tannock R (2006) Characterizing cognition in ADHD: beyond executive dysfunction. Trends Cog Sci 10: 117-123.
39. Thorell LB, Lindqvist S, Nutley SB, Bohlin G, Klingberg T (2009) Training and transfer effects of executive functions in preschool children. Dev Sci 12:106-113.
40. Aron AR, Poldrack RA (2006) Cortical and subcortical contributions to stop signal response inhibition: role of the subthalamic nucleus. J Neurosci 26:2424-2433.
41. Corbetta M, Kincade JM, Ollinger JM, McAvoy MP, Shulman GL (2000) Voluntary orienting is dissociated from target detection in human posterior parietal cortex. Nat Neurosci 3:292-7.
42. Aisa B, Mingus B, O'Reilly RC (2008) The emergent neural modeling system. Neural Networks, 21(8):1146-1152.
43. Tucker DM (1993) Spatial sampling of head electrical fields: The geodesic sensor net. Electroenceph & Clinical Neurophys 87:154–16.
44. Junghoefer M, Elber T, Tucker DM, Braun C (1999) The polar average reference effect: A bias in estimating the head surface integral in EEG recording. Clinical Neurophysiology, 110, 1149–1155.
45. Nelder JA, Mead RA (1965) Simplex method for function minimization. Comp J 7 :308–313.
46. Ratcliff R, Starns JJ (2009) Modeling confidence and response time in recognition memory. Psychol Rev 116:59-83.
47. Band GP, van der Molen MW, Logan GD (2003) Horse-race model simulations of the stop-signal procedure. Acta Psychol 112:105-42.