Separating Within-Trial Activities in the Stop Signal Task Using fMRI

by

Andre D. Chevrier

A thesis submitted in conformity with the requirements for the degree of Doctor of Philosophy

Graduate Department of Medical Science, Faculty of Medicine
University of Toronto

© Copyright by Andre D. Chevrier 2016
Separating Within-Trial Activities in the Stop Signal Task Using fMRI

Andre D. Chevrier
Doctor of Philosophy
Graduate Department of Medical Science
University of Toronto
2016

Abstract

Functional magnetic resonance imaging (fMRI) approaches tend to focus on a particular process of interest, and model neural activity as a single phase of activity on a given trial. This thesis presents new methods for separating rapidly unfolding within-trial activities in a stop signal task using fMRI.

Current models of inhibitory control address processes that are engaged after the presentation of go- and stop-signals. The first study presented here separated warning-from response-phase activities in healthy adults (A. Chevrier, Cheyne, Graham, & Schachar, 2015). This work revealed activities in posterior networks that signal the need for top-down control followed by fronto-posterior activities involved in implementing top-down control. Both of these phases contained activities that influenced inhibitory control.

Reinforcement learning models treat errors as a unitary phase of activity. The second study presented here separated error detection from post-error slowing on failed stop trials (A. Chevrier & Schachar, 2010). Error detection deactivated structures that are richly innervated by midbrain dopamine neurons (dorsal striatum and ventral anterior cingulate cortex (ACC)), followed by deactivations during post-error adjustment in structures that modulate the ascending dopamine response (ventral striatum and caudal orbitofrontal cortex), which encode error magnitude.
The third study involved the separation of error detection and post-error slowing in adolescents with and without attention deficit hyperactivity disorder (ADHD) (Chevrier, Bhaijiwala, Cheyne, Graham & Schachar, manuscript in preparation). A recent study of these data from our laboratory (Bhaijiwala, Chevrier, & Schachar, 2014) showed that previous evidence of ADHD subjects under-activating inhibition networks, is instead the result of deactivating these networks during preparation, when healthy controls were pre-activating. This suggests that inhibitory control deficits in ADHD result from dysfunctional reinforcement, which tunes appropriate networks in an inappropriate way. Consistent with this hypothesis, we found that ADHD subjects failed to fully deactivate dopamine targets on error detection, and inappropriately activated the reciprocal pathway on post-error slowing. Dysfunctional reinforcement would explain the lack of potency of behavioral interventions that are based on the known properties of normal reinforcement. Further, the effects of dysfunction in moment-to-moment reinforcement would accumulate over experience, suggesting a mechanism for the delayed maturation of function and cortical thickness observed in ADHD.
Acknowledgments

This thesis would not have been possible without the help of many people. Firstly, my committee has provided a base of expertise to tether lofty ideas to concrete principles. My supervisor, Dr. Douglas Cheyne, is a rare combination of knowledge, skill and organization, and was always there when I needed him. Dr. Simon Graham has a unique ability to indelibly convey pointed insights in the friendliest and most helpful ways, which always lead to practical improvements. Dr. Russell Schachar provided the original motivation for this work. He has been a great friend and mentor to me, and consistently facilitated a view of the broader implications of basic neuroscientific research.

To my fellow graduate student, Mehereen Bhaijiwala, this work would not have been possible without you. I would like to thank Linda Citren, Dr. Jen Crosibe, Tara Goodale, and Janet Shan, who have always been there to help even when I did not know what help I needed. I would also like to thank all the coworkers, friends and family members who have endured many conversations and provided invaluable feedback about the work presented here.
# Table of Contents

Acknowledgments ........................................................................................................iv

Table of Contents ...........................................................................................................v

List of Tables ..................................................................................................................ix

List of Figures ..................................................................................................................x

List of Abbreviations .....................................................................................................xi

Chapter 1 .........................................................................................................................1

1 Introduction ..................................................................................................................1

1.1 Motivation for current work .....................................................................................1

1.2 Overview of Chapters ...............................................................................................4

1.3 Contributions .............................................................................................................7

Chapter 2 .........................................................................................................................9

2 Literature review .........................................................................................................9

2.1 Functional magnetic resonance imaging (fMRI) .......................................................9

2.2 Executive function .....................................................................................................12

2.3 Inhibitory control .....................................................................................................13

2.4 The stop signal task (SST) ......................................................................................14

2.5 SSRT and the race model of inhibition ....................................................................17

2.6 fMRI of inhibitory control .....................................................................................18

2.7 Preparation ...............................................................................................................19

2.8 Conflict, errors and performance monitoring ........................................................21

2.9 Dopamine, reward and reinforcement learning .......................................................23

2.9.1 Reward ...............................................................................................................23

2.9.2 Reinforcement learning and goal-directed behavior ........................................27
6 Conclusion ........................................................................................................................................142

6.1 Summary of results .......................................................................................................................142

   6.1.1 Warning- and response-phase activities in the SST ..............................................................142

   6.1.2 Error detection and post-error slowing ..................................................................................143

   6.1.3 Error detection and post-error slowing in adolescent ADHD ............................................144

6.2 Speculations on ADHD, dopamine and the suspension of disbelief .........................................146

6.3 Future work ..................................................................................................................................162

7 References .......................................................................................................................................167
List of Tables

Table 3-1  BOLD activities during warning-phases 57
Table 3-2  BOLD activities during response-phases 58
Table 4-1  Trial types and deconvolution regressors 79
Table 4-2  Post-error slowing activities 88
Table 5-1  Activities in putative dopamine pathways 118
Table 5-2  Group activity differences during error detection (TD-ADHD) 120
Table 5-3  Group activity differences during post-error slowing (TD-ADHD) 121
List of Figures

Figure 2-1 Trial types and regressors in the SST 16

Figure 3-1 Trial types and regressors for isolating preparatory activity in the SST 50

Figure 3-2 Warning-phase activities 59

Figure 3-3 Response-phase activities 61

Figure 4-1 Stop signal task 76

Figure 4-2 Error detection deactivations 85

Figure 4-3 Selected deactivations on post-error slowing 87

Figure 5-1 Error detection activity in ventral ACC 115

Figure 5-2 Error detection activity in dorsal ACC 115

Figure 5-3 Error detection activity in ventral ACC and dorsal striatum 116
List of Abbreviations

ACC – Anterior cingulate cortex
ADHD – Attention deficit hyperactivity disorder
AFNI – Analysis of functional neuroimages
ANOVA – Analysis of variance
BA – Brodmann area
BOLD - Blood-oxygen-level dependent
DA – Dopamine
dACC – Dorsal ACC
DLPFC – dorsolateral prefrontal cortex
DMN – Default mode network
DSM - Diagnostic and Statistical Manual of Mental Disorders
ER-fMRI – Event-related fMRI
ERP – Event-related potential
fMRI – Functional magnetic resonance
FOV – Field of view
HPC – Hippocampus
HRF – Hemodynamic response function
IFG – Inferior frontal cortex
IPL – Inferior parietal lobe
ITI – Inter-trial interval
lp – Lateral posterior nucleus
OFC – Orbitofrontal cortex
MFG – Middle frontal gyrus
MPD – Multiple personality disorder
MRI – Magnetic resonance imaging
MTG – Middle temporal gyrus
NA – Nucleus accumbens
ODD – Oppositional defiant disorder
RP – Readiness potential
PET – Positron emission tomography
pHPC - Parahippocampus
PICS – Parent interview for child symptoms
PCC – Posterior cingulate cortex
RF – Radio frequency
ROI – Region of interest
RT – Reaction time
SFG – Superior frontal gyrus
SMA – Supplementary motor area
SPGR - Spoiled Gradient Recalled Acquisition
SPL – Superior parietal lobe
SPN – Stimulus preceding negativity
SSRT – Stop signal reaction time
SST – Stop signal task
STG – Superior temporal gyrus
TD – Typically developing
TE – Echo time
TR – Repetition time
WISC – Wechsler intelligence scale for children
Chapter 1

1 Introduction

1.1 Motivation for current work

Our mental faculties are not static entities located in discrete brain regions that we simply access as needed. Rather, they are spontaneously generated in distributed networks from moment-to-moment according to our changing needs, and are adaptively reinforced based on feedback. Our mental landscapes and corresponding functional brain networks are laid down over development, the result of the accumulation of moments when these networks come into play and are reinforced based on their success or failure.

In contrast to the generative, integrated and reentrant nature of our mental faculties, functional neuroimaging designs informed by cognitive models have been aimed at pinpointing a particular process of interest. These approaches have been both necessary and highly successful at identifying brain regions where the pattern of activity corresponds to the expected behavior of certain variables in our cognitive models, and at testing the predictions of competing models. However, any task involves multiple mental processes, and any given trial involves not only the process in question, but also requires a mechanism to generatively bring that process into play at the beginning of every trial, and to adjust its function based on the outcome of every trial.
The current practice of mapping function onto anatomy based on cognitive models in the absence of a known mechanism is reflective, is destined to be phrenology to the degree that this logic is reflective and therefore circular. Certainly model-based designs are necessary for localizing what parts of the brain are involved in a given function. However, cognitive models remain concepts that describe what must be taking place independent of the physical substrate, not mechanisms that are the result of their physical substrate. Many conceptual models remain too abstract to directly map onto brain function, which requires a physical mechanism in order to work. Cognitive models take a modular view of function, wherein different regions of the brain are assumed to carry out different aspects of these functions, with greater activity reflecting greater computational effort. Although these models often involve rigorous mathematical descriptions of a function, they must not be mistaken for the mathematics of the mechanism of that function. Some unnecessary phrenology arises in functional neuroimaging from mistaking the mathematics of the model for the mathematics of the mechanism.

Cognitive models are meant to be a rigorous account of our concepts of various mental faculties, and are not meant to describe their underlying mechanism. The limited mechanistic insight available from model-based neuroimaging is apparent from the fact that interpretations of activity differences in pathological populations are constrained to simply reflecting the narrative of deficient processing. For example, findings of under-activation are often interpreted as reflecting deficient function, and over-activation as some form of compensation for deficient function, neither of which provides much mechanistic insight into the nature of the deficit.
The problem of phrenology in model-based imaging designs is not solved by added complexity of analysis, such as connectivity or multivariate approaches, because the modular view of neural mechanisms remains. Functional imaging provides an opportunity to take a whole-brain view of functional networks coming into play and being reinforced on a moment-to-moment basis, allowing at least some qualitative inspection of whole-brain mechanisms that are outside the scope of cognitive models. This approach has not been heavily pursued because it does not provide the kinds of specific hypotheses and experimental predictions afforded by imaging cognitive models exclusively. However, contrary to prevalent expert opinion, the difficulty in assessing the scientific merit of inspecting neural activities that are outside the predictions of cognitive models does not make such an approach unscientific.

The scientific method admittedly demands specific experimental hypotheses and predictions. However, scientific enlightenment begins with new observations and the questioning of old assumptions. Certainly the development of new theories invites unforeseen risk, but pure empiricism risks total error. It is therefore necessary that empirical science provide novel data for the development of new theories, and for new theories to generate specific experimental predictions.

In the spirit of generating novel data for interpretation, this thesis presents the development of functional magnetic resonance imaging (fMRI) approaches to explicitly separate multiple phases of within-trial activity to gain a better whole-brain view of relevant circuits coming into play at the beginning of a trial, and being reinforced based on the outcome of that trial. Using a stop signal task, an imaging approach was
developed to separate how task-specific networks are prepared from how they are used, and from how they are reinforced based on the outcome of a trial. This work yielded new insights into the nature of top-down control, and of dopamine-driven reinforcement signals, that have not been accessible to previous approaches based on models of inhibitory control or reinforcement learning. Further, applying this method to the study of ADHD revealed previously inaccessible differences during preparation and feedback that suggest a potential link between altered dopamine signaling and deficient inhibitory control in ADHD.

Therefore, by allowing for stages of processing that are outside the purely conceptual variables of inhibitory control and reinforcement learning, the current approach helps to bring our understanding of these cognitive faculties more into the concrete realm of neural mechanisms.

1.2 Overview of Chapters

Chapter 2 consists of a literature review beginning with an outline of rapid event-related fMRI, followed by relevant cognitive functions and their models, how they are assessed and imaged in healthy and ADHD populations. This section concludes with the general hypotheses for each of the studies presented here.
Chapter 3 presents a method for separating warning- from response-phase activity in healthy adults (Chevrier et al., 2015). This approach revealed that warning-phases activated posterior networks that signal the need for top-down control, whereas response-phases activated fronto-posterior networks involved in engaging that top-down control. The ability to separate warning- from response-phase activities in the stop task identified factors that influence the speed-accuracy trade-offs before stop signals appear, which are outside the scope of existing models of inhibitory control.

Chapter 4 introduces an imaging approach to investigate reinforcement signals on failed stop trials by separating error detection from preceding response-phase and post-error slowing activities. Error detection deactivated an ascending dopamine pathway (substantia nigra, dorsal striatum, ventral ACC and hippocampus (HPC)), whereas post-error slowing deactivated the reciprocal pathway (caudal orbitofrontal, ventral striatum) (Chevrier & Schachar, 2010), which is known to respond to prediction error magnitude. These first two studies confirm that it is possible to separate the engagement of top-down control from its implementation, and the detection of errors from feedback-based reinforcement related to prediction error magnitude.

Chapter 5 describes the separation of error detection from post-error adjustment activities on failed stop trials in adolescents with and without attention deficit hyperactivity disorder (ADHD). The clinical study was conducted by fellow graduate student Mehereen Bhaijiwala, and the analysis and interpretations of results were the result of a cooperative effort, leading to one paper on preparation and inhibition (first author MB (Bhaijiwala et al., 2014)) and one paper on error processing (first author AC
ADHD is characterized by deficient inhibitory control associated with under-activation of inferior frontal gyrus (IFG), presumably when responses are successfully stopped (Smith, Taylor, Brammer, Toone, & Rubia, 2006). In a recent study, we found that ADHD subjects did not under-activate IFG when responses were successfully stopped. Rather, ADHD subjects deactivated IFG and activated regions implicated in the default mode network (DMN) during response-phases that precede stop phases, when typically developing (TD) adolescents were pre-activating IFG and deactivating the DMN (Bhaijiwala et al., 2014). Previous approaches have combined these phases of activity and deduced that IFG under-activates during response inhibition, and reflects deficient inhibitory control in ADHD.

Rather than simply reflecting the narrative of poor inhibitory control, the current findings suggest that inhibitory control deficits in ADHD might arise from a dysfunctional reinforcement process that tunes appropriate networks in an inappropriate way. Consistent with a dysfunctional tuning hypothesis, the clinical study presented here revealed that ADHD subjects failed to significantly deactivate the ascending dopamine pathway during error detection, and activated the reciprocal pathway that deactivated in TD controls during post-error slowing.

This result could potentially explain the known involvement of dopamine in the etiology of ADHD, despite the paucity of evidence of differences in dopamine-related genes or receptor availability (Del Campo et al., 2013). Given that top-down control emerges over the course of development, which consists of the accumulated effects of moment-to-moment reinforcement, these results could explain the delayed maturation of function.
and cortical thickness associated with ADHD (Shaw et al., 2006). This delayed maturation would be pronounced in inhibitory control, which continues to develop well into adulthood (Bedard et al., 2002; Velanova, Wheeler, & Luna, 2009; Williams, Ponesse, Schachar, Logan, & Tannock, 1999). Further, given that feedback-based reinforcement is maladaptive in ADHD, these results might also explain the lack of potency of behavioral interventions based on the known properties of healthy reinforcement learning systems.

Chapter 6 consists of a general discussion of the main findings, future work, and speculations on a reformulation of ADHD.

### 1.3 Contributions

Chapter 3, Dissociating two stages of preparation in the stop signal task using fMRI: Andre Chevrier designed the study, collected data, performed all analyses and wrote manuscript. Russell Schachar co-designed the study, provided funding, contributed to manuscript preparation and submission. Simon Graham consulted regarding imaging and statistical methods, contributed to manuscript editing, submission and responses to reviewer comments. Douglas Cheyne was primary supervisor, contributed to study design, analytical methods, and contributed to manuscript preparation and submission process.

Chapter 4, Error detection in the stop signal task: Andre Chevrier designed the study, collected data, performed all analyses and wrote manuscript. Russell Schachar co-designed the study, provided funding, contributed to manuscript preparation and submission. Simon Graham consulted regarding imaging and statistical methods, contributed to manuscript editing, submission and responses to reviewer comments.
Douglas Cheyne was primary supervisor, contributed to study design, analytical methods, and contributed to manuscript preparation and submission process.

Chapter 5, fMRI of error detection and post-error slowing in adolescent ADHD: Study was designed by Russell Schachar, Mehereen Bhaijiwala and Andre Chevrier. Mehereen Bhaijiwala collected data and wrote first draft of behavioral and clinical methods. Andre Chevrier performed analyses and wrote manuscript. Russell Schachar co-designed the study, provided funding, and contributed to manuscript preparation. Simon Graham consulted regarding imaging and statistical methods, and contributed to manuscript design and editing. Douglas Cheyne was primary supervisor, contributed to study design, analytical methods, and contributed to manuscript preparation.
Chapter 2

2 Literature review

2.1 Functional magnetic resonance imaging (fMRI)

fMRI is an application of magnetic resonance imaging (MRI) based on unique magnetic properties of blood that depend on oxygen content. MRI employs a strong, permanent, static magnetic field to align nuclei (hydrogen protons in free water) in the body. Radiofrequency (RF) pulses are applied to excite these nuclear spins into higher energy states. These spins then gradually return to equilibrium by emitting RF photons in a process referred to as ‘relaxation,’ which can be measured with a coil (Faraday’s law of induction). Additional magnetic fields that vary along a specific spatial dimension (e.g. gradually increasing from left to right) referred to as gradient fields. These gradients change the RF energy (and therefore frequency) of nuclear spins according to their position along the field gradient. By imposing different relaxation frequencies along a spatial dimension, it is possible to reconstruct the positions of the nuclei, providing a static structural view of brain matter (Hopf, 1985). fMRI is sensitive to functional changes in the brain because increased neuronal activity induces differences in magnetic properties between arterial (oxygen-rich) and venous (oxygen-poor) blood (Ogawa & Lee, 1990).

Increased neural activity is associated with increased local blood flow and oxygenation about 2 seconds later, peaking at around 6 seconds, before gradually returning to
baseline. This response to increased neural activity is known as the hemodynamic response function (HRF). Increased neural activity is associated with increased blood flow and oxygenation in nearby microvasculature that exceeds the local increase in energy demand, paradoxically leading to a decreased concentration of deoxygenated hemoglobin. Increased signal in fMRI is actually the result of this decreased concentration of deoxygenated hemoglobin, which is more magnetic and interferes more with the MRI signal than oxygenated hemoglobin. This activity-dependent increase in signal is referred to as blood-oxygen-level dependent (BOLD) signal (Huettel, 2010). Spatial reconstruction of BOLD responses can therefore show which regions of the brain are active during an experimental condition versus a control condition.

Temporal resolution of fMRI is typically on the order of 1-2 seconds, and spatial resolution on the order of 3-6 mm. Despite the speed of acquisition, the sluggish HRF that underlies the BOLD contrast further limits the ability to resolve rapid changes in neuronal activity. A conservative approach for dealing with limited temporal resolution is to present experimental manipulations in blocks (~30s) interspersed with rest conditions, which can be contrasted to reveal regions that were more active in the experiment than the rest condition (Maus, van Breukelen, Goebel, & Berger, 2010).

Event-related fMRI (ER-fMRI) approaches have been developed for isolating BOLD responses from individual trials presented at rapid intervals (Amaro & Barker, 2006; Friston, Zarahn, Josephs, Henson, & Dale, 1999). These approaches require either sufficient rest after each trial for the HRF to return to baseline, or the use of temporal jittering and trial randomization. If trials were presented at the same relative delay, then
it would be impossible to determine how much of the overlapping HRF’s are attributable to which event in time. ER-fMRI commonly uses deconvolution approaches to disentangle overlapping HRF activity, based on the fact that the measured signal is a convolution of HRF’s with regressors representing experimental stimuli and responses.

Deconvolution involves solving a matrix inverse problem, where the experimental design matrix consists of equations relating the sum of overlapping HRF’s at every time point to the measured signal. The ability to estimate activity in response to an event depends on the ability to invert the experimental design matrix. Therefore the experimental design should be optimized such that the equations for each time point should be maximally diverse; if HRF’s occurred at regular intervals, then these equations would all be the same, and the matrix non-invertible. This multiple regression problem is referred to as multicolinearity, because it is associated with a matrix of linear equations that are identical. In order to avoid multicolinearity in the design matrix, the experimental design must maximize the different relative onset times of experimentally-induced HRF’s (Dale, 1999).

In addition to separating rapidly presented trials, there are also methods for separating multiple phases of within-trial activity (Ollinger, Corbetta, & Shulman, 2001). Separating within-trial activity requires that events do not co-occur on every trial, and that they unfold in a known temporal sequence when they do co-occur. The studies presented in this thesis all involve the separation of within-trial activities during performance of an inhibition task, namely the SST.
The degree of multicollinearity of a design matrix can be checked prior to scanning with sample behavioral data by computing the correlation (r) between potentially overlapping regressors. A standard measure of colinearity is known as detection tolerance, calculated by $1-r^2$ (O’Brien, 2007). Detection tolerance values below 0.1-0.2 indicate that multicollinearity is a problem and that results of the matrix inversion will be unstable.

### 2.2 Executive function

Executive function is a broadly defined set of abilities including planning, working memory, sustained attention and inhibition (Aman, Roberts, & Pennington, 1998; Barkley, 1997; Welsh & Pennington, 1988). The proposed role of executive function is to handle situations that require more complex responses than those provided by relatively automatic psychological processes associated with learned behaviors and set schemas. Executive function is required in situations that require planning and decision making, correcting errors in performance, responding to novel contingencies requiring novel behavior, and the need to overcome more habitual behaviors (Norman & Shallice, 1986).

Executive functions are thought to be critically dependent on the prefrontal cortex (Welsh & Pennington, 1988). Functional imaging and lesion studies have helped delineate some of the division of labor in the frontal lobes (Alvarez & Emory, 2006; Stuss, 2011). The anterior cingulate cortex (ACC) is involved in the integration of experience with changing motivational states by monitoring performance and decision
making (Lezak, Howieson, & Loring, 1995). The dorsolateral prefrontal cortex (DLPFC) is more involved in holding information in working memory for planning, interrupting and changing task set, and abstract reasoning (Clark et al., 2008). The orbitofrontal cortex (OFC) is involved in the maintenance of appropriate behavior by assessing the subjective emotional value of experiences (Grabenhorst, Rolls, & Parris, 2008; Rolls & Grabenhorst, 2008).

2.3 Inhibitory control

Inhibitory control is thought to be a central component of executive function. Inhibitory control involves the suppression of prepotent behaviors (Aron, 2011; Band & van Boxtel, 1999; Bedard et al., 2002). A prepotent behavior is one that has been reinforced to the point of automaticity. The central role of inhibitory control in executive function arises from the fact that relatively automatic processing must be suppressed or delayed so that executive processing can take place.

Inhibitory control refers to three interrelated processes: suppression of attention to distracting stimuli, inhibiting prepotent responses, and interrupting responses that have already been initiated (Barkley, 1997). The ability to suppress distracting information has traditionally been studied using Stroop tasks (Gerstadt, Hong, & Diamond, 1994; Stroop, 1935a, 1935b), in which subjects must override a prepotent response tendency (e.g. presenting the word “blue” in a red-colored font, and reporting the color of the font, rather than the word itself). Non-verbal versions of Stroop tasks have also been
developed in order to avoid confounds associated with differences in reading ability (Gerstadt et al., 1994).

The ability to inhibit prepotent response tendencies has been studied using Go-NoGo tasks (Trommer, Hoeppner, Lorber, & Armstrong, 1988). In these tasks, subjects respond to the majority of stimuli, such as presentation of a shape on a computer screen, but must withhold their responses on a minority of trials in which a different shape is presented, or the same shape in a different color.

The ability to interrupt responses that have already been initiated is studied using a stop signal task (Band, van der Molen, & Logan, 2003; Logan, Cowan, & Davis, 1984; Logan, 1982; Logan, 1994; Verbruggen & Logan, 2008c), in which the signal to stop is presented at some delay after presentation of the response stimulus. The stop signal task is thought to be analogous to real life situations requiring the rapid and accurate execution of a behavior and the need to occasionally stop that behavior due to changes in external conditions or internal motivational states (Band & van Boxtel, 1999; De Jong, Coles, Logan, & Gratton, 1990; Logan et al., 1984; Logan, 1982; Logan, 1994; Verbruggen & Logan, 2008c, 2009). The studies presented in this thesis use the stop signal task, which will now be explained in greater detail.

2.4 The stop signal task (SST)

The SST involves a primary forced reaction time task and a secondary stop task. In the studies presented here, the primary task consists of presenting either the letter “X”
indicating that the subject should make a button press with their left thumb, or the letter "O" indicating that the subject should make a button press with their right thumb. The goal of the primary task is to respond as quickly and accurately as possible to left- and right-hand response cues. On an unpredictable subset of trials (33% of trials in the studies presented here) the response stimulus is followed after some delay by the presentation of a stop stimulus, indicating that no button should be pressed on that trial. In laboratory versions of the SST the stop signal is an auditory tone, whereas imaging versions tend to use a change in screen color (black to red) due to the high noise levels during fMRI scanning (Chevrier, Noseworthy, & Schachar, 2007; Chevrier & Schachar, 2010; Hu & Li, 2012).

The delay between presentation of the response stimulus and presentation of the stop stimulus, referred to as the stop signal delay, is initially set to 250 ms. The stop signal delay decreases by 50 ms when subjects fail to stop, making it easier to stop on the next stop trial, and increases by 50 ms when subjects successfully stop, making it more difficult to stop on the next stop trial. This dynamic tracking algorithm ensures that only half of the stop trials presented can be stopped on average, whereas responses on the other half of stop trials will be too far under way and cannot be stopped (Logan, 1994).

Based on the race model of inhibition explained below (Boucher, Palmeri, Logan, & Schall, 2007; G D Logan et al., 1984; Verbruggen & Logan, 2009), the SST allows for indirect estimation of the speed of the stop process, referred to as the stop signal reaction time (SSRT), despite the absence of any overt behavior on successful stop trials.
All of the fMRI studies presented here use the SST paradigm to separate various within trial activities reflecting preparation and reinforcement learning. The sequences of trial types as presented in the fMRI version of the SST used here are portrayed in Figure 2.1, along with the event types and regressors used in the deconvolution analyses. Further description of task timing can be found in sections 3.3.3 and 4.3.3.

Figure 2-1 Trial types and regressors in the SST. a) Sample sequence of 14 trials, which are followed by a 17.5 s rest every minute. b) Imaging model for isolating preparation activity. c) Model for separating error detection activity.
2.5 SSRT and the race model of inhibition

In the race model of inhibition proposed by Logan et al. (Band et al., 2003; De Jong et al., 1990; Logan et al., 1984), the outcome of a given stop trial is described by the outcome of a race between two opposing processes, namely a go process and a stop process. The go process is modeled as beginning with the presentation of go stimuli followed by stimulus recognition, response choice, and the preparation and execution of a motor response. The stop process is modeled as beginning with the presentation of the stop stimulus. The relative finishing times of these competing processes determines the winner of the race and therefore the outcome of a given stop trial. If the go process finishes before the stop process, then that trial will contain a response. If the stop process finishes first, then there will be no response on that trial.

The dynamic tracking algorithm in the SST adjusts the stop signal delay based on the speed of the go-process, preventing subjects from simply withholding responses until stop signals appear before deciding to respond. The variability in stop signal delay also permits the estimation of the speed of the stop process, called SSRT. By plotting the probability of inhibition as a function of stop signal delay, one can calculate SSRT based on the slope of this function, with steeper slopes indicating faster inhibitory control (Band et al., 2003; De Jong et al., 1990; Logan et al., 1984). The SSRT can be estimated with a high degree of accuracy by subtracting the mean stop signal delay from the mean go reaction time. An important feature of the SST is that it is capable of separating variability in the go process from variability in the stop process. An important assumption of the race model is that the relative finishing times of go and stop processes
are stochastically independent, which means that variability in the go process does not affect the stop process and vice-versa. The independence of the finishing times of go and stop processes in the SST has been confirmed with extensive behavioral and physiological studies (Band et al., 2003; Logan et al., 1984; Logan, 1982; Logan, 1994; Verbruggen & Logan, 2009).

The independence of going and stopping in the SST has often been over-interpreted to mean that there is no preparation to stop before the appearance of stop signals. However, the fact that go- and stop-processes do not interact does not necessarily mean that they cannot be independently prepared prior to the appearance of go- and stop-stimuli. The first study presented here (Chapter 3) involves the separation of two phases of preparatory activity, both of which contain activities that affect the speed and accuracy of stopping.

2.6 fMRI of inhibitory control

Neuroimaging studies of inhibitory control have subtracted activity on unsuccessful stop trials from that on successful stop trials based on the assumption that they only differ in terms of inhibition, or subtracted go trials from successful stop trials in attempts to control for stop trial activities that precede the stop process (Chikazoe, Konishi, Asari, Jimura, & Miyashita, 2007; Rubia, Smith, Brammer, & Taylor, 2003; Simmonds, Pekar, & Mostofsky, 2008). Both of these approaches have consistently found inhibition related activity in the right inferior frontal gyrus (IFG) and other fronto-parietal, temporal and
striatal regions (Chevrier et al., 2007; Jaffard et al., 2008; Li, Yan, Sinha, & Lee, 2008). Although multiple regions have been shown to activate during inhibitory control, imaging, meta-analytic and lesion deficit studies point to the right IFG and caudate nucleus as the primary neural substrates of inhibitory control (Chambers et al., 2007; Chikazoe, Konishi, Asari, Jimura, & Miyashita, 2007; Hampshire, Chamberlain, Monti, Duncan, & Owen, 2010; Hughes, Johnston, Fulham, Budd, & Michie, 2013; Iversen & Mishkin, 1970; Aron, Fletcher, Bullmore, Sahakian, & Robbins, 2003; Aron & Poldrack, 2006; Aron, Robbins, & Poldrack, 2004; Aron, 2007, 2011).

2.7 Preparation

Goal-directed behavior is generally more efficient when imperative stimuli are preceded by warning cues than when such cues are absent (Jennings & van der Molen, 2005; Sanders & Wertheim, 1973). Warning stimuli facilitate rapid processing of sensory, cognitive and motor aspects of behavior. Due to their rapid nature, preparatory activities have largely been studied using event related potential (ERP) approaches. Anticipatory behavior is reflected in two important brain potentials: the readiness potential (RP) reflects the timing of upcoming voluntary movements, and the stimulus preceding negativity (SPN) reflects anticipatory attention for upcoming stimuli (Böcker, Baas, Kenemans, & Verbaten, 2001; Brunia & Damen, 1988; Brunia, 1993; Gemba & Sasaki, 1984; Mnatsakanian & Tarkka, 2002; Niki & Watanabe, 1979; van Boxtel & Brunia, 1994). In order to separate these waves, electrophysiological studies have used a time-estimation task: a warning stimulus is followed by an imperative stimulus indicating
that the subject should respond, which is then followed by another feedback stimulus indicating whether the subject responded early, late or within the desired time window (Brunia & Damen, 1988; Jennings & van der Molen, 2005).

In time estimation tasks, the RP is greatest over the motor cortex contralateral to a finger movement. Single cell recordings in monkeys (Gemba & Sasaki, 1984) have demonstrated neural sources for the RP in primary, pre- and supplementary motor areas, as well as primary sensory cortex. Lesion evidence has also shown that the cerebellum is necessary for the emergence of the RP, while evidence from Parkinsonian patients also implicate the basal ganglia in the RP (Dick et al., 1989; Praamstra, Meyer, Cools, Horstink, & Stegeman, 1996). SPN’s occur before alerting and feedback stimuli, and involve frontal cortex, parietal cortex and insula, with centro-parietal maxima (Böcker et al., 2001; Brunia & Damen, 1988; Brunia, 1993; van Boxtel & Brunia, 1994).

Understanding preparatory processing is particularly important in tasks that engage motor response inhibition such as the SST. This is because the race model of inhibitory control does not address how go and stop processes are initiated beyond the presentation of go and stop stimuli (Boucher et al., 2007). A few studies have indirectly inferred preparation-related activity in the SST, implicating fronto-parietal networks and presupplementary motor area (Boucher et al., 2007; Chikazoe et al., 2009; Connolly, Goodale, Menon, & Munoz, 2002; Hu & Li, 2012; Jaffard et al., 2008; Postle, Zarahn, & D’Esposito, 2000). However, no previous fMRI study has explicitly separated preparation from other processes in the SST.
Variability in preparatory processes in stop tasks, when imaging groups with deficient executive control, have resulted in inconsistent activity differences compared to normal individuals (e.g. increases (Schulz, Newcorn, Fan, Tang, & Halperin, 2005), decreases (Mostofsky et al., 2006; Shafritz, Marchione, Gore, Shaywitz, & Shaywitz, 2004), or both (Silk, Vance et al. 2005; Konrad, Neufang et al. 2006)). Preparation is certainly involved in the ongoing trade-off between go- and stop-processes and can consequently affect experimental measures of inhibitory control. It is important to isolate preparation-related activities that precede response and stop phases in order to properly understand how inhibitory control unfolds over time.

2.8 Conflict, errors and performance monitoring

Another important part of inhibitory control is the ability to detect and adjust to errors, referred to as performance monitoring (Holroyd & Coles, 2002a). Errors in the stop task are generally followed by slower responses in order to increase the probability of stopping. Errors in a wide array of tasks are known to activate the ACC (Debener et al., 2005; Gehring & Fencsik, 2001; Gehring, Liu, Orr, & Carp, 2011; Hajcak, Moser, Yeung, & Simons, 2005; Herrmann, Römmler, Ehlis, Heidrich, & Fallgatter, 2004; Holroyd & Coles, 2002b; Meyer, Riesel, & Proudfit, 2013; Riesel, Weinberg, Endrass, Meyer, & Hajcak, 2013; Yeung, Botvinick, & Cohen, 2004). Activity in the dorsal ACC on errors is thought to index the impact of midbrain dopamine responses to prediction errors, causing a reallocation of prefrontal resources based on changing conditions (Holroyd & Coles, 2002a; Kerns et al., 2005; Orr & Hester, 2012). Although early
accounts of dorsal ACC function focused on sensitivity to errors and the monitoring of conflicting response tendencies (Bioulac, Michelet, Guehl, Aouizerate, & Burbaud, 2005; Botvinick, Braver, Barch, Carter, & Cohen, 2001; Botvinick, 2007; Braver, Barch, Gray, Molfese, & Snyder, 2001; Gehring & Fencsik, 2001; Swick & Turken, 2002; Ullsperger & von Cramon, 2001; van Veen, Cohen, Botvinick, Stenger, & Carter, 2001; Yeung et al., 2004), recent evidence indicates that activity in the dorsal ACC is consistent with an ideal Bayesian observer, involved in identifying prefrontal resources required based on changing task set (Ide, Shenoy, Yu, & Li, 2013).

In addition to task-related processing in networks that deal with the details of errors, errors also invoke neuromodulatory processes that alter excitability and plasticity in diffuse brain networks, which are crucially dependent on midbrain dopamine neurons as well as other neuromodulatory systems. These effects have been studied using prediction errors in the context of reward and reinforcement learning tasks, which involve the presentation of pleasant and unpleasant stimuli, or overt rewards and punishments. Prediction errors are defined as the difference between an actual and expected outcome. By controlling the expectation and delivery of reward, it is possible to identify structures where the level of activity corresponds with changes in variables in models of reinforcement learning based on prediction error signals. The level of activity in the OFC and ventral striatum have been shown to respond to the magnitude of prediction errors, and activity in the dorsal striatum has been shown to respond to the occurrence but not magnitude of prediction errors (Knutson, Taylor, Kaufman, Peterson, & Glover, 2005; O’Doherty, 2004).
Midbrain dopamine neurons have been shown to alter their firing characteristics based on the occurrence and magnitude of prediction errors rather than context-dependent information about the nature of the error. Positive prediction errors briefly increase dopamine activity, which reinforces connectivity in currently active networks by the long term potentiating effects of dopamine on target synapses. By contrast, negative prediction errors briefly (~100 ms) deactivate midbrain dopamine neurons, interrupting and negatively reinforcing ongoing processing. Positive prediction errors strengthen associations, whereas negative prediction errors lead to extinction. Therefore, reinforcement learning signals from midbrain dopamine neurons, reviewed in the following section, play an important role in the detection of errors and the subsequent adjustment of functional networks and corresponding behavior.

2.9 Dopamine, reward and reinforcement learning

2.9.1 Reward

Our brains are built for identifying, learning about and pursuing rewards in order to propagate our genes (Dawkins, 1976; Schultz, 2015). Rewards consist of external and internal attributes. Rewards have external sensory properties that are used to identify physical attributes. Gustatory rewards contain nutrients necessary for survival that are directly identifiable by our senses. Sexual attraction provides another primary reward, which serves to maintain activities necessary for mating, producing and caring for offspring. All other rewards serve to reinforce the primary rewards of survival, selection and reproduction. Other rewards may be context-dependent, requiring more cognitive
identification. The identification of reward properties serves to harness attention and pursuit, which can be measured in terms of behaviors such as orienting responses. Attention affords the prioritization of behavior for the efficient pursuit of rewards.

The internal component of rewards, namely ‘value,’ is more elusive and difficult to objectively quantify, as it is subjective by its nature. However, value is precisely what we understand as reward. Value is the result of internal preferences that are individual, but individual differences can be objectively assessed by measuring differences in effort, approach and decisions. The pursuit of more complex goals requires more complex cognition and learning about secondary reinforcers, which predict rewards but are not rewards themselves. Secondary rewards generate cognitive top-down attention that requires memory and learning, and lead to the same responses in dopamine pathways as explicit rewards.

Brains that have developed more over evolution are able to process more and more abstract secondary rewards. These might involve complex sensory properties such as in the appreciation of art or fine wine, which can lead to more sophisticated tastes for more richly nutritious food and opportunities to aid sexual selection. These can also consist of internal decisions or actions that provide no immediate primary reward, but which can induce novel behaviors that can lead to novel sources of primary rewards in the future (Barto, 2004; Meijer & Robbers, 2014; Singh, Lewis, Barto, & Sorg, 2010). The lack of narrow focus associated with more complex cognition has led to the development of skills, such as art, science and engineering, which have led to unexpected progress in human society.
The most immediate function of primary rewards is for the maintenance of homeostasis, which is greatly aided by learned hunger and thirst reactions (Hebb, 1949). These responses can readily adapt to changes in internal chemical balances, for example causing salt water to be palatable in sodium-deprived states (Robinson & Berridge, 2013). More complex reinforcement of learning allows for more advance notice of potential imbalances, and the deployment of economic choices for obtaining necessary objects and conditions for survival. Homeostatic balance therefore imposes primal constraints on the development of brains that value objects and situations, and the interaction between thinking, behavior, and the pleasurable effects of rewards.

A second function of reward is to propagate our genes by obtaining mates and caring for children, regulated by hormones that make sex and attachment pleasurable. This brings us to the third function of reward, namely ‘pleasure.’ Homeostasis only explains the response to a limited set of primary rewards, whereas pleasure affords an explanation of secondary rewards, as well as some objective definition of ‘value.’ As the primary effect of rewards, pleasure determines the attention and effort we will exert in order to receive it. Secondary reinforcers can also develop stronger behavioral responses than primary rewards, exemplified in monkeys that will forego liquid rewards in order to view sexual pictures (Deaner, Khera, & Platt, 2005). The involvement of dopamine in reward originated from the study of approach behaviors elicited by stimulation of regions involving midbrain dopamine neurons (Corbett & Wise, 1980).

Other innate mechanisms might explain other aspects of pleasure, such as parental attachment or the preference in babies for viewing human faces over similar non-face
stimuli (Valenza, Simion, Cassia, & Umiltà, 1996). It has been proposed that these kinds of innate preferences can aid in the development of attention to particularly important stimuli such as social cues (Schultz, 2015).

Sensory properties alone are not sufficient for explaining the effects of rewards on behavior. Rewards can only be explained by the subjective value attributed to them. Value is reflected in preferences, which are measurable in terms of behaviors such as choices and reaction times. Rewards cannot be defined externally, but are based completely on internal brain activity that relates rewards to evolutionary survival (Singh et al., 2010).

Dedicated processing of reward information in the brain involves a limited set of structures including midbrain dopamine neurons, striatum, amygdala and orbitofrontal cortex. Modulation of activity based on reward properties are also found in most other parts of the brain, usually in association with relevant sensory, cognitive or motor attributes. However, rewards do not have dedicated receptors like our sensory systems. In order to obtain a reward signal, the brain must process heterogeneous and multimodal information in distributed sensory and cognitive systems. The lack of dedicated receptors, although invoking an initial cost in speed of processing when learning about rewards, allows for much greater generalization and adaptation. We only want to focus on the color or shape of a reward if that enhances the pleasurable effects. This mechanism maximizes the number of objects that can be rewarding, and increases the odds of obtaining rare rewards, without the need for dedicated hardware for different situations based on their likelihood.
2.9.2 Reinforcement learning and goal-directed behavior

The diverse rewards that we are able to pursue are the result of learning. Learning and adaptation allow for survival in diverse environments amidst changing conditions, which would otherwise require dedicated biological resources to address. Pavlovian conditioning involves identifying a stimulus such as a bell with a reward (Rescorla, 1967). Pavlovian conditioning is responsible for the development of stimulus-driven attention, which directs top-down control that can direct behavior toward the pursuit and acquisition of the reward. Pavlovian conditioning allows for the attention-harnessing effects of pleasurable stimuli to extend to the learning of higher order reinforcers that are used to make complex economic choices.

Operant conditioning involves pairing an existing behavior with a stimulus that predicts a reward (Tuckerman, 2003). Learning allows the reward-predicting stimulus to evoke a conditioned response, such as salivation or approach behavior, when the behavior is followed by a reward. When a reward is delivered after a pigeon randomly pecks at a lever, the pecking behavior will increase. The generalization afforded from learning by operant conditioning can have unexpected drawbacks, such as the superstitious conditioning of random behaviors that are paired with the random delivery of rewards (Justice & Looney, 1990; Neuringer, 1970).

Goal-directed behavior requires the updating of information about the values of rewards when circumstances change the availability or nature of rewards. Goal-directed behavior requires some form of representation of the outcome during performance of the behavior, and of the causal relation between the behavior and the outcome (Dickinson &
Balleine, 1994). Goal-directed behaviors can become habits after repeated and stereotyped performance of the same instrumental behavior, which become inflexible stimulus-response actions. The formation of habits frees up our mental resources, allowing us to move on to more complex behaviors and reward contingencies.

2.9.3 Prediction error learning

In order for reward values to be updated based on changing conditions, it is necessary to compare changes in conditions with our past predictions of value. This process is formalized in theories of conditioning using the concept of a reward prediction error. Prediction errors are defined mathematically as the difference between the actual reward that is received, ‘A,’ and the reward that was expected, ‘E’ (i.e. \( PE = A - E \)). Animal learning theories describe how values are updated based on prediction errors (Rescorla & Wagner, 1972). In its basic form, predictions are updated on trial \( n+1 \) by adding the previous prediction on trial \( n \) multiplied by a constant \( \alpha \), referred to as the learning rate (i.e. \( E(n+1) = E(n) + \alpha \cdot PE(n) \)). In this framework, all conditioning is based on error-driven learning. When the learning rate is equal to 1, the new prediction is updated to fully reflect the new reward contingency in a single trial. However, learning rates are generally smaller than 1 in order to make more probabilistic predictions, leading to asymptotic learning (den Ouden, Friston, Daw, McIntosh, & Stephan, 2009).

Outcomes leading to rewards that are greater than expected constitute positive prediction errors, which strengthen associations and behaviors that led to the reward. By contrast, outcomes leading to rewards that are less than expected constitute negative prediction errors. Negative prediction errors weaken behaviors and the attentional salience of
learned reward-predicting cues, which can result in extinction of unnecessary behaviors. When outcomes exactly match predictions, the prediction error is zero, the prediction itself becomes stable, and no more reinforcement learning occurs. In this way, prediction errors facilitate the maximization of rewards (positive prediction errors) and the avoidance of undesirable outcomes (negative prediction errors).

Extensions of basic Pavlovian and operant conditioning theories can help to deal with the non-stationary conditions associated with natural environments. When conditions change rapidly, we need to learn more quickly, which can be modeled by updating the learning rate in proportion to the magnitude of the latest prediction error (Pearce & Hall, 1980). Alternatively, similar reinforcing stimuli might be drawn from different distributions, which can lead to the over-learning of special cases that might not be representative of future instances. Generalizing prediction errors to samples from different distributions can be improved by normalizing the learning rate by the standard deviation of the appropriate distribution (Preuschoff & Bossaerts, 2007).

Theories of operant conditioning have also been extended to include higher-order reinforcers that precede the appearance of primary rewards. These models, referred to as temporal difference models, not only update learning parameters after each trial, but also within a given trial, in order to capture the most relevant predictive aspects of goal-directed behavior (Sutton & Barto, 1981). Prediction errors in these models are updated at every step in a sequence of learned predictors, based on the unique and unambiguous information that each stimulus lends to predicting the outcome. Temporal difference models are capable of using basic operant conditioning principles to learn complex tasks.
involving sequences of actions in order to obtain rewards (Barto, Sutton, & Anderson, 1983; Tesauro, 1994).

2.9.4 Dopamine prediction error activity

The unexpected delivery of liquid or food rewards in monkeys results in a rapid (<100 ms) and brief (<200ms) activation of the majority (60-75%) of midbrain dopamine neurons (Ljungberg, Apicella, & Schultz, 1992; Satoh, Nakai, Sato, & Kimura, 2003). It has been shown that this activity is consistent with the intensity of prediction error signals expected from animal learning theories. The dopamine response codes the difference between actual and expected rewards. Rewards that are better than expected lead to increased firing, rewards that are worse than expected lead to depressed activity, and rewards that are exactly as expected evoke no response (Bayer & Glimcher, 2005; Fiorillo, Tobler, & Schultz, 2003; Ljungberg, Apicella, & Schultz, 1991; Mirenowicz & Schultz, 1994; Pan & Hyland, 2005; Schultz, Apicella, & Ljungberg, 1993; Schultz, Dayan, & Montague, 1997; Schultz, 1998).

Dopamine responses have also been found to exhibit necessary properties for temporal difference learning (Day, Roitman, Wightman, & Carelli, 2007; Montague, Dayan, & Sejnowski, 1996) and adjusting to non-stationary reward contingencies, and samples drawn from different distributions (Fiorillo et al., 2003). Phasic responses of dopamine neurons have been demonstrated to exhibit the properties of prediction errors required by formal theories of animal learning (Waelti, Dickinson, & Schultz, 2001). Therefore, it appears that dopamine neurons in some way encode prediction error activity that matches the reinforcement term in models of reinforcement learning.
2.9.5 Negative prediction errors

Chapters 4 and 5 of this thesis deal with the identification of BOLD responses in dopamine pathways on stop task errors, which are a form of negative prediction error. The suppression of dopamine activity on negative prediction errors occurs in the context of a low baseline level of activity, and therefore has less dynamic range than positive prediction errors (Bayer & Glimcher, 2005; Fiorillo et al., 2003). However, the cessation of dopamine activity for longer periods of time could have stronger effects on postsynaptic neurons, which appears to compensate for this limited dynamic range; voltammetry measured dopamine concentrations in response to negative prediction errors appear to be symmetric with those on positive prediction errors (Bayer & Glimcher, 2005). Therefore, phasic dopamine activity appears to encode a bidirectional reward prediction error response to rewards as well as higher order reinforcers, and to negative prediction errors as well as positive prediction errors. The novel approach to imaging negative prediction errors in the SST outlined in Chapter 4 (Chevrier & Schachar, 2010) revealed novel patterns of reinforcement activity in reward pathways that have not been available to previous approaches using explicit reward or reinforcement learning tasks.

2.9.6 fMRI of reward prediction errors

The reward prediction error response conveyed by midbrain dopamine neurons is detectable using the BOLD response measured by fMRI during performance of tasks involving prediction errors of known magnitude. A large number of studies (i.e. hundreds) have identified reward related activity in the main reward related brain regions (e.g. (Knutson, Adams, Fong, & Hommer, 2001; Thut et al., 1997)). The fMRI
response to reward prediction errors in the ventral striatum has been referred to as “probably the most solid reward response in the brain” (Schultz, 2015). This response increases with dopamine agonists and decreases with dopamine antagonists, demonstrating that dopamine has both necessary and sufficient roles in fMRI responses to reward in the ventral striatum (Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006). fMRI responses in ventral striatum, orbitofrontal and anterior cingulate regions have also been shown to vary with pleasantness and momentary happiness ratings of various reward prediction errors (Rolls, Grabenhorst, & Parris, 2008; Rutledge, Skandali, Dayan, & Dolan, 2014). Therefore, fMRI is capable of reliably measuring certain aspects of the midbrain reward prediction error response. The whole-brain view afforded by fMRI affords some perspective on the interaction between networks that are involved in reward prediction errors and the cortical task-related networks that are reinforced by errors.

2.9.7 Interactions between reward and task-related networks

True goal-directed behavior involves model-based learning which in addition to reward processing conveyed by midbrain dopamine pathways, requires distributed cortical resources for processing task-related details. Model-based learning involves the acquisition and maintenance of models, and the influence of models on the predictions of reinforcement learning mechanisms. Our models of the world involve details associated with contexts, environments and reward contingencies (Erhel & Jamet, 2013; Lecun, Chopra, Hadsell, Ranzato, & Huang, 2007; Pivec, Dziabenko, & Schinnerl, 2003). The acquisition and maintenance of these models involves cortical resources, whereas reward information is integrated in midbrain dopamine neurons once these
models are established. Even in higher-order cognitive tasks, the phasic dopamine response does not encode any information relating to task-specific details, but does reliably encode the reward prediction error (de Lafuente & Romo, 2011; Ljungberg et al., 1991; Satoh et al., 2003; Schultz et al., 1993; Takikawa, Kawagoe, & Hikosaka, 2004).

Previous research has rigorously studied reinforcement activities related to prediction error magnitude in dopamine pathways, or activities in task-related networks related to various experimental manipulations of cognitive variables. The separation of error detection from post-error slowing presented in Chapter 4 of this thesis provides new perspectives on how prediction error activities in midbrain dopamine networks affect distributed task-related networks. The results suggest that this approach successfully isolated a whole-brain view of the event detection response from the subsequent evaluative response for the first time.

2.9.8 Dopamine event detection response

The initial reward prediction error response measured from single cell recordings in midbrain dopamine pathways is very brief (< 100 ms), and appears to reflect the identification of salient outcomes before the value of the outcome can be identified (Kobayashi & Schultz, 2008; Nomoto, Schultz, Watanabe, & Sakagami, 2010; Tobler, Dickinson, & Schultz, 2003). This event detection response is followed by the reward magnitude related response, with positive prediction errors leading to increased activity and negative prediction errors leading to suppressed activity. The transition from the event detection response to the prediction error response depends on the ease with which
the value of the outcome can be identified. In tasks that use stimuli that are difficult to identify, this delay can be up to 200 ms, which clearly separates the event detection response from the prediction error response (Nomoto et al., 2010).

In tasks that do not require saccadic eye movements to identify error magnitude this transition can take less than 50 ms, causing these two responses to combine together into a single response (Kobayashi & Schultz, 2008; Tobler et al., 2003). Chapter 4 of this thesis presents a method for separating error detection from post-error slowing, which is related to prediction error magnitude (Chevrier & Schachar, 2010). The task we used was a stop signal task, in which the identification of errors does not require saccadic eye movements and therefore can be rapidly identified (< 50 ms). Despite the poor temporal resolution of fMRI, the design nonetheless permitted the separation of event detection from reward prediction error activity, which could not be separated even in single cell recording approaches. The event detection response has been proposed to enhance the impact of the prediction error response in reward learning (Mackintosh, 1975; Pearce & Hall, 1980; Wolfram Schultz, 2015). The whole-brain view of error detection activities afforded by fMRI presented in Chapter 4 offers new insights into how the brain might use prediction error activities to adjust functional connectivity in cortical and subcortical networks that carry out task-related processing. Some interpretation of the distributed neural effects of the error detection response is possible based on the known powerful effects of phasic dopamine activity on the excitability and plasticity of their postsynaptic targets.
2.9.9 Neuromodulatory effects of dopamine

The limited set of structures identified during error detection using fMRI (Chevrier & Schachar, 2010) is consistent with the limited topography of the projections of midbrain dopamine neurons (Gauthier, Parent, Lévesque, & Parent, 1999; Jimenez-Castellanos & Graybiel, 1987; Matsuda et al., 2009). The efferent projections of midbrain dopamine neurons on the dendritic spines of their striatal and cortical targets form synaptic triads with other cortical inputs that facilitate learning and memory (Calabresi, Picconi, Tozzi, & Di Filippo, 2007; Dani & Zhou, 2004; Jay, 2003). This arrangement accomplishes Hebbian plasticity based on selectively active corticostriatal transmission (Freund, Powell, & Smith, 1984; Goldman-Rakic, Leranth, Williams, Mons, & Geffard, 1989; Reynolds, Hyland, & Wickens, 2001; Schultz, 1998). These dopaminergic projections play important roles in the long term potentiation and depression of their targets in the striatum, prefrontal cortex and hippocampus (Calabresi et al., 2000; Gurden, Takita, & Jay, 2000; Kerr & Wickens, 2001; Otani, Blond, Desce, & Crépel, 1998; Otmakhova & Lisman, 1996).

Although a great deal is known about the plasticity effects of dopamine neurons on their postsynaptic targets at a cellular level, no previous study had provided a whole-brain view of error detection followed by prediction error related activities. In Chapter 4, I discuss the potential implications of the whole-brain view of error detection and prediction error activities in terms of their known plasticity effects. In Chapter 5, I discuss the potential implication of altered error detection and prediction error magnitude activities in adolescents with attention deficit hyperactivity disorder (ADHD).
2.10 Deficient executive function in ADHD

ADHD is a common and persistent neurodevelopmental psychiatric disorder characterized by significant problems with executive functions such as sustained attention, inhibitory control and error monitoring (Biederman & Faraone, 2005; Sroubek, Kelly, & Li, 2013). These problems lead to learning difficulties and impulsive behaviors that are not appropriate for one’s age (Childress & Berry, 2012; Diamond, 2013). Several models have been proposed to explain the etiology of ADHD.

Some models of ADHD have proposed that inhibitory control is the primary deficit, which prevents effective functioning of other aspects of executive control (Barkley, 1999). Without the inhibition of automatic processing, there is no delay in which executive processing can occur. ADHD is associated with deficient inhibitory control as measured by the SSRT (Crosbie et al., 2013; Crosbie & Schachar, 2001; Lipszyc & Schachar, 2010; Soreni, Crosbie, Ickowicz, & Schachar, 2009), and decreased activation in inhibition networks during performance of stop tasks (Dickstein, Bannon, Xavier Castellanos, & Milham, 2006; Mulligan et al., 2011; Paloyelis, Mehta, Kuntsi, & Asherson, 2007; Rubia et al., 2010; Schulz et al., 2004; Tamm, Menon, Ringel, & Reiss, 2004; Vaidya et al., 2005).

Other models of ADHD have proposed that error processing is the primary deficit in ADHD (Sagvolden, Johansen, Aase, & Russell, 2005), such that the history of reinforcement does not lead to appropriate changes in behavior. Diminished error processing in ADHD has been characterized by greater response variability, increased error rates in speeded tasks and diminished response slowing following errors.
Castellanos-Ryan et al., 2014; Schachar et al., 2004; Sergeant & Van der Meere, 1988). ERP studies have found reduced error-related negativity (ERN) in ADHD (Liotti, Pliszka, Perez, Kothmann, & Woldorff, 2005; Shen, Tsai, & Duann, 2011; Shiels & Hawk, 2010; van Meel, Heslenfeld, Oosterlaan, & Sergeant, 2007). The ERN is a midline wave that follows the detection of errors and is thought to index the impact of midbrain reinforcement signals in the ACC (Gehring et al., 2011; Holroyd & Coles, 2002b). fMRI studies have found ADHD to be associated with reduced error signaling in prefrontal, parietal, cingulate and thalamic regions (Hester et al., 2012; Rubia, Smith, et al., 2009; Vasic et al., 2012).

Regardless of the primary executive deficit, there is convergent evidence that ADHD symptoms are the result of atypical reinforcement learning due to impairment in the midbrain dopamine system (Bellgrove, Hawi, Gill, & Robertson, 2006; Luman, Oosterlaan, & Sergeant, 2008; Sagvolden et al., 2005; Tripp & Wickens, 2009; Volkow et al., 2009). Stimulant medications that increase extra-cellular dopamine levels improve behavior in ADHD subjects (Castells et al., 2011; Wigal, 2009). The use of stimulants in ADHD subjects has also been shown to normalize brain activation on errors (Groen, Mulder, Wijers, Minderaa, & Althaus, 2009; Rubia, Halari, Mohammad, Taylor, & Brammer, 2011), but these effects have not been found in brain regions that are richly connected with dopamine neurons.

Studies using positron emission tomography (PET) have shown that ADHD subjects show reduced binding potential for dopamine transporter in midbrain regions where dopamine neurons originate, resulting in altered dopamine signaling (Jucaite, Fernell,
Lesions in parts of the striatum that receive dopamine input increase the risk of ADHD traits (Max et al., 2002). ADHD traits are also associated with dopamine manipulations in animal models of ADHD (Leo et al., 2003; Miller et al., 2012; Sontag, Tucha, Walitza, & Lange, 2010; van der Kooij & Glennon, 2007; Viggiano, Vallone, Ruocco, & Sadile, 2003). However, the causal relationship between ADHD symptoms and altered dopamine remains uncertain; results from dopamine-depleted mice would predict reduced dopamine function associated with ADHD symptoms, whereas results from dopamine transporter knock-out mice would predict increased dopamine function (Ohno, 2003).

Reward task studies have reported abnormal activity in the ACC and striatum during reward anticipation and delivery in ADHD subjects (Furukawa et al., 2014; Hauser et al., 2014), and a failure of OFC to encode prediction error magnitude as in healthy control subjects (Wilbertz et al., 2012). ADHD subjects have also been shown to require higher levels of reward to engage similar levels of motivation and of neural activity reflecting reward magnitude (Umemoto, Lukie, Kerns, Müller, & Holroyd, 2014). However, no previous study has separated error detection from post-error adjustments related to error magnitude in individuals with ADHD. The end goal of the studies presented here is to identify activities in structures that directly project to and receive from midbrain dopamine neurons during error detection and post error adjustments in adolescents with and without ADHD. This approach revealed novel activity differences in these pathways that validates, and offers new perspectives on previous theories of altered dopamine processing in ADHD.
2.11 General hypotheses

This thesis presents the development of fMRI methods for separating within-trial activities in the SST in healthy adults, and the application of these methods to the study of adolescent ADHD. The goal is to provide a novel view of how top-down control networks come into play at the beginning of every trial, how they are reinforced based on errors, and how ADHD adolescents differ from healthy adolescents.

Chapter 3 describes a method for separating two rapidly unfolding (<1s) phases of preparatory activity that precede the appearance of stop signals. These phases of activity correspond to the presentation of initial warning stimuli, and response phases that follow, both of which precede the appearance of stop signals. Based on previous ERP research outlined in the introduction of Chapter 3, we hypothesized that warning- and response-phases will involve: left followed by right parietal activity; deactivation followed by activation of primary motor cortex; posterior followed by fronto-posterior activity; and cortical followed by cortical-subcortical activity. We performed regression analyses and hypothesized that both of these phases will contain activities that have distinct influences on the speed and accuracy of going and of stopping, providing insight into factors that influence the speed-accuracy trade-off that must occur on every trial, but are not addressed in current models of inhibitory control.

Chapter 4 describes a method for separating activity during error detection from that during post-error adjustments (i.e. response slowing). Given that post-error adjustments should be related to prediction error magnitude, we hypothesized that greater post-error
slowing after inhibition errors should deactivate the ventral striatum and caudal OFC, which are known to encode prediction error magnitude. Further, given that error detection should occur regardless of the magnitude of the error, we hypothesized that error detection should reflect the event-detection response that has been posited based on observations from single cell recordings in midbrain dopamine networks.

Chapter 5 presents the separation of error detection and post-error slowing activities in ADHD and typically developing (TD) adolescents. Based on previous evidence of abnormal reinforcement signaling in dopamine pathways in ADHD, we hypothesized that ADHD subjects would exhibit a diminished event detection response on error detection, and a diminished or opposite response (activation instead of deactivation) in ventral striatum and caudal OFC during post-error slowing.
3 Dissociating two stages of preparation in the stop signal task using fMRI

doi:10.1371/journal.pone.0130992.t002

by Andre Chevrier, Douglas Cheyne, Simon J. Graham, and Russell J. Schachar
3.1 Abstract

Often we must balance being prepared to act quickly with being prepared to suddenly stop. The stop signal task (SST) is widely used to study inhibitory control, and provides a measure of the speed of the stop process that is robust to changes in subjects’ response strategy. Previous studies have shown that preparation affects inhibition. We used fMRI to separate activity that occurs after a brief (500 ms) warning stimulus (warning-phase) from activity that occurs during responses that follow (response-phase). Both of these phases could contribute to the preparedness to stop because they both precede stop signals. Warning stimuli activated posterior networks that signal the need for top-down control, whereas response phases engaged prefrontal and subcortical networks that implement top-down control. Regression analyses revealed that both of these phases affect inhibitory control in different ways. Warning-phase activity in the cerebellum and posterior cingulate predicted stop latency and accuracy, respectively. By contrast, response-phase activity in fronto-temporal areas and left striatum predicted go speed and stop accuracy, in pre-supplementary motor area affected stop accuracy, and in right striatum predicted stop latency and accuracy. The ability to separate hidden contributions to inhibitory control during warning-phases from those during response-phases can aid in the study of models of preparation and inhibitory control, and of disorders marked by poor top-down control.
3.2 Introduction

Preparation is important to the accurate and rapid execution of any wilful behaviour, including the ability to stop ongoing behaviour, referred to as inhibitory control. Inhibitory control involves stopping actions that are inappropriate or are no longer required, and is crucial to goal-directed behaviour in changing conditions (Jennings & van der Molen, 2005). Inhibitory control has been studied extensively using the stop signal task (SST) (Logan, 1994). SST trials begin with a brief warning stimulus signalling that a trial is about to begin, followed by a choice-response stimulus. Occasionally (e.g. 33%) and unpredictably, a stop signal follows the go signal, indicating that the subject should not make that response. The latency of the stop process (stop signal reaction time (SSRT)) is estimated by subtracting the mean stop delay from the mean go reaction time. Although models of inhibitory control rely heavily on the independence of go- and stop-processes (Logan, 1994), it has been shown that SSRT is affected by the level of preparation (Li, Krystal, & Mathalon, 2005), and by preparatory brain activities (Chikazoe et al., 2009). This apparent contradiction has arisen partly because the independence of going and stopping in models of inhibitory control refers only to the relative finishing times of go- and stop processes, which are modelled as being initiated by go- and stop-signals, respectively. The independence of the relative finishing times of go- and stop-processes has been interpreted to mean that these processes do not begin until after the presentation of go- and stop signals, when the race model is in fact silent about the onset of these processes. The assumption that go- and stop-processes do not begin until the presentation of go- and stop-stimuli appears to be true in the modality-specific networks that ultimately carry out or inhibit responses.
However, from the moment of the earliest warning stimuli that precede stop signals, central structures must engage and maintain top-down control over modality-specific networks until a decision is made to go or to stop (Jaffard et al., 2008). Here we attempt to gain a better view of the hidden factors that affect the decision to go or stop while top-down control is being engaged, before any stop signal appears.

The warning stimuli used in the SST are brief (500 ms) compared to the temporal resolution of fMRI (e.g. 2s). As a consequence, previous studies have used indirect methods to infer preparation-related activity in stop tasks. For example, Chikazoe et al. (Chikazoe et al., 2009) included trials in which it was known that no stop signal would appear. By subtracting activity on these trials from go-trials that might include stop signals, they found activity in inhibition-related structures, suggesting that subjects were pre-activating these regions in anticipation of potential stop signals. Therefore there is strong evidence that inhibition-related activity comes into play before stop signals appear. Here we attempt to capture these preparatory activities directly and use regression analyses to assess their effects on go- and stop-processes.

Other approaches have contrasted compound trials in various ways and have inferred preparatory activities that affect inhibition, particularly in fronto-parietal networks and presupplementary motor area (Chikazoe et al., 2009; Connolly et al., 2002; Hu & Li, 2012; Jaffard et al., 2008). For example, some studies have separated preparation from inhibition by using long fore-period delays (Postle et al., 2000), and contrasting successful with failed stop trials across variable fore-period delays (Hu & Li, 2012;
Jaffard et al., 2008). However, using long fore-period delays can be problematic because this kind of preparation cannot be maintained for more than a few seconds before resetting working memory (Jennings & van der Molen, 2005), which would invoke preparation-phase activities that are not related to preparation. Further, contrasting failed and successful stop trials confounds differences in monitoring and feedback-reinforcement processes on these trials with differences in preparation (Rubia et al., 2003). Hu and Li (Hu & Li, 2012) contrasted failed and successful stop trials using two separate deconvolution analyses; one that was time-locked to warning stimuli and one that was time locked to response stimuli. However, warning- and response phase activities are not statistically independent and so should be separated simultaneously in a single deconvolution analysis. In this study, we accomplish such a simultaneous deconvolution by capturing response-phase activities using separate regressors that were time-locked to left- and right-hand responses instead of a single regressor time-locked to the presentation of response-stimuli.

Another approach has been to present response stimuli after stop signals in order to isolate the circuits involved in preparing to stop a response that the subject already knows will have to be stopped (Connolly et al., 2002). However, this cannot reflect the kind of preparedness to stop that occurs before stop signals appear. We present a method to generate a whole-brain view of two phases of activity that affect our preparedness to stop when it remains uncertain whether the response on that trial will have to be stopped or not. Imaging preparation to stop in the context of uncertainty is important because we often do not have advance notice of the need to stop in real-life situations. The central decision to stop or to go is an important part of overall inhibitory control and in certain
circumstances, or in certain disorders, may be more important than the modality specific skeleto-motor mechanisms that ultimately carry out or inhibit the overt response.

The value of directly separating preparation from other within-trial processes, instead of contrasting compound trials, was demonstrated in a recent study of attention deficit hyperactivity disorder (ADHD). ADHD is associated with inhibitory control deficits, and ADHD subjects have been shown to under-activate the inferior frontal gyrus (IFG) on stop trials (Dickstein et al., 2006; Morein-Zamir et al., 2014; Mulligan et al., 2011).

The natural interpretation of under-activation of the IFG on stop trials has been that it reflects the under-performance of inhibitory control in ADHD, and that it manifests during the act of stopping a response. However, by separating activities before and after stop signals that had been combined in previous studies, it was demonstrated that this apparent under-activation is in fact the result of ADHD subjects deactivating the IFG before stop signals appeared, when healthy control subjects were pre-activating the IFG (Bhaijiwala et al., 2014). Contrary to previous assertions, ADHD subjects were not under-activating the IFG when stopping their responses; IFG differences in ADHD instead reflect a dysfunctional attempt to pre-activate the IFG during preparation.

Therefore, rather than simply reflecting the narrative of poor inhibitory control in ADHD, the direct separation of within-trial activities suggests a dysfunctional reinforcement mechanism that tunes appropriate networks in an inappropriate way. This vital kind of distinction can contribute to the understanding and treatment of cognitive dysfunction, and could not have been found by the conventional approach of contrasting compound trials based on cognitive models of inhibitory control. Therefore, directly separating preparation from other within-trial activities can provide insights into the
neural mechanisms of behavioral deficits that might be beyond the predictions of cognitive models.

Here we present an approach to separate activity that occurs after warning stimuli from activity that occurs during response phases that immediately follow. Both of these phases precede stop signals, and both of these phases should affect inhibition (Chikazoe et al., 2009; Connolly et al., 2002; Hu & Li, 2012; Jaffard et al., 2008; Li et al., 2005; Smittenaar, Guitart-Masip, Lutti, & Dolan, 2013). We used separate regressors for warning stimuli and for left- and right-hand responses, and employed a temporal jittering strategy that optimized the separability of event-types in the deconvolution analysis (as in (Chevrier et al., 2007; Chevrier & Schachar, 2010)). Despite the brief nature of preparation in the SST, it is possible to separate rapid sequences of activity as long as they do not always co-occur, and as long as the component processes these activities reflect unfold in a known temporal order (Ollinger et al., 2001). The regressors used here, reflecting warning- and response-phases in the SST, undoubtedly unfold in a known sequence in time, and do not always co-occur because one sixth of trials were successful stop trials, which contained warning stimuli but no overt response. However, although stop trials contain no response, they should contain response-phase preparatory activity even if no response regressor is used for these trials in the deconvolution analysis. To this end, we showed in a previous study that in the proposed approach, response-phase activity on successful stop trials is captured by the stop stimulus, along with activities reflecting response cancellation (Chevrier et al., 2007). Further, despite the brief temporal separation between warning stimuli and responses (~ 1s), a similar approach in a previous study isolated prediction error deactivations in an ascending
dopamine pathway (Chevrier & Schachar, 2010) which only last 100 ms, an order of magnitude shorter than the preparatory periods being investigated here. Therefore the current approach should be capable of separating activity during warning-phases from those during response-phases that follow.

In this study, we will refer to activities related to warning stimuli as “warning-phase” activities, and activities related to responses as “response-phase” activities. Significant warning- and response-phase activities were correlated with individual differences in the speed and accuracy of going and of stopping to identify their distinct influences on inhibitory control. Despite the overall independence of the finishing times of go- and stop-processes in the SST, we expect that distinct preparatory activities during warning- and response-phases will have distinct influences on the speed and accuracy of go- and stop-processes.

Given the close temporal proximity of warning- and response-phase activity, and the fact that response-phase activities should also be present on stop trials that contain no response regressor, it is important to first determine whether the patterns of activation found here are consistent with what would be expected based on previous research using event-related potentials (ERP). Several differences between warning- and response-phases that have consistently been observed using ERP should also be present here if the current approach were successful. Firstly, warning-phases tend to activate left parietal regions, whereas response phases activate right parietal regions (Khonsari et al., 2007; van Boxtel & Brunia, 1994). Secondly, primary motor cortex deactivates during warning-phases and activates during response phases (Brunia & Damen, 1988). Thirdly,
warning-phases primarily activate posterior regions, followed by fronto-posterior activities during response-phases (Mnatsakanian & Tarkka, 2002). Fourthly, preparation requires contributions from subcortical structures, but begins cortically (Gemba & Sasaki, 1984). Therefore, if the current approach successfully separated these phases of activity, then the results should involve: left followed by right parietal activity, deactivation followed by activation in primary motor cortex, posterior activity followed by fronto-posterior activity, and cortical activity followed by cortical-subcortical activity. After whole-brain correction, the current results contained all of these expected patterns of activity, providing a high degree of confidence that the approach can indeed separate warning- from response-phase activity. Regression analyses revealed distinct warning- and response-phase activities that had distinct influences on inhibitory control.

3.3 Materials and methods

3.3.1 Subjects and task design

Fourteen healthy, right-handed subjects (mean age 29.4 years) participated in the study. Subjects had normal vision and reported no medication use, medical illness or psychological problems. Subjects gave informed written consent to participate in the study, which was approved by our institutional research ethics board. Stop task trials began with a 500 ms fixate (warning) stimulus followed by a go (response) stimulus that remained on the screen for 1000 ms (Figure 3.1). The go (response) stimulus consisted of either the letter “X,” which indicated that the subject should make a button press with their left thumb, or the letter “O,” which indicated that the subject should make a button
press with their right thumb. On one third of trials, the response-stimulus was followed by a stop signal (change in background colour from black to red), indicating that the subject should try to stop their response on that particular trial. Stop signals followed response-stimuli after an adaptive delay that ensured subjects could only successfully stop on approximately half of the stop trials presented (Logan, 1994). The initial stop delay was 250 ms, and increased or decreased by 50 ms when subjects succeeded or failed to stop, respectively.

**Figure 3-1 Trial types and regressors for isolating preparatory activity in the SST**

Trials began with a fixation dot (warning stimulus) in the middle of the screen for 500 ms followed by a response stimulus for 1 s. The response stimulus was either an “X”, indicating that the subject should press a button with their left thumb (a), or an “O”, indicating that the subject should press a button with their right thumb (b). On one third of trials (c), the response stimulus was followed by a stop signal (screen colour change from black to red), indicating that the subject should not respond on that trial. The box at the top right of the figure describes the stimuli that are used to estimate warning- and response-phase activities.
A blank screen appeared during the inter-trial interval (ITI), which was either 1 or 2 seconds, resulting in trial lengths of 2.5 or 3.5 s. The ITI’s were jittered with random combinations of spread-spectrum binary coding sequences. These binary sequences are mutually orthogonal, which maximises the diversity of temporal offsets from one trial to the next. Maximizing the diversity of temporal offsets of subsequent trials serves to maximize the number of independent equations in the deconvolution analysis, which enhances the separation of event types in the model. This timing also ensured that warning-stimuli and responses to go-stimuli only occurred in the same fMRI time-point on about half of trials. Further, although every trial contains a warning-stimulus followed by an imperative response-stimulus, not every trial contains a response (i.e. one out of six trials were successful stop trials), resulting in sufficient contrast over the course of the experiment to separate activity related to warning-stimuli from activity related to responses to go-stimuli. A 17.5 second rest followed every fourteenth trial in order to generate a reliable estimate of baseline activity. The task was performed in the MRI environment for a total of 21 minutes and 40 seconds, yielding a total of 322 trials.

3.3.2 Behavioral measures

Mean reaction times on successful go trials were calculated for all subjects, as were percent accuracy for both go trials and stop trials. The latency of the stop process is not directly observable in the SST because there are no responses on successful stop trials. However, the stop signal reaction time (SSRT) can be estimated by subtracting the mean stop delay from the mean go reaction time (Logan, 1994). The SSRT is a reliable
measure of inhibitory control, with shorter SSRT’s indicating faster inhibition than longer SSRT’s.

3.3.3 fMRI acquisition and subject-level analysis

Imaging was performed with a GE LX 1.5T MRI scanner (General Electric, Milwaukee, LX hardware and software platform). Anatomical data were acquired with a standard high-quality SPGR sequence (96 slices, 1.5 mm thick, FOV = 20 cm, 256x256 matrix). Functional data were collected using a spiral gradient echo pulse sequence (TE/TR/θ = 40/2000/90). Functional acquisitions were 24 slices, 6 mm thick, FOV = 20cm, reconstructed at 64x64 pixel resolution resulting in a final voxel size of 3.125 x 3.125 x 6 mm³. Subjects’ responses were collected using a LumiTouch fibre-optic button box (Lightwave Medical, Burnaby, British Columbia, Canada) interfaced to a laptop running the stop task paradigm at the MRI system console.

Functional data were analyzed using AFNI (Cox, 1996). Images were motion corrected using a standard coregistration algorithm (Woods, Grafton, Watson, Sicotte, & Mazziotta, 1998). None of our subjects moved more than 2mm over the course of the scan, and could therefore all be included in subsequent analyses. A general linear model of stimulus vectors convolved with the hemodynamic response function (HRF) was used in the deconvolution analysis. Estimates of baseline and linear drift were generated simultaneously with 6-point HRF’s (12 second duration) for each event type outlined below (HRF delay = 2TR = 4 seconds).

The following event types were included in the deconvolution analysis: Fixate = warning-phase; X = left hand response; O = right hand response; Stop = successful
inhibition; Fail = error processing on failed inhibition trials. Go-trials were modelled using Fixate and response (X or O) events. Stop trials were modelled using Fixate and Stop events. Failed stop trials were modelled using warning (Fixate), response (X or O) and error processing (Fail) events. Figure 3.1 portrays the relevant trial types and regressors used to separate warning- from response-phase activity. Warning-phase activity was isolated with the Fixate map. Response-phase activity was isolated using the contrast \( \frac{1}{2}(X+O) \), which enhanced activities that would be common to both responses, while suppressing lateralized activity related to the handedness of the response (as in (Chevrier et al., 2007)).

Multicollinearity of a deconvolution analysis can be optimized by minimizing the inverse average error of the model matrix, which in the current experimental design was below \( 10^{-15} \). Further, the inherent correlation between warning- and response-phase regressors can be computed prior to data collection by determining the angle between them using the formula for the dot product (i.e. \( \theta = \cos^{-1} \left( \frac{\Sigma \text{warning}_i \times \text{response}_i}{|\text{warning}| \times |\text{response}|} \right) \)), where \( i \) ranges over all time fMRI time points, warning, and response; equal 1 when the event occurs during time point \( i \) and 0 when it does not occur, and \( |x| \) denotes the magnitude of vector \( x \). A grazing angle (near 0) signifies high collinearity whereas angles approaching orthogonality (i.e. 90°) signify low collinearity and a basis for meaningful contrast. Using this method we found that the angle between warning- and response-phase regressors was 62.8°, corresponding to a correlation of around 0.5. Using this value, we were able to quantifiably estimate whether warning- and response-phase activities can be separated. We found the detection-tolerance (calculated as \( 1-r^2 \)) to be 0.75. Tolerance values less than 0.1 or 0.2 would indicate that
Multicollinearity is a problem (O’Brien, 2007). It appears the current approach should be capable of separating warning- from response-phase activity as the tolerance is nearly a factor of 4-8 times better than required. From a qualitative perspective, when regressors are multicollinear, the respective maps generally show activation and deactivation in the same anatomical locations. If the approach was successful, then regions that are known to activate during both of these phases (e.g. primary visual areas) should exhibit positive activity in both of these maps, whereas regions that are known to deactivate on the presentation of warning-stimuli and then activate on the presentation of response-stimuli (e.g. primary motor cortex) should show deactivation followed by activation. Therefore as a final test of the multicollinearity between warning- and response-regressors, their unthresholded group activation maps were visually inspected to determine if this was the case. The low multicollinearity between regressors despite their close temporal proximity was due to the experimental design: firstly, temporal jittering using spread spectrum binary coding sequences maximized the number of independent equations in the deconvolution; secondly, one out of six trials (successful stops) contained fixate cues, but no overt response (X or O); thirdly, the interplay of 2.5 and 3.5 s trial lengths with a TR of 2 s ensured that warning- and response-regressors did not always occur in the same fMRI time-point even if they both occurred in the same trial.

Intensity maps for all event types, and the \( \frac{1}{2}(X+O) \) contrast were generated for each subject by taking the area under the HRF, estimated by the sum of HRF coefficients generated by AFNI’s 3dDeconvolve program. These maps were then warped into Talairach space, and blurred (6mm FWHM).
3.3.4 Group-level analysis

Activity maps reflecting warning- and response-phases (Fixate and ½(X+O) maps) for all subjects were entered into a group ANOVA to determine activities that were common across the group. This ANOVA output (distributed as a t* statistic with 65 degrees of freedom due to the number of subjects in the study, and the number of event-types in the model) was then cluster/thresholded according to Gaussian field theory using AFNI’s AlphaSim program in order to correct for multiple statistical comparisons as in (Chevrier et al., 2007). This analysis required that significant voxels be part of a larger cluster of at least four original voxels (234.4 mm$^3$) with a minimum z-score of 3.3 for an overall $\alpha < 0.05$. Therefore every region that significantly activated has an effect size greater than 3.3. These regions represent statistically significant (whole-brain corrected) activations for the group. In order to test whether activities in these regions might affect the balance between going and stopping, we sorted the group of fourteen subjects according to their go reaction time, go accuracy, stop signal reaction time, and percent stop accuracy. We then performed regression analyses to identify warning- and response-phase activities that predicted individual differences in the speed or accuracy of going or of stopping.

3.4 Results

3.4.1 Subject performance

Behavioral data were consistent with normal adults in non-MRI environments. Go responses were fast (597.7 ± 53.7 ms) and accurate (99.6 ± 0.98 %), and the mean stop
signal reaction time (SSRT) was within normal range (210.3 ± 48.0 ms) for adults. Approximately half of the stop trials presented contained erroneous responses (48.8 ± 2.4 %), indicating that the tracking algorithm was functioning as expected. The speed of the go-process, measured by mean go reaction time, was not significantly correlated with the speed of the stop-process, as measured by the SSRT ($r^2 = 0.053; p = 0.57$) in our subjects. Faster inhibitory control was therefore not simply a result of slower responses, which is consistent with the well-established independence of the relative finishing times of go- and stop-processes in the SST.

### 3.4.2 Warning- and response-phase activities:

Single subject deconvolution models had a low degree of multicollinearity, as indicated by inverse average errors on the order of $10^{-15}$. fMRI results correspondingly showed a clear differentiation between activities during warning- compared to response-phases of the task. Visual inspection of warning- and response-phase maps revealed that regions that are known to activate during both of these phases (e.g. primary visual areas) showed positive activity in both of these maps, whereas regions that are known to deactivate and then activate (e.g. primary motor) showed negative activity in the warning-phase and increased activity during the response-phase. Whole-brain correction revealed warning-related activity changes in posterior networks that have previously been implicated in motor preparation (Table 3.1). Decreases in BOLD activity were seen in right primary sensory and motor regions (Figure 3.2), posterior cingulate and lingual gyri. Increases in BOLD activity were seen in right superior parietal cortex and cerebellum. No significant activity changes were present in prefrontal cortex, thalamus or basal ganglia. By contrast, response-phases contained significant BOLD activities in prefrontal cortex,
thalamus and basal ganglia as well as in posterior cortical networks and lateral
cerebellum (Table 3.2).

<table>
<thead>
<tr>
<th>Structure</th>
<th>BA</th>
<th>Location</th>
<th>Estimate</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>L parietal</td>
<td>7</td>
<td>-29, -55, 49</td>
<td>3.1</td>
<td>-</td>
</tr>
<tr>
<td>R occipital</td>
<td>18</td>
<td>26, -86, 1</td>
<td>9.0</td>
<td>-</td>
</tr>
<tr>
<td>L fusiform</td>
<td>37</td>
<td>-37, -61, -10</td>
<td>9.6</td>
<td>-</td>
</tr>
<tr>
<td>L cerebellum (vermis)</td>
<td>-</td>
<td>-7, -73, 18</td>
<td>5.3</td>
<td>SSRT</td>
</tr>
<tr>
<td>R primary motor</td>
<td>4</td>
<td>19, -26, 58</td>
<td>-3.0</td>
<td>-</td>
</tr>
<tr>
<td>R primary sensory</td>
<td>3</td>
<td>16, -42, 63</td>
<td>-3.4</td>
<td>-</td>
</tr>
<tr>
<td>R inferior parietal</td>
<td>40</td>
<td>47, -27, 22</td>
<td>-3.7</td>
<td>-</td>
</tr>
<tr>
<td>R insula</td>
<td>13</td>
<td>49, -18, 15</td>
<td>-2.8</td>
<td>-</td>
</tr>
<tr>
<td>L posterior cingulate</td>
<td>31</td>
<td>-6, -57, 21</td>
<td>-4.5</td>
<td>-</td>
</tr>
<tr>
<td>L posterior cingulate</td>
<td>31</td>
<td>-7, -53, 31</td>
<td>-3.7</td>
<td>-</td>
</tr>
<tr>
<td>L posterior cingulate</td>
<td>23/31</td>
<td>-10, -45, 24</td>
<td>-2.8</td>
<td>PC</td>
</tr>
<tr>
<td>R posterior cingulate</td>
<td>23</td>
<td>9, -57, 15</td>
<td>-3.0</td>
<td>-</td>
</tr>
</tbody>
</table>

**Table 3-1 BOLD activities during warning-phases**

BA–Brodmann area; SSRT–regions that significantly correlated with stop signal reaction time; PC–regions that significantly correlated with stop accuracy (i.e. percent correct stop trials). Locations indicate maximum activity coordinates for all significant clusters, and are given in Talairach coordinates. Activity estimates in arbitrary units.

### 3.4.3 Regression analyses

Whole-brain corrected warning-phase activities were regressed with subjects’ mean go reaction time, go task accuracy, stop accuracy and SSRT in order to determine whether activity in these regions predicted individual differences in behavior. No warning-phase activations showed a significant correlation with mean go reaction time or go task accuracy. By contrast, warning-related activity in the vermis of the cerebellum was significantly correlated with SSRT ($r^2 = 0.38; p = 0.018$), indicating that greater activity in this part of the cerebellum during warning-phases predicted a longer latency of inhibition, but this region was not significantly correlated with stop accuracy. However,
warning-phase activity in one part of the posterior cingulate cortex (PCC) (BA 31) was significantly correlated with stop accuracy ($r^2 = 0.27; p = 0.031$), indicating that greater activation, or, more properly, less deactivation, predicted a higher likelihood of successfully stopping on stop trials. Warning-phase activities that predicted individual differences in behavior are portrayed in Figure 3.2, and indicated in Table 3.1 a).

### Table 3-2 BOLD activities during response-phases

<table>
<thead>
<tr>
<th>Structure</th>
<th>BA</th>
<th>Location</th>
<th>Estimate</th>
<th>Correlations</th>
</tr>
</thead>
<tbody>
<tr>
<td>R inferior parietal</td>
<td>40</td>
<td>42, -45, 50</td>
<td>6.3</td>
<td>-</td>
</tr>
<tr>
<td>L SMA</td>
<td>6</td>
<td>-7, 8, 60</td>
<td>7.1</td>
<td>-</td>
</tr>
<tr>
<td>L pre-SMA</td>
<td>6/24</td>
<td>-3, 1, 47</td>
<td>5.2</td>
<td>PC</td>
</tr>
<tr>
<td>L ACC</td>
<td>24</td>
<td>-10, -15, 34</td>
<td>4.7</td>
<td>-</td>
</tr>
<tr>
<td>R ACC</td>
<td>32</td>
<td>7, 32, 24</td>
<td>5.5</td>
<td>-</td>
</tr>
<tr>
<td>R superior frontal</td>
<td>10</td>
<td>20, 50, 24</td>
<td>6.9</td>
<td>PC</td>
</tr>
<tr>
<td>R superior frontal</td>
<td>10</td>
<td>30, 47, 0</td>
<td>13.3</td>
<td>RT, PC</td>
</tr>
<tr>
<td>L occipital</td>
<td>18</td>
<td>-25, -86, 3</td>
<td>9.4</td>
<td>-</td>
</tr>
<tr>
<td>R inferior temporal</td>
<td>20</td>
<td>43, -13, -28</td>
<td>5.6</td>
<td>-</td>
</tr>
<tr>
<td>R lingual</td>
<td>17</td>
<td>15, -89, -3</td>
<td>11.2</td>
<td>-</td>
</tr>
<tr>
<td>L putamen</td>
<td>-</td>
<td>-22, 10, 11</td>
<td>5.6</td>
<td>RT, PC</td>
</tr>
<tr>
<td>L globus pallidus (internal)</td>
<td>-</td>
<td>-15, -8, -4</td>
<td>9.1</td>
<td>RT, PC</td>
</tr>
<tr>
<td>R putamen</td>
<td>-</td>
<td>20, 5, 0</td>
<td>9.1</td>
<td>SSRT, PC</td>
</tr>
<tr>
<td>R thalamus (ventral lateral)</td>
<td>-</td>
<td>15, -10, 16</td>
<td>5.6</td>
<td>SSRT</td>
</tr>
<tr>
<td>L cerebellum</td>
<td>-</td>
<td>-30, -67, -28</td>
<td>7.5</td>
<td>-</td>
</tr>
<tr>
<td>R precuneus</td>
<td>7</td>
<td>-7, -65, 35</td>
<td>-8.2</td>
<td>-</td>
</tr>
<tr>
<td>L cuneus</td>
<td>18</td>
<td>0, -78, 25</td>
<td>-7.0</td>
<td>-</td>
</tr>
<tr>
<td>L superior temporal</td>
<td>22</td>
<td>-46, -21, -8</td>
<td>-6.0</td>
<td>RT, PC</td>
</tr>
<tr>
<td>R insula</td>
<td>21</td>
<td>44, -9, -7</td>
<td>-9.4</td>
<td>-</td>
</tr>
<tr>
<td>R orbitofrontal</td>
<td>11</td>
<td>2, -12, -11</td>
<td>-9.1</td>
<td>-</td>
</tr>
<tr>
<td>L orbitofrontal</td>
<td>11</td>
<td>-25, 21, -12</td>
<td>-12.2</td>
<td>-</td>
</tr>
<tr>
<td>L medial frontal</td>
<td>10</td>
<td>-8, 59, 2</td>
<td>-9.8</td>
<td>-</td>
</tr>
<tr>
<td>R ACC</td>
<td>25</td>
<td>-1, 4, -5</td>
<td>-10.5</td>
<td>-</td>
</tr>
</tbody>
</table>
Figure 3-2 Warning-phase activities

Arrows indicate regions where the level of warning-phase activity significantly predicted individual differences in SSRT (cerebellum), and PC (posterior cingulate BA 23/31). Abbreviations: SSRT—stop signal reaction time; PC—percent correct inhibition. Numbers in bottom right corners refer to the z coordinate in Talairach space. Increased activity is coloured red-yellow and decreased activity blue. Slices are portrayed in radiological space (right = left).

Whole-brain corrected response-phase activities were regressed with mean go reaction time, go task accuracy, SSRT and stop accuracy to determine whether the speed or accuracy of behaviour could be predicted by individual differences in response-phase activity. Response-phase activities that predicted individual differences in behaviour are portrayed in Figure 3.3 and indicated in Table 3.1 b). No response-phase activities were significantly correlated with individual differences in go task accuracy. However, left putamen ($r^2 = 0.36; p = 0.013$), left globus pallidus ($r^2 = 0.40; p = 0.009$), left superior temporal gyrus BA 22 ($r^2 = 0.35, p = 0.015$) and right superior frontal gyrus BA 10 ($r^2 = 0.36; p = 0.013$)
0.24; p = 0.043) activities were significantly correlated with mean go reaction time, indicating that greater response-phase activity in these regions predicted longer response times. In addition, every region that significantly predicted go reaction time, also predicted stop accuracy (left putamen: $r^2 = 0.41$, p = 0.0085, left globus pallidus: $r^2 = 0.65$, p = 0.00032, left superior temporal gyrus BA 22: $r^2 = 0.61$, p = 0.00059, and right superior frontal gyrus BA 10: $r^2 = 0.29$, p = 0.026). Although go reaction time was strongly correlated with stop accuracy ($r^2 = 0.66$, p = 0.00043), there were two regions of response-phase activity that predicted greater stop accuracy but not longer reaction times: left pre-supplementary motor area (pre-SMA) BA 6/24 ($r^2 = 0.23$, p = 0.049) and right superior frontal gyrus BA 10 ($r^2 = 0.3$, p = 0.026), which means that more response-phase activity in these areas predicted a higher probability of stopping on stop trials without affecting the speed of the go-process. The only response-phase activities that predicted individual differences in stop signal reaction time were in right putamen ($r^2 = 0.33$, p = 0.018) and ventral lateral thalamus ($r^2 = 0.16$, p = 0.019), indicating that subjects who exhibited greater response-phase activity in these areas had a longer latency of inhibition. Finally, response-phase activity in right putamen, which predicted longer stop signal reaction time, also predicted higher stop accuracy ($r^2 = 0.27$, p = 0.024), whereas activity in ventral lateral thalamus did not ($r^2 = 0.2$, p = 0.062).
Figure 3-3  Response-phase activities
Arrows indicate regions that significantly predicted individual differences in behaviour. a) Right superior frontal BA 10 predicted RT and PC; right putamen predicted SSRT and PC; right ventral lateral thalamus predicted SSRT; left putamen, internal globus pallidus and left superior temporal BA 22 predicted RT and PC. b) Left superior frontal BA 10 predicted PC. c) Left pre-SMA predicted PC. Abbreviations: RT—go reaction time; PC—percent correct inhibition; SSRT—stop signal reaction time. Reprinted from

3.5  Discussion
The current approach revealed warning- and response-phase activities in central structures that are involved in preparing to decide whether to stop or to go, which can bring relevant modality-specific inhibition networks into play. We used a stop task that involved the presentation of warning stimuli before the presentation of imperative response stimuli, which, in turn, were occasionally followed by a stop signal. We developed an imaging approach capable of separating this sequence of activities, and applied regression analyses to identify preparatory activities that influenced go- and stop-processes before stop signals appeared. The results add detail to previous studies.
that have demonstrated the influences of preparation on inhibitory control in health and disease (Chikazoe et al., 2009; Connolly et al., 2002; Schwerdtfeger et al., 2013; Jaffard et al., 2008; van Veen, Krug, & Carter, 2008).

Separating these phases of activity appears to have distinguished two important stages of engaging top-down control: first signalling the need for top-down control during warning-phases, followed by the implementation of that top-down control during response-phases. The presence of activity in posterior cortical regions and cerebellum during warning-phases reflects the bottom-up call for the upcoming need to exert top-down control at the beginning of a trial. Activities in fronto-posterior and cortical-subcortical networks during response-phases that follow reflect the implementation of that top-down control, which must balance the need to respond quickly with the occasional need to stop that speeded response. The current approach identified preparatory activities that influenced inhibitory control in the same regions that have been found in previous approaches and in several regions not found in previous approaches that have not separated warning- and response-phase activities.

3.5.1 Warning- and response-phases can be separated with fMRI

One indication of the success of the current approach is that warning- and response-phase maps contained entirely unique distributions of activity. Previous approaches that have employed separate deconvolution analyses have generated warning- and response-phase maps with greatly overlapping distributions of activity. A further indication of the success of the current approach is that the patterns of warning- and response-phase
activities we found are highly consistent with observations from previous ERP studies that have not previously been distinguished using fMRI.

Several patterns of activity identified here are highly consistent with previous observations from cued-response tasks using ERP, providing a high degree of confidence that the method succeeded in separating warning- from response-phase activities. Firstly, the observation of left followed by right parietal activity has been demonstrated directly with ERP (Khonsari et al., 2007; van Boxtel & Brunia, 1994) but only indirectly with fMRI (Brass & von Cramon, 2004). Secondly, the current finding of deactivation in primary motor cortex after warning stimuli followed by positive activity after imperative stimuli has previously been observed using ERP in cued-response tasks (Brunia & Damen, 1988; Gemba & Sasaki, 1984) but has not been distinguished with a whole-brain fMRI design. Deactivation of primary motor cortex during warning-phases that precede response-stimuli has been interpreted as inhibition of the afferent aspect of reflexive behaviour in order to ensure a more central regulation of motor output (Brunia, 1993). Thirdly, warning stimuli activated only neocortical regions (and cerebellum), which is consistent with previous assertions from scalp recordings in monkeys that preparation requires contributions from subcortical structures, but begins cortically (Gemba & Sasaki, 1984). Fourthly, the finding of activity in exclusively posterior networks after warning stimuli followed by fronto-posterior activities after imperative stimuli has been suggested (Mnatsakanian & Tarkka, 2002) but has not previously been shown in a whole-brain context.
3.5.2 Regions that affect the balance between going and stopping - Regression analyses

Whole-brain corrected preparatory activities were correlated with individual differences in the speed and accuracy of the go- and stop-processes. Despite the independence of the finishing times of go- and stop-processes, regression analyses identified distinct warning- and response-phase activities that covertly affect distinct components of inhibitory control before stop signals appear.

A previous study by Chikazoe et al. found that preparatory activities in frontoparietal regions was proportional to the preparation cost associated with longer warning-phases (Chikazoe et al., 2009). However, their approach combined frontal, parietal and other areas as a single region of interest, so it is difficult to determine which brain regions drove this correlation. The only other previous study that has found preparatory activities that were correlated with behavioural measures of inhibitory control stated that subjects with shorter SSRT, as determined by a median split, exhibited greater activity in orbitofrontal cortex (Hu & Li, 2012). However, this result was obtained by contrasting the difference between failed and successful stop trials from two separate deconvolution analyses; one time-locked to the warning stimulus and one time-locked to the go-signal. This result was likely driven by the fact that this part of orbitofrontal cortex strongly deactivates on failed stop trials (A. Chevrier & Schachar, 2010), which would lead to apparently strong activation using the contrast (successful stop)-(failed stop). Although these authors did attempt to rule out error-related processing by excluding from analysis any areas that exhibited greater activity during failed than successful stop trials, this would not be the case for the orbitofrontal cortex. We show
that warning- and response-phases are better separated using a single deconvolution analysis because these phases of activity are not statistically independent and should therefore be solved for simultaneously.

We found that initial warning stimuli deactivated the posterior cingulate cortex, with less deactivation predicting a less accurate stop process (i.e. lower percent successful inhibition). A previous study by Jaffard et al. (Jaffard et al., 2008) that investigated activity associated with warning-phases also found deactivation in this part of the posterior cingulate cortex, and ascribed this activity to general alerting processes. Regardless of the general role of this structure, we show that activity here affects the accuracy of inhibitory control. This area of the posterior cingulate is part of the ‘task-negative network’. This network is known to be less anti-correlated with task-specific activity as working memory load increases (Leech, Kamourieh, Beckmann, & Sharp, 2011). Therefore subjects who deactivated this area to a lesser extent during warning-phases may have used a strategy with greater working memory load, causing a slower stop process. Warning stimuli also invoked activity in vermis of the cerebellum, where greater activity predicted a slower stop process. Given that this part of the cerebellum is involved in response readiness (Yakusheva et al., 2007), it appears that subjects who were more oriented towards the go-component of the task required more notice in order to successfully stop. This is consistent with previous findings that varying fore-period (i.e. warning) delays affects inhibitory control by affecting response readiness (Jaffard et al., 2008; C. S R Li et al., 2005).
Response-phase activity in left superior temporal gyrus BA 22 and right superior frontal gyrus BA 10 predicted longer go reaction time and higher stop-accuracy. Superior frontal gyrus serves as an interface between high level representations of externally and internally generated states, and is crucial to the ongoing evaluation of the consequences of decisions (Goldman-Rakic, 1987). Left superior temporal gyrus projects to superior frontal BA 10 (Burman, Reser, Yu, & Rosa, 2011), and is involved in the understanding and generalizations of words, and therefore likely served to invoke a representation of the stop-rule in working memory before stop signals appeared. By contrast, pre-SMA BA 6/24, which has been associated with preparatory inhibition activity in previous studies (Chikazoe et al., 2009; Connolly et al., 2002), predicted stop accuracy, but not go-reaction time. Pre-SMA is involved in procedural learning and the development of skills requiring elaboration of motor behaviour (Ackermann, Daum, Schugens, & Grodd, 1996). Therefore, neocortical response-phase activities that predicted variability in stop task performance are consistent with a network that generates internal representations of task goals (superior temporal cortex), which engage top-down control (superior frontal cortex) over pre-SMA and motor circuits that in turn project to subcortical structures, where control of primary motor output can be achieved (Jaffard et al., 2008).

Subcortical response-phase activities in the left putamen and the internal segment of globus pallidus predicted longer go reaction time and higher stop accuracy. Activity in left putamen is associated with the retention and recall of visuomotor response mappings (Buch, Brasted, & Wise, 2006; Grol, de Lange, Verstraten, Passingham, & Toni, 2006). The significant correlation of activity in left putamen with go reaction time is consistent with previous observations that activity here is related to the degree of response
preparation (Mars, Coles, Hulstijn, & Toni, 2008). Response-phase activity in the internal segment of the left globus pallidus also predicted longer go reaction time and greater stop accuracy. The initiation of skeletomotor movements is under the inhibitory influence of the internal segment of globus pallidus (Alexander, DeLong, & Strick, 1986), which might be the skeletomotor correlate of fixation neurons that have been indentified in oculomotor versions of the SST (Boucher et al., 2007). Taken together, these correlations indicate that preparatory activity in left putamen and globus pallidus serve to regulate the speed of the go-process by both preparing the go-response (left putamen) and delaying the onset of that response (globus pallidus). These results indicate that subjects who activated left striatum more strongly after imperative response-stimuli used a strategy that involved longer delays before the initiation of a response, which explains the correlation with longer response times in these subjects.

In contrast with the left striatum, in which response-phase activity predicted longer go reaction times, response-phase activities in similar structures in the opposite hemisphere (i.e. right putamen and ventral lateral thalamus) predicted longer SSRT. These areas are known to be more strongly activated under conditions that emphasize speed over accuracy (Vincent van Veen et al., 2008). Greater preparatory activity in right putamen and ventral lateral thalamus appears to reflect a strategy that emphasizes going over stopping. Therefore, despite the independence of the relative finishing times of go- and stop-processes, there are hidden preparatory processes that are involved in establishing speed-accuracy trade-off.
3.5.3 Potential applications to the study of pathological states

Recent studies of pathological populations that have investigated inhibition-related preparatory activities have found that preparatory activity is more important to the success or failure on a given stop trial than actual stop-phase activity (Bhaijiwala et al., 2014; Cameron et al., 2012; Schwerdtfeger et al., 2013). The additional detail afforded by the current approach could therefore add insight into the nature of inhibitory control deficits in a wide array of disorders.

Several neurological and psychiatric disorders are marked by apparently similar inhibitory control deficits. The addition of neuroimaging data can help to discriminate subtle differences between these kinds of populations that are not observable from behavioural data alone, or imaging approaches that contrast compound trials based on predictions of cognitive models. This study shows two distinct phases of activity that precede the appearance of stop signals, both of which contain regions whose activity regulates distinct components of the go- and stop-processes. The current approach could therefore be used to help identify where in the brain, and at which point in a trial, activity underlying deficient inhibitory control becomes atypical, and how one disorder might differ from another despite apparently similar behavioural deficits. For example, disorders associated with poor top-down control such as ADHD and traumatic brain injury might reflect a failure to recruit top-down control, rather than a failure to implement top-down control (Nigg & Casey, 2005). The ability of the current approach to separate activities that recruit top-down control during initial warning-phases from those that actually implement top-down control during response-phases can be exploited to address these kinds of questions.
Regression analyses suggest candidate regions that might be differentially involved in various disorders marked by apparently similar inhibitory control deficits. For example, in disorders associated with working memory problems such as schizophrenia (Bittner et al., 2014), the current approach could differentiate atypical activity associated with abnormal disengagement of the task-negative network during initial warning-phases (posterior cingulate), from failure to invoke a representation of the stop rule in working memory (superior temporal) or of working memory itself (superior frontal) during response phases. Such a hypothesis could be tested by applying the current approach to a cued-response task that manipulates working memory load, such as n-back or selective stopping tasks. This kind of information is potentially useful for understanding the unique neurophysiology of various disorders, detecting individuals at risk, tracking progress of disorders, monitoring interventions, and discovering genetic and metabolic correlates of abnormal brain states.

### 3.6 Conclusions

This is the first fMRI study to use a design that permits direct estimation of neural responses during brief warning- and response-phases in the SST. Our results indicate that the time course of preparing to stop begins at the onset of the initial warning stimulus, and continues up to the point of response execution or response interruption. Warning-phase activities appear to reflect signalling the need for top-down control, and response-phase activities reflect the implementation of that control. Both of these phases contained activities that affect going and stopping. The dependence of inhibitory control
on distinct warning- and response-phase activities demonstrated here points to the importance of separating these activities from activities after stop signals appear when imaging inhibitory control. The ability to separate these processes with fMRI can aid in the study of top-down control deficits, especially those in which failure to signal top-down control could mimic symptoms resulting from disruptions in actually implementing top-down control, such as ADHD and traumatic brain injury.
4 Error detection in the stop signal task

doi: 10.1016/j.neuroimage.2010.06.056

by Andre Chevrier and Russell J. Schachar
4.1 Abstract

Previous error detection research has focused on error processing functions in the anterior cingulate cortex or on putative reinforcement learning roles of midbrain dopamine pathways. We studied error detection in 14 healthy adult volunteers using a novel fMRI design in the stop signal task (SST), a task which invokes numerous errors in performance and frequent instances of post-error slowing. The imaging design accommodated variability immediately before errors (handedness of response) and immediately after (degree of post-error slowing) using distinct within-trial regressors. This approach revealed a whole-brain view of error detection in a reinforcement-learning pathway. Error detection deactivated the midbrain in the vicinity of dorsal substantia nigra where dopamine neurons originate, and the primary targets of dopamine neurons: dorsal striatum and ventral anterior cingulate. Error detection also deactivated posterior hippocampus, which is highly sensitive to long-term synaptic plasticity effects of dopamine. Errors that led to slowed responses deactivated structures in the reciprocal pathway that are known to exert control over dopamine output, and which have been shown to encode error magnitude: ventral midbrain, ventral striatum, and caudal orbitofrontal cortex. Consistent with the role of these structures in modulating dopamine output, post-error slowing also increased activities in the same structures that deactivated on error detection. These results are consistent with the view that errors deactivate structures that receive input from dopamine neurons, followed by deactivations related to requisite behavioral adjustments in structures that exert control over dopamine output.
4.2 Introduction

Errors play an important role in goal-directed behavior. When the consequences of our behavior are better than expected, associations and response patterns are strengthened (Hebb, 1949). But, when outcomes are worse than expected (e.g. performance errors), adjustment is essential (Rabbitt, 1966). Performance errors initiate neural training signals that alter our response tendencies (Holroyd & Coles, 2002a; Ljungberg et al., 1991). Errors briefly deactivate midbrain dopamine neurons (~100 ms), which carry predictive error signals to various parts of the brain for reinforcement learning (Schultz et al., 1995). Dopaminergic projections are most highly concentrated in the dorsal striatum and ventral bank of the anterior cingulate cortex (ACC) (Goldman-Rakic, 1995; Smith & Bolam, 1990). Dopamine neurons also receive inputs, primarily from the ventral striatum and caudal orbitofrontal cortex (OFC) (Eblen & Graybiel, 1995). These projections exert control over dopamine output, which can modulate stimulus-response associations (Draper, 1974; Horvitz, 2002; Servan-Schreiber, Printz, & Cohen, 1990).

Given that these pathways and activity patterns are derived largely from single cell studies in animals, we aimed to determine if distinct patterns of activity could be revealed in these pathways with fMRI. Activities during response phases and successful stop trials were reported in a previous paper (Chevrier et al., 2007).

Previous imaging studies have shown that activity in structures that modulate dopamine output (caudal OFC and ventral striatum) reflect the magnitude of errors (Knutson et al., 2005; O’Doherty, Dayan, Friston, Critchley, & Dolan, 2003). In reward tasks, error magnitude refers to the difference between the amount of money or food received and
the amount that was expected. Mathematical models used in the study of learning define post-error adjustments in stimulus-response associations in direct proportion to error magnitude (Rescorla & Wagner, 1972). Error magnitude thereby dictates the degree to which behavior is altered after errors. Error magnitude has typically been studied by parametrically manipulating reward expectation and delivery, and then measuring changes in reinforcement signals expected from various learning models. However, basic operant conditioning theory would predict that reinforcement signals that change associations based on reward task errors should also adjust stimulus-response associations on reaction task errors. Instead of manipulating error magnitude using rewards, we use a stop signal task (SST) that generates many errors, which lead to measurable changes in behavior. The SST has one predominant type of error (failure to inhibit a response) and one type of post-error adjustment (response slowing). Therefore we can use the magnitude of adjustment to infer error magnitude, because of the proportional relationship between error magnitude and adjustment. In addition to previous reinforcement learning approaches, imaging studies have largely investigated error-related activities in the ACC based on the notion that the ACC plays a role in monitoring conflicting stimulus-response associations and in detecting performance errors (Botvinick et al., 2001; Fassbender et al., 2004; Garavan, Ross, Murphy, Roche, & Stein, 2002; Gehring & Fencsik, 2001; Kerns et al., 2005; Ridderinkhof, Van Den Wildenberg, Segalowitz, & Carter, 2004; Swick & Turken, 2002; Ullsperger & von Cramon, 2001). However closer inspection of these functions suggest that post-error adjustments and activities in the ACC respond more to the recent history of reinforcement in order to guide goal-directed behavior (Amiez, Joseph, & Procyk, 2005;
In this study, we attempt to capture a whole-brain view of reinforcement signals when errors are detected.

Imaging error detection is a challenge because the deactivation to errors is very brief (~100ms) (Ljungberg et al., 1991), whereas fMRI measures a prolonged (~20s) response with a temporal resolution of seconds. However, fMRI can separate rapid sequences of processes if they do not always occur together, and unfold in a known temporal sequence (Ollinger et al., 2001). We developed an approach that satisfied these criteria by using a stop signal task (SST) in which the sequence of processes on errors is known. Response processes precede error detection, which precedes post-error slowing (Boucher et al., 2007; Logan et al., 1984). We separated error detection from preceding response-activity and subsequent adjustment-activity, and hypothesized that error detection should deactivate structures that receive the most input from midbrain dopamine neurons, i.e. the dorsal striatum and ventral ACC. Secondly, if post-error slowing is related to error magnitude in the SST, then greater post-error adjustments should deactivate dopamine-modulating structures that respond to error magnitude, i.e. the OFC and ventral striatum.
4.3 Materials and Methods

4.3.1 Subjects

Fourteen healthy subjects (8 male, 6 female) ranging in age from 22-35 years (mean 29.4) were studied. All were right-handed, had normal or corrected-to-normal vision and reported no medication use, medical illness or psychological problems. Subjects gave informed written consent to participate in the study, which was approved by our institutional research ethics board.

![Diagram](image)

**Figure 4-1 Stop signal task**

Trials began with a fixation dot (500 ms) followed by a go-stimulus (1000ms), which was either an “X” (left-thumb button response), or an “O” (right-thumb response). 33% of trials contained an adaptive stop signal, half of which could not be stopped.
4.3.2 Behavioral task

The Stop Signal Task (SST) is portrayed in Figure 4.1. Each trial began with a fixation point for 500 ms followed by the stimulus (the letter “X” or “O”) for 1000 ms.

Participants responded with their left thumb if the go stimulus was an X or right thumb if it was an O. A blank screen appeared between trials. Stop-trials (33%) involved a change in screen color from black to red, instructing participants to withhold that particular response. Stop-signal delay was 250 ms, and increased or decreased by 50 ms when subjects succeeded or failed to stop, respectively, ensuring an approximately equal number of failed as successful stop-trials (Logan et al., 1984). Trials were either 2.5 or 3.5 seconds, and jittered with random combinations of 7-bit spread-spectrum binary coding sequences, which are inherently orthogonal. The advantage of using these sequences to jitter our trial lengths lies in the fact that they maximise the number of independent equations in the model matrix, thereby maximising the ability of the deconvolution approach to separate the various event-types. Every fourteenth trial was 19 seconds in order to ensure a well-defined baseline of activity, and to be consistent with the kind of distribution of trial lengths established by Ollinger et al. (Ollinger et al., 2001) for separating processes within a trial, which requires three distinct trial lengths, one of which should be relatively rare and long in duration. The task involved 322 trials (21min 40sec). Go response time was observable from the 67% of trials in which no stop signal appeared. Stop-signal reaction time (SSRT) was estimated by subtracting the mean stop-signal delay from the mean go response time (Logan et al., 1984). We divided errors into those with and without slowing according to a median split of reaction times on the following trial, similar to previous imaging studies that have
divided errors into those that were followed by slowed responses and those that were not (Garavan et al., 2002; Gehring & Fencsik, 2001; Li et al., 2008; Li et al., 2008).

4.3.3 Scanning parameters and statistical analysis.
Imaging was done with a GE LX 1.5T magnetic resonance scanner (General Electric, Milwaukee, USA) using a standard high-quality SPGR anatomic (96 slices, 1.5mm thick, 20cm FOV, 256x256 matrix). Functional data were collected using a gradient echo sequence with a spiral read-out (TE/TR/θ = 40/2000/90, 24 slices, 6 mm thick, FOV 20cm, resolution 3.125²x6 mm³). Response data was collected using a Lumitouch fibre-optic button box (Lightwave Medical, Burnaby, British Columbia). Functional data were analyzed using AFNI (Cox, 1996), and motion corrected using a standard coregistration algorithm. We used a general linear model of stimulus vectors convolved with the hemodynamic response function (HRF) using AFNI’s 3dDeconvolve program. Regressors included estimates of baseline and linear drift along with 6-point HRF’s (12s duration) for each event type (HRF delay = 2TR = 4 seconds).
<table>
<thead>
<tr>
<th>Trial type</th>
<th>Events</th>
</tr>
</thead>
<tbody>
<tr>
<td>GO (left)</td>
<td>X</td>
</tr>
<tr>
<td>GO (right)</td>
<td>O</td>
</tr>
<tr>
<td>Successful stop</td>
<td>S</td>
</tr>
<tr>
<td>Failed stop LT (left)</td>
<td>X, D</td>
</tr>
<tr>
<td>Failed stop LT (right)</td>
<td>O, D</td>
</tr>
<tr>
<td>Failed stop GT (left)</td>
<td>X, D, P</td>
</tr>
<tr>
<td>Failed stop GT (right)</td>
<td>O, D, P</td>
</tr>
</tbody>
</table>

**Table 4-1 Trial types and deconvolution regressors**

GO = go trial; LT = error followed by less than median slowing; GT = error followed by greater than median slowing; X = left thumb response; O = right thumb response; S = successful inhibition; D = error detection; P = post-error slowing.

Trial types and model regressors are portrayed in Table 4.1. Go trials were modelled with a response (X or O) stimulus; successful stop trials were modelled with a stop (S) stimulus; failed stop trials followed by less than median slowing were modelled with a response (X or O) and an error detection (D) stimulus; and errors with greater than median slowing were modelled with a response (X or O), an error detection (D) and a post-error slowing (P) stimulus. Response regressors were time-locked to responses, stop regressors were time-locked to stop signals on successful stop trials, and error detection and adjustment regressors were time-locked to responses on failed stop trials.

The temporal sequence of erroneous responses, error detection and post-error adjustment was not specified in the model, but rather was known a priori – error detection must follow the erroneous response, and post-error adjustment must follow error detection.

BOLD responses are separable if there are a sufficient number of independent equations to uniquely estimate each time course (Ollinger et al., 2001). Response, error detection,
and post-error slowing do not always co-occur, and are therefore separable\(^1\).

Multicollinearity in the experimental design was assessed using the output of AFNI's 3dDeconvolve program, which provides estimates of the matrix-model inverse error for each subject.

Intensity estimate maps for each subject were generated by taking the area under the HRF’s (sum of HRF coefficients), warped into Talairach space, Gaussian blurred (6mm) and resampled at 1 mm\(^3\) resolution. These maps were then analysed with a group random effects ANOVA, with event types (X, O, S, D, P) as levels. The deconvolution approach separated independent estimates of response, error detection and post-error slowing activities at the single-subject stage of analysis, eliminating the need for any contrasting of these regressors in the group ANOVA. ANOVA output (distributed as a t\(^*\) statistic

\[^1\] Example demonstrating the dissociability of response, error detection and post-error slowing regressors in the current experimental design: First, assume that the hemodynamic response function is instantaneous and only lasts 1 TR. Second, assume that go-trials lead to a signal increase of +2 (arbitrary units) for a given voxel, and assign this estimate to the variable ‘x,’ yielding equation 1: \(x = 2\). Third, assume that unsuccessful stop trials followed by less than median slowing lead to a signal decrease of -2. If we assign error detection activity to the variable ‘y,’ then we have equation 2: \(x + y = -2\). Lastly, assume that errors followed by greater than median slowing lead to a net signal increase of +3. If we assign post-error slowing activity to the variable ‘z,’ then we have equation 3: \(x + y + z = 3\). This set of equations undeniably has a unique solution \((x = 2; y = -4; z = 5)\). This set of equations also has a unique solution for any other possible values of equations 1, 2 and 3. The extension of this example involving single point HRF’s to the deconvolution of realistic multiple time-point HRF’s is straightforward because trial type was randomized, which ensured that the sequence of combinations of regressors would not be multicollinear.
with 65 degrees of freedom due to the number of subjects and regressors in the model) was converted to Z-scores and corrected for multiple comparisons according to Gaussian field theory (program AlphaSim in AFNI http://afni.nimh.nih.gov/afni/) for an overall \( \alpha < 0.05 \) as in (Chevrier et al., 2007). Using AFNI’s AlphaSim program, we performed Monte Carlo simulations based on single subject full-width half-maximum (FWHM) estimates from the deconvolution output. The output of this analysis required that significant voxels be part of a larger cluster of at least three original voxels (175.8 mm\(^3\)) with minimum Z score of 3.3 for an overall \( \alpha < 0.05 \). This whole-brain corrected analysis identified regions of significant activity changes during error detection and post-error adjustment and allowed us to compare these whole-brain corrected areas with those that were predicted by current models of reinforcement learning as described in the Introduction. The whole-brain corrected areas that emerged from this analysis were used to generate ROI’s in order to estimate the means and standard deviations of activity in these areas across the other event types in the model. This approach allowed us to investigate activity in regions involved in error detection and post-error slowing during the other phases of the task.

We assessed several assumptions of our model in subsequent analyses. Imaging designs that separate within-trial processes always run the risk of incorrectly attributing the modulation of one process to the presence of a second. If one were to add a second within-trial process (e.g. error detection) to a first process (e.g. button press), and if the responses were in opposite directions (e.g. activation followed by deactivation), then it would be possible to test whether the measured response to the second process might instead reflect the absence of the first process. Therefore we tested whether error
detection deactivations could possibly be explained as the absence of activity that might only occur on go-trials. Similarly, we tested whether post-error slowing deactivations could reflect the absence of activities associated with faster responses after errors. Finally, in order to compare results from the current approach with previous approaches that have not separated errors into detection and adjustment components, we performed an analysis of the data in which errors were modelled as a unitary process. Such approaches have consistently found positive activity in dorsal ACC on errors. Therefore, when errors are not split into detection and adjustment regressors, we should also find increased activity in dorsal ACC.

4.4 Results

Subject performance was consistent with normal adults in non-MRI environments. Go responses were fast (597.7±53.7ms) and the mean stop signal reaction time was normal (210.3±48.0ms) (Schachar, Tannock, Marriott, & Logan, 1995; B. R. Williams et al., 1999). Roughly half of the stop trials (48.8±2.4%) contained erroneous responses (mean error response time = 517.6±75.0ms). The median split approach succeeded in separating errors into those that were followed by slowed responses from those that were not: mean post-error slowing was 69.7 ± 38.1 ms (measured relative to mean go-RT); the mean of those with less than median slowing was -35.0±36.0ms, and greater than median slowing was 104.7±67.1ms.
Assessment of the deconvolution model for multicollinearity showed that the event types were indeed separable, with single subject inverse average errors on the order of $6 \times 10^{-16}$. The separability of our event types is quite logical when one considers: 1) only one fifth of responses contain error detection, 2) only half of these are followed by greater post-error slowing 3) trial type was randomised, 4) trial duration was jittered, and 5) there were 17.5 second rests after every fourteenth trial. Error detection deactivated (whole-brain corrected) the midbrain directly on the atlas coordinates of dorsal substantia nigra (Figure 4.2a: 7,−15,−8), dorsal striatum (Figure 4.2b: right 15,14,3; left −20,12,−1), ventral ACC (Figure 4.2c: −9,−13,33; -9,13,30; and -9,28,12), and posterior hippocampus (Figure 4.2d: 27,−20,−14; −34,−23,−13). Hemodynamic response functions to error detection in these regions can be found in Supplementary Figure 4.3 The only other deactivation on this map was in lingual gyrus (13,-86,-4). The only activation was in right superior temporal gyrus (45,−26,−1). Error detection deactivations (except lingual gyrus) were used to determine the boundaries of ROI’s in order to assess activity in these regions during the other phases of the task. Error detection deactivations were significantly greater than activation during Go-phases in the dorsal striatum ($p = 0.00081$), hippocampus ($p = 0.024$), and rostral ACC (ACC2 $p = 0.022$, ACC3 $p = 0.0047$). This contrast failed to reach conventional levels of significance in dorsal substantia nigra ($p = 0.086$) and caudal ACC (ACC1 $p = 0.084$).
Figure 4-2 Error detection deactivations

Brain activities shown on the left (blue = deactivation) are the result of a whole-brain statistical correction. The only other significant activities on error detection were in lingual and temporal gyri. Graphs on the right represent activity estimates in these ROI’s across all regressors in the experimental design: Press X and Press O - motor responses; Stop - successful stops; Error detect - inhibition errors; and PE-Slowing - errors followed by greater slowing. Darker bars indicate the condition (Error detect) from which these regions were identified. a) dorsal substantia nigra (inferior 8mm), b) dorsal striatum (superior 3mm) c) ACC (left 7mm), and d) hippocampus (posterior 23mm).

A complete list of significant activity changes on post-error slowing is shown in Table 4.2. Post-error slowing significantly (α<0.05, corrected) deactivated the predicted structures, which were used to generate ROI estimates of activity changes across the other phases of the task (Figure 4.3). Post-error slowing deactivated ventral striatum (Figure 4.3b: 21,1,-10; -27,3,-7), and caudal OFC (Figure 4.3c: 26,19,-15; -18,17,-14). Post-error slowing also showed whole-brain significant deactivations in bilateral parts of ventral midbrain (Figure 4.3a: 15,-13 -9; -19,-12,-10). Hemodynamic response functions to post-error slowing in these regions can be found in Supplementary Figure 4.4. Post-error slowing deactivations were significantly greater than activations during error detection in ventral midbrain (p = 0.049) and ventral striatum (p = 0.02). This contrast failed to reach conventional levels of significance in caudal OFC (p = 0.074). It is
important to note that the ventral midbrain deactivations on post-error slowing were spatially distinct from the more dorsal and caudal midbrain region that deactivated on error detection. In fact, the post-error slowing map not only deactivated the ventral midbrain regions shown in Figure 4.3a, but also showed significant increases in activity in the same part of substantia nigra that deactivated on error detection. These (relatively ventral vs dorsal) midbrain responses were spatially distinct, separated by a gap of no less than 3 mm. Therefore, ventral midbrain deactivations on post-error slowing (Figure 4.3a) are anatomically distinct from the relatively dorsal deactivation seen on error detection (Figure 4.2a), and have distinct activity profiles across the various event-types in the model.
Figure 4-3  Selected deactivations on post-error slowing

Images on the left portray whole-brain significant deactivations (blue) in predicted structures when errors were followed by greater than median response slowing. These post-error slowing deactivations were accompanied by widespread activity changes related to processing behavioral adjustment after errors (see Table 4.2 for a complete list of significant post-error slowing activities). Graphs on the right are defined as in Figure 4.2. Darker bars indicate the condition (PE Slowing) from which these regions were identified. a) ventral midbrain (inferior 7mm), b) ventral striatum (inferior 6mm) and c) caudal OFC (anterior 22mm). The negative y-axis is extended to show the full pattern of activity in a) and c).
Table 4-2 Post-error slowing activities

These activities were the result of a whole-brain correction of the post-error slowing condition in the group ANOVA analysis. Post-error slowing activity was separated from other processes at the single-subject deconvolution stage, and did not require the contrasting of any event-types.

In order to contrast our results with the results of previous studies that have found increased activity in dorsal ACC, we performed an analysis in which errors were not split into detect and adjustment triggers. This analysis confirmed that when error detection and adjustment are not modelled with distinct regressors at the deconvolution...
stage, then increased activity is indeed seen in dorsal ACC (6,33,25). The ventral ACC
deactivations we saw on error detection were absent when errors were not split into
detection and adjustment regressors. When we examined the activity in dorsal ACC
using the current model, we found increased activity during response phases (Z=1.7),
weak deactivation on error detection (Z= -0.5), and activation on post-error slowing
(Z=3.2), which combine to form the net increase normally found in the dorsal ACC on
errors.

4.5 Discussion

4.5.1 Error detection and post-error slowing deactivate structures that receive from and project to the midbrain, respectively

The present study used a task that generates a large number of errors and a
deconvolution strategy that separated activity associated with error detection from that
associated with post-error adjustment. We reported activities during response-phases and
successful stops in a previous paper (Chevrier et al., 2007). This is the first study to
separate within-trial response, error detection, and post-error adjustment activities. First,
we identified areas involved in error detection and in post-error slowing using a whole-
brain corrected statistical analysis. Second, we inspected the activities in whole-brain
corrected areas across all phases of the task in order to identify unfolding patterns of
activity.
The main findings were that error detection deactivated the midbrain in the vicinity of dorsal substantia nigra, dorsal striatum and ventral bank of the ACC. This network is in keeping with observations of phasic suppression of dopamine neurons on errors (Crutcher & DeLong, 1984; Fiorillo et al., 2003; Hollerman & Schultz, 1998; Ljungberg et al., 1991; Schultz et al., 1993; Schultz et al., 1995), and the hypothesized role of this pathway in modulating error processes (Amiez et al., 2005; Brown & Braver, 2005; Draper, 1974; Holroyd & Coles, 2002a; Niki & Watanabe, 1979; Shima & Tanji, 1998; Takenouchi et al., 1999). This pathway is known to arise in substantia nigra and project most densely to the dorsal striatum and ventral ACC (Goldman-Rakic, 1995; Smith & Bolam, 1990). Error detection also deactivated posterior hippocampus, which is highly sensitive to the long-term synaptic plasticity effects of dopamine (Li, Cullen, Anwyl, & Rowan, 2003). The only other activities were in lingual and temporal gyri, both of which process visual stimuli (Mesulam, 1998), and likely reflect sensory processing of the stop signal.

Error trials that led to greater slowing deactivated structures that modulate dopamine output (Eblen & Graybiel, 1995) and encode error magnitude (Menon et al., 2007; Murray et al., 2008; Tobler, Fletcher, Bullmore, & Schultz, 2007). Ventral striatum deactivated bilaterally (Figure 4.3b) in locations that vary in activity with the magnitude of errors (Murray et al., 2008; O’Doherty et al., 2003). Bilateral deactivations were present in caudal OFC (Figure 4.3c), which processes error magnitude (O’Doherty et al., 2003; Remijnse et al., 2006). Ventral midbrain deactivated bilaterally (Figure 4.3a) during post-error slowing in regions which correspond closely (< 3mm distance) with previous findings of midbrain responses to rewarding and aversive stimuli (Knutson et
al., 2005; Menon et al., 2007; Murray et al., 2008). Deactivation of these structures is consistent with the proportional relationship between error magnitude and adjustment (Dayan & Niv, 2008; Gallistel & Gibbon, 2000; Rescorla & Wagner, 1972), and previous evidence of deactivation in these structures related to error magnitude (D’Ardenne, McClure, Nystrom, & Cohen, 2008; Menon et al., 2007; J. O’Doherty et al., 2004). Unlike the stark isolation of structures that receive the most dopamine input on error detection, post-error slowing showed widespread activity changes (Table 4.2) because many brain areas should be involved in adjusting behavior after errors.

4.5.2 Limitations and technical considerations

In weighing the novelty and significance of these findings, several possible limitations of the approach should be considered. For example, we assumed that errors followed by less adjustment were detected, but did not lead to substantial neural processing of faster responses. This assumption proved to be correct because we saw no activities on error detection apart from those predicted by reinforcement-based learning. We also tested whether error detection deactivations might represent the absence of successful Go-trial activity. This test was not meant to prove that response activity is the same on errors as on go-trials, although there is strong evidence that it is very similar (Boucher et al., 2007; Logan et al., 1984; van den Wildenberg et al., 2010). Rather, this test confirmed that even if response activity were entirely absent, this still would not explain the level of deactivation we found on error detection. Similarly, we confirmed that post-error slowing deactivations could not be explained by the potential absence of activity related to faster responses. Finally, the tri-phasic response on errors with greater slowing is consistent with observations of decreased activity when errors occur, followed by
increases related to altered behavior (Crutcher & DeLong, 1984; Niki & Watanabe, 1979; Schultz et al., 1993).

Although reinforcement-learning models predict activities in dopamine pathways on errors, we did not expect to see differentiation in the midbrain. Substantia nigra is a small structure, and the dopaminergic aspect (pars compacta) is admittedly not dissociable from the motor aspect (pars reticulata) based the imaging resolution of fMRI alone. However whole-brain correction revealed significant and focal activities on distinct midbrain structures that are consistent with previous observations of more complex responses to errors in ventral compared to dorsal substantia nigra (Menon et al., 2007; Murray et al., 2008; Tobler et al., 2007). Additional confidence that this midbrain activity reflects dopamine activity comes from the fact that the only other activities on error detection were in the primary targets of midbrain dopamine nuclei, and that the activity profile in these areas across the other event-types is consistent with known responses of dopamine neurons under similar conditions (Crutcher & DeLong, 1984; Niki & Watanabe, 1979; Schultz et al., 1993). Several studies have addressed the challenge of imaging midbrain activity by using smaller voxels and employing small-volume corrections that allow for cluster/threshold combinations below whole-brain correction (e.g. (D’Ardenne et al., 2008)). It is true that smaller voxels increase the contrast to noise ratio by reducing partial volume effects. However, smaller voxels are more sensitive to motion, and greatly reduce the signal to noise ratio: the correlation between neural activity and BOLD response drops catastrophically as voxel size drops below 3mm (Kim et al., 2004). Apparently the current approach allowed partial volume effects to summate across subjects and result in detail beyond any single subject’s data.
4.5.3 Comparison with conflict-monitoring theories of ACC function

The majority of error-processing studies have investigated conflict-monitoring hypotheses, in which the ACC serves a putative role in monitoring response conflict and detecting erroneous behaviors (Botvinick et al., 2001; Chevrier et al., 2007; Garavan et al., 2002; Gehring & Fencsik, 2001; Holroyd et al., 2004; Li et al., 2008; Rubia et al., 2003; Ullsperger & von Cramon, 2001). The current finding of exclusively deactivation on error detection is conspicuous when viewed from the perspective of previous imaging research, which has consistently shown activity increases on errors. This apparent contradiction is easily resolved when one considers the difference between previous motivations, experimental designs and structures identified, with those in the current study.

The current approach was motivated by previous evidence of suppressed dopamine activity on errors (Fiorillo et al., 2003; Hollerman & Schultz, 1998), which is known to be highly comparable across error types (Servan-Schreiber et al., 1990). Midbrain dopamine neurons send the densest projections to the dorsal striatum and ventral ACC (Goldman-Rakic, 1995; A. D. Smith & Bolam, 1990). Previous single cell recording and even fMRI studies (Delgado, Locke, Stenger, & Fiez, 2003) have shown that the dorsal striatum responds to the occurrence, but not to the magnitude of errors. There is less direct evidence for this pattern in the ACC, but the rationale for this prediction is sound because of the known anatomical projections from dopamine neurons to the ventral ACC. The fact that all regions that deactivated on error detection,
also activated to some degree on post-error slowing, explains the lack of evidence from all previous studies, which have combined these phases of activity.

In order to confirm that the difference between current and previous results was a consequence of different models, we also performed an analysis that combined error detection and post-error slowing, and found the same activity in dorsal ACC others have found (see Figure 4.4b of (Chevrier et al., 2007)). But when analyzed with the current model, dorsal ACC activated during response and post-error slowing phases, and slightly deactivated on error detection. Therefore, rather than contradicting previous work, our data actually elaborates on the pattern of activity in the ACC on errors. When errors are modeled as a unitary process, slight decreases when errors are detected combine with strong increases during response- and post-error slowing phases, which add to net positive activity in the dorsal ACC. But when errors are split into detection and adjustment, one sees decreases in ventral ACC on error detection and increases related to post-error slowing in ventral and dorsal ACC. Therefore previous approaches have been appropriate for delineating conflict- and error-related activities in dorsal ACC, but have not been appropriate for isolating error detection and post-error slowing activities in ventral ACC. This ventral/dorsal delineation is consistent with the findings of Braver et al. (Braver et al., 2001) in which ACC activity related to erroneous outcomes was observed to be inferior to ACC activity related to response conflict. Our results are consistent with recent evidence that ACC activity can be more complex than conflict-monitoring theory would predict, and that adjustments after errors are better explained by accounts of reinforcement-guided action (Amiez et al., 2005; Bioulac et al., 2005; Holroyd & Coles, 2008; Kennerley et al., 2006; Li et al., 2008; Nakamura et al.,

94
Our results also suggest why previous studies have localized error detection in variable parts of the ACC, with some pointing to rostral, and others to caudal areas of the ACC (Bush et al., 2002; Garavan et al., 2002; Ullsperger & von Cramon, 2001). Figure 4.2c shows activities in caudal (ACC 1), middle (ACC 2) and rostral (ACC 3) regions during the various phases of the task. All three regions deactivated to a similar extent when errors were detected. By contrast, response-phase activities (Press X; Press O) were more intense in the caudal direction (ACC 1 > 2 > 3) and post-error adjustment activities (PE-Slowing) were more intense in the rostral direction (ACC 3 > 2 > 1). This pattern is consistent with previous topographical observations of the ACC, with rostral parts representing predictions about the presence or absence of rewards (negative reward processing on post-error slowing), and caudal regions related to the execution of learned instrumental behaviors (response-phase) (Takenouchi et al., 1999). This rostro-caudal division of labor is also consistent with evidence that response conflict activates caudal ACC whereas error detection and post-error adjustments activate the rostral cingulate zone (Debener et al., 2005; Fassbender et al., 2004; Garavan et al., 2002; Ridderinkhof et al., 2004; Ullsperger & von Cramon, 2001).

Another trend that has been observed in imaging studies is that errors are often preceded by drops in attention, evident in response speeding prior to inhibition errors (Li, Yan, Bergquist, & Sinha, 2007), or response slowing prior to discrimination task errors (Eichele et al., 2008). This effect is associated with suppressed activation in prefrontal
structures that regulate attention and working memory (Eichele et al., 2008), and suppressed deactivation in regions thought be involved in resting states (Li et al., 2007), which are interpreted as decreasing vigilance. This effect is distinct from the deactivations we show on errors in two important ways. Firstly, drops in vigilance are associated with suppressed activity changes preceding go-stimuli, whereas we show significant drops below baseline, which additional statistical tests confirmed could not be explained as the simple reduction of preceding activities. Secondly, activity decreases related to less controlled responses have not been reported in the ventral ACC regions reported here.

4.5.4 Connection between reinforcement signals and task-specific networks

Our results support a reinforcement learning approach to errors in the SST. Reinforcement learning is usually studied with reward tasks in which error magnitude is under experimental control. However, the same circuits that process prediction errors in reward tasks should also adjust instrumental behaviors on performance errors (Dayan & Niv, 2008; Gallistel & Gibbon, 2000; Rescorla & Wagner, 1972). The SST maintains an error rate of 50%, which maximises natural variations in predictions and subsequent error magnitudes (Dayan & Niv, 2008; Fiorillo et al., 2003). Post-error slowing is a logical index of error magnitude because reinforcement learning defines adjustment in proportion to error magnitude both in reaction task (Gallistel & Gibbon, 2000) and reward task errors (Dayan & Niv, 2008). One important question about the neural basis of learning is how limbic information about changing motivational context comes to be expressed in structures that convey instrumental behavior. In this section, we examine
the known functional roles of the deactivations we saw on error detection, and those on post-error adjustment. We then suggest how these patterns combine to impart limbic information about changing motivational context in structures that express instrumental behavior.

The proposition that error-related dopamine signals affect the recruitment of prefrontal task-controllers by the ACC has been suggested by Holroyd and Coles (Holroyd & Coles, 2002a). However, previous investigations of this proposed mechanism have asserted that dopaminergic deactivations on errors somehow lead to positive activity in the ACC. Perhaps this assertion arose because previous imaging research had consistently found activation of the dorsal ACC on errors. By contrast, we show that suppressed dopamine activity on errors actually does deactivate the ventral ACC, followed by activation related to post-error adjustment. Braver and Cohen (Braver & Cohen, 1999) suggested that drops in prefrontal activity, like we observed in ventral ACC on error detection, could interrupt ongoing processing and reset information held in working memory. On a cellular level, a phasic drop in afferent dopamine input increases the gain of pyramidal neurons, which renders them susceptible to a broader range of inputs from more distal dendritic terminals to facilitate a search for more adaptive strategies (Servan-Schreiber et al., 1990; Yang & Seamans, 1996). Similarly, deactivations in the hippocampus could interrupt ongoing representations of the motivational context. The dorsal striatum also deactivated on error detection. Dorsal striatum receives inputs from structures involved in expressing instrumental behaviors, which feed back to their point of origin in the frontal lobes (Middleton & Strick, 2000; Middleton & Strick, 2002). These loops support working memory by reinforcing
recurrent excitation in thalamocortical networks. The primary output gate of the striatum, the globus pallidus, exerts a tonic inhibitory influence on the thalamus. Inhibitory projections from the striatum to the globus pallidus can therefore disinhibit thalamic activity, and directly facilitate the flow of excitation through thalamocortical loops (Dominey, Arbib, & Joseph, 1995; Frank, Loughry, & O’Reilly, 2001; Houk, 1997). Therefore deactivation of the dorsal striatum when errors are detected can interfere with reinforcement of active motor programs that require adjustment. Deactivation of this pathway might not only serve to interrupt ongoing behavior, but could also induce spike-timing dependent plasticity in these neurons so that post-error adjustment activities can influence subsequent behavior (Bliss and T. Lomo, 1973; Hebb, 1949; Song, Miller, & Abbott, 2000). Deactivation of the dorsal striatum on errors would temporarily decrease the firing rate of post-synaptic neurons, thereby decreasing the probability of “post-pre” events, and enhancing the net plasticity effects of “pre-post” events conveying post-error adjustment activities. This interpretation is consistent with the emergence of maladaptive behaviors associated with drugs that increase dopamine activity in the dorsal striatum, which might prevent its deactivation when errors are detected (Djamshidian, Cardoso, Grosset, Bowden-Jones, & Lees, 2011).

Post-error slowing deactivated structures that modulate dopamine neurons, caudal OFC and ventral striatum. It has been suggested that the OFC and ventral striatum might convey new behaviors to the dorsal striatum because they encode motivational context, and send direct and indirect (through the midbrain) projections to the dorsal striatum. These projections have been referred to as a “limbic/motor interface,” which could funnel information from the limbic system to the motor system (Mogenson, Jones, &
Yim, 1980; Nauta, Smith, Faull, & Domesick, 1978). However, the exact functional relationship between the cortex and striatum is still uncertain. Procedural learning theories postulate that habits are represented directly in the striatum, which is adaptively trained by the cortex (Graybiel, Aosaki, Flaherty, & Kimura, 1994; Packard & Knowlton, 2002). Atallah et al. showed that rodents could learn when the dorsal striatum was disengaged with drugs, but could not express this learning until the dorsal striatum was once again functioning, which would seem consistent with behaviors being encoded directly in the dorsal striatum. However, once a new behavior is carried out with an intact dorsal striatum, it is no longer needed for the new behavior (Atallah, Lopez-Paniagua, Rudy, & O’Reilly, 2007). Therefore the impact of teaching signals on the dorsal striatum establishes new associations in other networks, not in the striatum itself. This result would support theories that the striatum disambiguates competing cortical representations in order to choose and implement appropriate strategies (Frank et al., 2001; Houk, 1997; Pasupathy & Miller, 2005; Seger & Cincotta, 2006).

If the OFC and ventral striatum do not convey explicit information to the dorsal striatum, then how might limbic information impact the motor system on errors? Firstly, there exists a series of lateral connections between the striatum and midbrain (“striatonigrostriatal system”), from limbic, through cognitive, and finally to motor regions of the basal ganglia (Haber, Fudge, & McFarland, 2000). These connections could infuse motor programs passing through the dorsal striatum with both limbic and cognitive information, before returning to their point of origin in the cortex. Secondly, rather than imprinting new behaviors in the dorsal striatum, the OFC and ventral striatum have a “threshholding” influence on the dorsal striatum, by suppressing the
transmission of weak inputs and amplifying the transmission of strong inputs (Horvitz, 2002). The proposition that this thresholding influence can establish unidirectional (limbic → cognitive → motor) associations in distributed cortical networks is consistent with the fact that post-error adjustment was associated with widespread activity changes throughout the cortex. Such a mechanism is consistent with previous assertions that dopamine deactivations on errors can interrupt processes involved in the error (Amiez et al., 2005; Brown & Braver, 2005), and induce plasticity conditions that can alter future behavior (Kandel, Klein, Castellucci, Schacher, & Goelet, 1986). Therefore the relationship between the cortex and striatum is likely more complex than one simply training the other. Rather, the thresholding influence of ventral cortico-striatal activities on convergent cortical and striatoniigrostriatal information passing through the dorsal striatum could generate new functional interactions in networks involved in expressing instrumental behaviors. Although highly speculative, the higher-order components of such an interference pattern (Willshaw, Buneman, & Longuet-Higgins, 1969) could provide a compact subcortical mechanism for generating the sort of higher-order cortical interactions that are the basis of brain function in general (Friston, 2002). Such a mechanism would explain why disruptions in subcortical structures, like the degradation of nigrostriatal projections in Parkinson’s disease, lead to maladaptive functional interactions in cortical networks. Future research is necessary to determine the precise nature of the relationship between the cortex, striatum and reinforcement signals during learning and adaptive behavior.
4.6 Conclusions

Our results indicate that error detection and post-error slowing involve distinct phases of deactivation in structures that receive the most projections from, and send the most projections to midbrain dopamine neurons, respectively. By separating error-related activity from preceding and following activities, the approach offers a whole-brain view of error detection that has not previously been reported, and possibly a new way to investigate parametric manipulations of value, drugs, and disease. We show patterns of activities during response, error detection and adjustment phases in central reinforcement learning structures that would be expected in a wide-array of tasks involving reinforcement-guided action. The current approach is potentially well suited to the study of disorders marked by poorly controlled behavior arising from suspected dopamine dysfunction, from Parkinson’s disease to attention deficit hyperactivity disorder and substance abuse.
5  fMRI of error detection and post-error slowing in adolescent ADHD

Manuscript in preparation

by Andre Chevrier, Mehereen Bhaijiwala, Douglas Cheyne, Simon Graham and Russell Schachar
5.1 Abstract

Error detection is associated with deactivations in dorsal striatum and ventral ACC, and post-error adjustment is associated with deactivations in ventral striatum and orbital frontal cortex (OFC). These activities are crucial to reinforcement learning. Children with ADHD show impairments during error processing. However, it is unclear at present if the deficit begins during the detection of an error or the adjustment processes that follow. These processes are not separable based on behavioral data alone. This study used a stop signal task to separate neural activity during error detection and post-error slowing in adolescents with and without ADHD. Typically developing (TD) adolescents deactivated dorsal striatum and ventral ACC during error detection, and ventral striatum and OFC during post-error adjustment, similar to healthy adults. By contrast, ADHD adolescents only partially deactivated the dorsal striatum and ventral ACC during error detection but activated dorsal ACC normally. ADHD adolescents then activated instead of deactivated ventral striatum and OFC during post-error adjustment. Atypical activity in ventral ACC and normal activity in dorsal ACC, combined with opposite recruitment of the same prefrontal resources during preparation and post-error adjustment, suggest that ADHD adolescents may possess similar insight to TD adolescents regarding the detection of errors, and the identification of appropriate neural resources. However, the reinforcement of these resources appears to be opposite in ADHD compared to TD adolescents. This atypical reinforcement might explain the delayed maturation of function and cortical thickness in ADHD, and the lack of potency of behavioral interventions that are based on the known properties of normal reinforcement processing.
5.2 Introduction

Errors are common when performing tasks that require both speed and accuracy.
Success in such circumstances requires the ability to monitor performance, detect errors
(Shiels & Hawk, 2010) and make appropriate adjustments in order to reduce the
likelihood of future errors. Typically, errors result in slower responses on the following
trial, referred to as post-error slowing (Rabbitt, 1966). Several pathological conditions
(Manoach & Agam, 2013) including attention deficit hyperactivity disorder (ADHD) are
categorized by apparent deficits in error processing. Individuals with ADHD are
thought to have difficulties with error processing because they exhibit greater response
variability, make more errors in speeded tasks and slow less following errors
(Castellanos-Ryan et al., 2014; Schachar et al., 2004; Sergeant & Van der Meere, 1988;
Shiels & Hawk, 2010). In this study, we used fMRI to separate neural activity when
errors are detected from the adjustment that immediately follows in order to gain a better
view of how error processing differs in ADHD compared to TD adolescents.

Previous studies of error processing in ADHD have conceptualized errors as a unitary
process (i.e. not as error detection followed by post-error adjustment), and have
consistently shown reduced error signaling in ADHD. Electrophysiological approaches
have focused on the error-related negativity (ERN), which occurs 200-400s after errors
are detected (Gehring et al., 2011; Hajcak et al., 2005). The ERN is thought to reflect
error detection activity in the dorsal ACC, which is a strong limbic-cognitive interface
(Holroyd & Coles, 2002a). Using a stop signal task, Liotti et al. (Liotti et al., 2005)
found that the ERN was markedly reduced in ADHD subjects, which these investigators
interpreted as reflecting a “global deficit in cognitive control operations subserved by
dorsal ACC in ADHD.” Other fMRI studies have also found that ADHD subjects show reduced activity on errors in ACC and IFG (Vasic et al., 2012), precuneus and posterior cingulate cortex (Rubia, Smith, Brammer, Toone, & Taylor, 2005), and prefrontal, thalamic and parietal regions (Rubia, Halari, Mohammad, et al., 2011).

There is substantial evidence that deficient error signaling in ADHD arises as a result of atypical reinforcement learning due to impairment in the midbrain dopaminergic system (Bellgrove et al., 2006; Luman et al., 2008; Sagvolden et al., 2005; Volkow et al., 2009). Impaired release of phasic dopamine in ADHD has been proposed to result in weaker anticipatory and reinforcement signals to reward and prediction errors, thereby diminishing attention to necessary information for appropriate decision making, and the impact of feedback information on subsequent behavior (Sagvolden et al., 2005; Tripp & Wickens, 2008).

Several lines of evidence point to atypical dopamine signaling as the primary deficit in ADHD. PET studies have shown that ADHD subjects show reduced binding potential for dopamine transporter in midbrain regions where dopamine neurons originate, which results in altered dopamine signaling (Jucaite et al., 2005). Lesions in ventral putamen, which is richly innervated by dopamine neurons, increase the risk of ADHD traits (Max et al., 2002). Altered dopamine is also implicated from animal models, but the nature of the influence remains unclear; findings in dopamine-depleted mice implicate reduced dopamine function associated with ADHD, whereas findings in dopamine transporter knock-out mice suggest increased dopamine function (Ohno, 2003). Further, stimulant medications that are known to increase extra-cellular dopamine levels also improve
behavior in ADHD subjects (Steele et al., 2006), and have been shown to normalize brain activation on errors (Rubia, Halari, Mohammad, et al., 2011).

Dopamine activity reflects variables associated with prediction errors rather than the details of the task. Therefore, a more direct approach for investigating dopamine driven activity has been to use reward tasks. In reward tasks, the magnitude of prediction errors, calculated as (value of actual outcome – value of expected outcome), is under experimental control. Activity related to error magnitude has been demonstrated in the midbrain, where dopamine neurons originate, and in OFC and ventral striatum (D’Ardenne et al., 2008; Knutson et al., 2005; O’Doherty et al., 2004), which have the strongest modulatory influence on midbrain dopamine activity. Reward task studies in ADHD have reported abnormal activity in medial prefrontal, and ventral and dorsal striatum during reward anticipation and delivery (Furukawa et al., 2014; Hauser et al., 2014), and a failure of OFC to encode prediction error magnitude as in healthy control subjects (Wilbertz et al., 2012). However, no previous ADHD study has modeled errors as a biphasic process, comprised of error detection followed by post-error adjustment related to error magnitude.

In a previous study, we developed an approach to separate neural activity associated with the detection of an error from that associated with adjustment that follows, which is related to error magnitude (Chevrier & Schachar, 2010). Allowing for a biphasic response to errors revealed adjustment-related activity in caudal OFC and ventral striatum, which modulate activity in the ascending dopamine pathway, and are known to respond to prediction error magnitude. This is consistent with the fact that adjustments
to errors should be in direct proportion to the magnitude of the prediction error in instrumental tasks as well as reinforcement learning and reward tasks (Gallistel & Gibbon, 2000). Allowing for a biphasic model of error processing in an instrumental rather than reward task also revealed several patterns of activity in dopamine-driven networks that have not been accessible to previous approaches. We found that error detection, when separated from error magnitude processing, was associated with deactivation of substantia nigra in the midbrain, where dopamine neurons originate, and their primary targets, namely the dorsal striatum and ventral ACC (Goldman-Rakic, 1995; Smith & Bolam, 1990), and in the hippocampus (HPC), which is the most sensitive to the modulatory influence of dopamine (Eblen & Graybiel, 1995). We also found that deactivations on error detection were followed by increased activity in these same regions during post-error adjustment. This biphasic response explains why previous approaches that have modeled errors as a unitary process, in which deactivations during error detection would be combined with activation related to error magnitude, have not found this pattern of activity.

In the present study, we use the same methodology for separating neural activity associated with error detection from that associated with adjustment that follows to investigate error detection and post-error slowing in adolescents with and without ADHD. We predict that TD adolescents will show similar activation patterns to what was observed in healthy adults (Chevrier & Schachar, 2010), whereas children with ADHD will show atypical activity during both error detection and error modification compared to TD adolescents.
5.3 Methods

5.3.1 Participants

28 subjects were included in this study (14 adolescents diagnosed with ADHD and 14 TD control adolescents) between the ages of 9-18 years. Participants gave informed, written consent and the study was approved by the Hospital for Sick Children institutional research ethics board. ADHD participants (n=14) on stimulant medication were asked to stop administration 24 hours prior to the scan in order to eliminate drug-induced BOLD changes (Dodds et al., 2008). History of stimulant medication use was documented in order to determine any treatment effects on task performance or neural activity.

Participants and their parents were interviewed separately and together using the parent interview for child symptoms (PICS-IV (Ickowicz et al., 2006)). Intelligence was assessed using the Wechsler intelligence scale for children (WISC-IV). ADHD subjects met the diagnostic and statistical manual of mental disorders (DSM-IV) criteria for ADHD, defined as having at least six of nine inattentive symptoms, six of nine hyperactive-impulsive symptoms, or both according to at least two of three informants (parents, teacher and/or patient self-report). ADHD subjects also showed moderate to severe impairment in both school and home settings (Global Assessment Scale (Shaffer et al., 1983) score of less than 60). Participants were excluded if they had any comorbid psychiatric or neurological disorder other than oppositional defiant disorder (ODD) or learning disability within the previous 12 months (e.g., obsessive compulsive disorder,
Tourette syndrome, major depressive, anxiety or pervasive developmental disorder), an IQ score of below 80 on both verbal and performance scales or any medical issues that would impact fMRI participation. Subjects with contraindications for MRI (metal braces or metal fragments in their body) were also excluded.

Nine ADHD subjects were diagnosed with ADHD combined subtype, five met the criteria for the inattentive subtype, and two also met DSM-IV criteria for ODD. Control subjects were assessed in a comparable manner and reported no psychiatric or medical disorders. All subjects were right-handed and had normal vision and hearing.

5.3.2 Behavioral task

The stop-signal task (SST) (Logan et al., 1984) involves a primary choice reaction time task and a secondary stop task. Each trial began with a fixation point which appeared in the center of a black screen for 500 ms, followed by the go-stimulus for 1000 ms. Participants were instructed to respond as quickly and as accurately as possible with their left thumb using an fMRI-compatible response box when the letter “X” appeared on the screen or with their right thumb when the letter “O” appeared. In 33% of the trials, a stop signal (background color change from black to red) followed the go stimulus by some delay. Participants were told to stop their response if they saw the stop signal. They were previously instructed to wait for stop signals.

The initial stop-signal delay was 250 ms and was adjusted dynamically following each stop signal. When the participants successfully inhibited a response, the delay was increased by 50 ms on the next stop trial and when they failed to stop a response, the delay was decreased by 50 ms. Dynamic adjustment of the stop signal onset tracked to
the delay at which, on average, subjects could stop 50% of responses when a stop signal was presented. Inter-trial interval (ITI) was jittered such that trials were either 2.5 or 3.5 seconds to ensure no multicollinearity of event types. The trials were jittered using random combinations of spread-spectrum binary coding sequences to maximize the number of independent equations in the deconvolution analysis, which enhanced the separation of the event-types in the experimental design. Every fourteenth trial was followed by a 17.5 s rest in which no stimuli were presented to establish a well-defined baseline of neural activity.

Trial order was pseudorandomized so that the current type of trial did not predict the subsequent kind of trial. The task involved 224 trials, requiring a total scan time of 15 minutes. The mean go response time (RT) was observable from the 67% of trials in which no stop signal appeared. The stop-signal reaction time (SSRT) was estimated by subtracting the mean delay on stop signal trials from the mean go RT on no-signal trials.

5.3.3 Scanning Parameters and Data Analysis

Imaging was done with a GE LX 1.5T MRI scanner (GE Healthcare, Waukesha, WI). Anatomical data were acquired with a standard high-quality SPGR sequence (120 slices, 1.5-mm thick, FOV 24 cm, 256 x 256 matrix). Functional data were collected using a GRE-EPI sequence with an 8-channel head coil (TE = 40; TR = 2,000; Flip angle =90 degrees; 24 slices; 6-mm thick; FOV 24 cm; 100-kHz readout bandwidth). These images were reconstructed to a 64 x 64 pixel resolution and final voxel size of 3.75 x 3.75 x 6 mm³. Behavioural data were collected using a Lumitouch fibre-optic button box (Lightwave Medical, Burnaby, BC, Canada) interfaced to a laptop running the SST.
Functional data were analyzed using AFNI (Cox, 1996). Images were motion corrected using a standard co-registration algorithm and estimated motion parameters were inspected to ensure that the amount of absolute motion did not exceed 2 mm$^3$ or angular displacement greater than 2 degrees. We used a general linear model of stimulus vectors convolved with the hemodynamic response function (HRF) using AFNI’s 3dDeconvolve program. Estimates of baseline and linear drift were generated along with 6-point HRF’s.

The following event types were used in the deconvolution analysis as in (Chevrier & Schachar, 2010): 1) fixate (F), time-locked to the presentation of warning-stimuli at the beginning of every trial, left- (X) and right-hand response (O) stimuli, time locked to responses, successful inhibition (SI), time-locked to the presentation of stop signals, and error detection (Detect) and post-error slowing (PES) stimuli, both time-locked to responses on failed stop trials. Go trials were modeled using fixate (F) and left- (X) or right-hand (O) response stimuli time-locked to the response. Successful stop trials were modeled using fixate (F) and successful inhibition (SI) stimuli. Failed stop trials followed by less than median response slowing were modeled with fixate (F), response (X or O) and successful inhibition (SI) stimuli. Failed stop trials followed by greater than median response slowing were modeled with fixate (F), response (X or O), error detection (Detect) and post-error slowing (PES) stimuli. These events can be separated because they do not always co-occur and unfold in a known temporal sequence (Ollinger et al., 2001).

Activation maps for each event-type were generated for individual subjects by taking the area under the HRF, warped into Talairach space, Gaussian blurred (6-mm FWHM), and
re-sampled at 1 mm³ resolution. The single subject activation maps were passed on to a random effects ANOVA analysis that was conducted separately for the ADHD and healthy control groups in order to identify the general pattern of whole-brain corrected activity for each group. Maps for ADHD and healthy controls were examined to identify qualitative differences in their patterns of activity.

Group difference maps were generated using a nested repeated-measures 3-factor ANOVA (group membership, event types, and participants) in order to identify significantly different activities between TD and ADHD adolescents. Group difference contrasts (Control-ADHD) for error detection and post-error slowing activities from the ANOVA output were distributed as a t* statistic with 138 degrees of freedom due to the number of regressors and subjects in the study.

Output from all the analyses (ADHD, control, and control-ADHD) were converted to raw Z scores and corrected for multiple comparisons using AFNI's AlphaSim program (Forman et al., 1995) as in Bhaijiwala et al. (2014). This analysis required significant voxels be part of a larger cluster of at least 6 original voxels (540 mm³) with a minimum Z score of 2.32.

Region of interest (ROI) analyses were performed to investigate activity in the putative ascending dopamine pathway during error detection (ventral ACC, dorsal striatum, HPC and midbrain), and in the descending pathway during post-error slowing (ventral striatum, caudal OFC). Locations were determined by the nearest peak of activation in TD adolescents to corresponding regions from Chevrier & Schachar (2010), and confirmed by a user with experience in functional neuroanatomy.
5.4 Results

28 participants were included in the current study. The 14 TD adolescents (age 15.43 ± 1.65) and 14 ADHD subjects (age: 13.56 ±2.19) showed significant age difference \( t(26) = 2.604, p=0.015 \). Post-error slowing in TD (10.27 ±9.03 ms) and ADHD (13.97 ±3.63 ms) groups was not significantly different (p=0.88). Go reaction time for controls (542.2 ±105.7 ms) and ADHD (632.2 ±145.2 ms) groups was also not significantly different (p=0.08), nor were the accuracy of go-responses (p=0.42) or the percent of successful stop trials (p=0.38). All brain activations reported below are whole-brain corrected for multiple comparisons for an overall \( \alpha < 0.05 \) unless otherwise specified.

TD adolescents deactivated the dorsal striatum (figure 5.3) on error detection (left (-20, 12, 0) and right (18, 7, 5), given in Talairach x, y, z coordinates), similar to what has been observed in healthy adults (Chevrier & Schachar, 2010). TD adolescents also deactivated the ventral ACC during error detection in similar locations that have been reported in healthy adults (Chevrier & Schachar, 2010) (ACC1: -18, -13, 33; ACC2: -14, 6, 29; ACC3: -17, 23, 22) (see Figure 5.1). Deactivations were also present in the midbrain and parts of the HPC similar to those reported in our previous paper, but these regions did not survive whole-brain correction for multiple comparisons, and were only marginally significant (i.e. \( p < 0.1 \)). Table 5.1 lists ROI analyses of activities in putative ascending and descending dopamine pathways during error detection and post-error slowing.

The only other deactivations in TD adolescents during error detection were in pre- and postcentral gyri, right superior frontal gyrus (SFG), and cerebellar tonsil. Error detection
was also associated with activation of right dorsal ACC, medial frontal and visual areas, and anterior nucleus of the thalamus. Bilateral activations were present in superior and inferior frontal, and inferior parietal regions. A complete list of activities in TD adolescents during error detection can be found in Supplementary table 5.1.

ADHD adolescents deactivated similar parts of pre- and post-central gyri, and activated similar parts of dorsal ACC as TD adolescents (see Figure 5.2). In contrast to TD adolescents, who deactivated three rostro-caudal regions of ventral ACC during error detection, ADHD adolescents only showed significant deactivation in the most caudal region of ventral ACC, which was only about one third of the intensity of deactivation in TD adolescents (see Table 5.1, Figure 5.3). Also, rather than deactivating the dorsal striatum as in TD adolescents, ADHD adolescents deactivated the ventral striatum and genual medial frontal/ACC BA 10/24/32, which were not present in TD adolescents (Figure 5.1).

The ADHD group showed some deactivation in a small part of left dorsal striatum (Supplementary figure 5.1) during error detection, but this ROI was actually centered on ventral striatum. By contrast, ADHD adolescents exhibited increased activity in the right dorsal striatum (caudate) and thalamus. The ADHD group also exhibited both activations and deactivations in several fronto-parietal and visual regions that were activated in the TD group. Overall, increased activity during error detection was far more expansive in ADHD compared to TD adolescents. A complete list of whole-brain corrected error detection activity in ADHD adolescents can be found in Supplementary table 5.2.
Figure 5-1  Error detection activity in ventral ACC

Error detection activity in TD and ADHD adolescents (shown at left 15mm in Talairach coordinates). Both groups deactivated pre- and postcentral regions. TD adolescents deactivated three regions of ventral ACC (crosshairs positioned on ACC 2 from Table 5-1), whereas ADHD adolescents deactivated ventral striatum and genual ACC/medial frontal cortex. Red = activation, blue = deactivation.

Figure 5-2  Error detection activity in dorsal ACC

Error detection activity in TD and ADHD adolescents (shown at right 5mm in Talairach coordinates). Both groups activated dorsal ACC similarly. This activity extended to the presupplementary motor area in TD but not ADHD adolescents.
Figure 5-3  Error detection activity in ventral ACC and dorsal striatum

Error detection activity in a) ventral ACC (shown at left 15 mm, crosshairs on ACC 2) and b) dorsal striatum (superior 5mm, crosshairs on right dorsal striatum), showing significant deactivation in TD but not ADHD adolescents. DS, dorsal striatum. Error bars denote standard error. Images in radiological coordinates (left = right).
TD adolescents exhibited deactivations during post-error slowing in ventral striatum and OFG (Supplementary figure 5.1) similar to what has been reported in healthy adults (Chevrier & Schachar, 2010), but these deactivations were only marginally significant (see Table 5.1). Deactivations were also present in left caudate tail, middle and superior frontal, and middle temporal gyri. Post-error slowing was also associated with bilateral activations in the insula, precentral and superior temporal gyri, and in the right ACC, parahippocampus, and cerebellar culmen and vermis. A complete list of whole-brain corrected activities in TD adolescents during post-error slowing can be found in Supplementary table 5.3.

In contrast to observations of deactivations in ventral striatum and OFC in healthy adults and adolescents, ADHD adolescents in this study activated these regions to a similar degree that they deactivated in control subjects (Supplementary figure 5.2). However, these activations did not survive whole-brain correction (see ROI analysis below). The only deactivation on this map was in right IFG, near regions that were positively activated in the TD group. Similar to the error detection map, post-error slowing was associated with widespread activity increases in the ADHD group that were not present in the TD group, involving frontal, temporal, thalamic and cerebellar regions. A complete list of whole-brain corrected activity during post-error slowing in ADHD adolescents can be found in Supplementary table 5.4.
<table>
<thead>
<tr>
<th>Structure</th>
<th>Location</th>
<th>Error detection</th>
<th>Post-error slowing</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>x    y    z</td>
<td>TD</td>
<td>ADHD</td>
</tr>
<tr>
<td><strong>Ascending pathway:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L ventral ACC 1</td>
<td>-18  -13  33</td>
<td>-1.83</td>
<td>***</td>
</tr>
<tr>
<td>L ventral ACC 2</td>
<td>-14  6   29</td>
<td>-2.03</td>
<td>***</td>
</tr>
<tr>
<td>L ventral ACC 3</td>
<td>-17  23  22</td>
<td>-2.10</td>
<td>***</td>
</tr>
<tr>
<td>L dorsal striatum</td>
<td>-18  7  5</td>
<td>-1.87</td>
<td>**</td>
</tr>
<tr>
<td>R dorsal striatum</td>
<td>20   12  0</td>
<td>-2.40</td>
<td>***</td>
</tr>
<tr>
<td>L HPC</td>
<td>-23  -18  -18</td>
<td>-2.51</td>
<td>*</td>
</tr>
<tr>
<td>R HPC</td>
<td>19   -15  -18</td>
<td>-3.49</td>
<td>**</td>
</tr>
<tr>
<td>Midbrain</td>
<td>10   -12  -10</td>
<td>-2.21</td>
<td>*</td>
</tr>
<tr>
<td><strong>Descending pathway:</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>L OFC</td>
<td>-21  16  -16</td>
<td>0.45</td>
<td>*</td>
</tr>
<tr>
<td>R OFC</td>
<td>25   35  -11</td>
<td>0.62</td>
<td>*</td>
</tr>
<tr>
<td>L ventral striatum</td>
<td>-13  -9  -4</td>
<td>1.41</td>
<td>**</td>
</tr>
<tr>
<td>R ventral striatum</td>
<td>10   -4  -5</td>
<td>1.75</td>
<td>**</td>
</tr>
</tbody>
</table>

Table 5-1 Activities in putative dopamine pathways

BOLD activity estimates (arb. units) from TD and ADHD groups in putative dopamine pathways during error detection and post-error slowing. Locations were selected according to peak activations in TD adolescents nearest to deactivations identified in healthy adults from Chevrier & Schachar (2010). *** indicates activities that were within whole-brain corrected clusters of deactivation. ** indicates activities that were ROI significant (p < 0.05). * indicates regions that were only marginally significant (p < 0.1).

Comparison of TD and ADHD activities during error detection (control-ADHD) revealed opposite patterns of activity in frontal, temporal and occipital regions, with ADHD subjects activating regions that were deactivating in TD adolescents. This difference was most intense in left superior frontal gyrus BA 10, with ADHD subjects strongly activating and TD adolescents only marginally activating this region. By contrast, ADHD subjects deactivated left cingulate, superior/middle frontal and cerebellar regions that activated in TD adolescents. ADHD subjects also exhibited
intense deactivations in occipital regions that intensely activated in TD adolescents, and activations in left insula and IFG, right thalamus and STG, and bilateral IFG that showed little activity in TD adolescents. A complete list of activity differences during error detection can be found in Table 5.2.

With the exception of right IFG, which deactivated in ADHD subjects but activated in TD controls, every other significant group difference in activity during post-error slowing was the result of ADHD subjects activating regions that were deactivating in TD adolescents. These regions included middle and superior frontal and temporal gyri and ventral striatum. This effect was most pronounced in superior frontal and posterior visual regions, parahippocampus, and thalamic midline nuclei, with ADHD subjects intensely activating regions that TD adolescents strongly deactivated (right occipital and lingual gyri), moderately deactivated (parahippocampus and thalamus) or did not significantly deactivate at all (left, superior frontal, fusiform and superior temporal gyri). A complete list of whole-brain corrected activity differences during post-error slowing can be found in Table 5.3.
<table>
<thead>
<tr>
<th>Structure</th>
<th>BA</th>
<th>Position</th>
<th>Difference</th>
<th>Z-score</th>
<th>TD</th>
<th>ADHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bi cingulate cortex</td>
<td>31</td>
<td>-1 -25 40</td>
<td>5.0</td>
<td>3.7</td>
<td>2.6</td>
<td>-2.4</td>
</tr>
<tr>
<td>L SFG/MFG</td>
<td>8</td>
<td>-21 21 45</td>
<td>5.3</td>
<td>3.0</td>
<td>1.5</td>
<td>-3.9</td>
</tr>
<tr>
<td>L cingulate cortex</td>
<td>31/27</td>
<td>-26 -21 37</td>
<td>-2.1</td>
<td>-2.8</td>
<td>-1.2</td>
<td>1.0</td>
</tr>
<tr>
<td>R cingulate cortex</td>
<td>31/24</td>
<td>18 -22 35</td>
<td>-2.6</td>
<td>-3.2</td>
<td>-1.2</td>
<td>1.6</td>
</tr>
<tr>
<td>L precentral gyrus</td>
<td>6</td>
<td>-33 -8 28</td>
<td>-2.4</td>
<td>-2.5</td>
<td>-3.9</td>
<td>1.5</td>
</tr>
<tr>
<td>R precentral gyrus</td>
<td>6</td>
<td>28 3 29</td>
<td>-2.1</td>
<td>-2.9</td>
<td>-4.7</td>
<td>2.6</td>
</tr>
<tr>
<td>L MFG/IFG</td>
<td>46/9</td>
<td>-39 18 25</td>
<td>-5.3</td>
<td>-2.8</td>
<td>-2.8</td>
<td>2.6</td>
</tr>
<tr>
<td>L MFG</td>
<td>10</td>
<td>-36 52 9</td>
<td>-11.6</td>
<td>-3.3</td>
<td>-4.7</td>
<td>6.8</td>
</tr>
<tr>
<td>L IFG</td>
<td>44</td>
<td>-52 13 14</td>
<td>-6.2</td>
<td>-3.5</td>
<td>-2</td>
<td>6.0</td>
</tr>
<tr>
<td>L IFG</td>
<td>10/46</td>
<td>-35 35 14</td>
<td>-6.0</td>
<td>-3.8</td>
<td>-3.3</td>
<td>2.6</td>
</tr>
<tr>
<td>L IFG</td>
<td>10</td>
<td>-33 52 15</td>
<td>-10.2</td>
<td>-2.5</td>
<td>-2.2</td>
<td>8.0</td>
</tr>
<tr>
<td>R IFG</td>
<td>46</td>
<td>46 40 10</td>
<td>-8.5</td>
<td>-3.2</td>
<td>-2.0</td>
<td>6.3</td>
</tr>
<tr>
<td>R IFG</td>
<td>47</td>
<td>41 15 8</td>
<td>-8.1</td>
<td>-5.1</td>
<td>-1.3</td>
<td>6.7</td>
</tr>
<tr>
<td>R IFG</td>
<td>44/45</td>
<td>50 16 20</td>
<td>-6.6</td>
<td>-3.4</td>
<td>-1.1</td>
<td>5.5</td>
</tr>
<tr>
<td>R STG</td>
<td>22</td>
<td>62 -47 13</td>
<td>-10.0</td>
<td>-2.4</td>
<td>0.9</td>
<td>11.0</td>
</tr>
<tr>
<td>R NA/pHPC</td>
<td>-</td>
<td>26 -3 -11</td>
<td>-5.9</td>
<td>-2.3</td>
<td>-1.2</td>
<td>4.7</td>
</tr>
<tr>
<td>L insula</td>
<td>13</td>
<td>-33 4 -8</td>
<td>-6.6</td>
<td>-2.4</td>
<td>0.1</td>
<td>6.6</td>
</tr>
<tr>
<td>R occipital gyrus</td>
<td>18</td>
<td>26 -86 12</td>
<td>20.1</td>
<td>2.8</td>
<td>11.6</td>
<td>-9.3</td>
</tr>
<tr>
<td>R thalamus, lp</td>
<td>-</td>
<td>22 -19 11</td>
<td>-3.1</td>
<td>-2.4</td>
<td>-0.7</td>
<td>2.5</td>
</tr>
<tr>
<td>L cerebellar tonsil</td>
<td>-</td>
<td>-25 -58 -33</td>
<td>-10.4</td>
<td>-3.1</td>
<td>1.7</td>
<td>-8.6</td>
</tr>
<tr>
<td>R cerebellar tonsil</td>
<td>-</td>
<td>19 -60 -31</td>
<td>-8.6</td>
<td>-2.8</td>
<td>1.3</td>
<td>7.3</td>
</tr>
</tbody>
</table>

Table 5-2 Group activity differences during error detection (TD-ADHD)

Difference, TD and ADHD columns show BOLD signal estimates (arb. units). Locations are given in Talairach coordinates, Left(-) /Right(+) Posterior(-)/Anterior(+) Inferior(-)/Superior(+) BA, Brodmann area; Bi, bilateral; L, left; R, right; SFG, superior frontal gyrus; MFG, middle frontal gyrus; IFG, inferior frontal gyrus; STG, superior temporal gyrus; NA, nucleus accumbens; lp, lateral posterior nucleus.
<table>
<thead>
<tr>
<th>Structure</th>
<th>BA</th>
<th>Location</th>
<th>Difference</th>
<th>Z-score</th>
<th>TD</th>
<th>ADHD</th>
</tr>
</thead>
<tbody>
<tr>
<td>L MFG</td>
<td>6</td>
<td>-21 5 51</td>
<td>-6.8</td>
<td>-2.8</td>
<td>-4.7</td>
<td>2.8</td>
</tr>
<tr>
<td>L SFG</td>
<td>10</td>
<td>-19 65 25</td>
<td>-13.9</td>
<td>-1.7</td>
<td>-1.0</td>
<td>12.4</td>
</tr>
<tr>
<td>R MFG</td>
<td>8</td>
<td>34 35 41</td>
<td>-12.4</td>
<td>-2.9</td>
<td>-7.4</td>
<td>4.8</td>
</tr>
<tr>
<td>R IFG</td>
<td>44</td>
<td>45 17 8</td>
<td>12.3</td>
<td>2.4</td>
<td>7.2</td>
<td>5.1</td>
</tr>
<tr>
<td>L STG</td>
<td>22</td>
<td>-60 -36 18</td>
<td>-11.0</td>
<td>-3.1</td>
<td>-1.4</td>
<td>8.7</td>
</tr>
<tr>
<td>L MTG</td>
<td>21</td>
<td>-60 -31 -8</td>
<td>-10.5</td>
<td>-2.8</td>
<td>-5.6</td>
<td>5.1</td>
</tr>
<tr>
<td>R MTG</td>
<td>21</td>
<td>62 -30 -13</td>
<td>-10.5</td>
<td>-2.5</td>
<td>-5.3</td>
<td>5.6</td>
</tr>
<tr>
<td>L pHPC</td>
<td>36</td>
<td>-39 -21 -15</td>
<td>-11.7</td>
<td>-3.3</td>
<td>-2.4</td>
<td>9.2</td>
</tr>
<tr>
<td>L pHPC</td>
<td>34</td>
<td>-27 6 -16</td>
<td>-12.8</td>
<td>-2.8</td>
<td>-5.2</td>
<td>8.0</td>
</tr>
<tr>
<td>Bi precuneus</td>
<td>7</td>
<td>1 -62 50</td>
<td>-24.1</td>
<td>-2.6</td>
<td>-4.9</td>
<td>19.4</td>
</tr>
<tr>
<td>L fusiform gyrus</td>
<td>37</td>
<td>-43 -63 -4</td>
<td>-10.1</td>
<td>-1.7</td>
<td>-0.5</td>
<td>9.8</td>
</tr>
<tr>
<td>R inferior occipital</td>
<td>18</td>
<td>33 -84 -14</td>
<td>-30.4</td>
<td>-2.7</td>
<td>-14.4</td>
<td>17.3</td>
</tr>
<tr>
<td>L lingual gyrus</td>
<td>19</td>
<td>-20 -82 -15</td>
<td>-18.8</td>
<td>-3.7</td>
<td>-3.1</td>
<td>16.5</td>
</tr>
<tr>
<td>R lingual gyrus</td>
<td>18</td>
<td>5 -85 -8</td>
<td>-34.1</td>
<td>-2.6</td>
<td>-16.3</td>
<td>21.2</td>
</tr>
<tr>
<td>R thalamus, mid</td>
<td>-</td>
<td>1 -17 20</td>
<td>-9.8</td>
<td>-2.8</td>
<td>-3.0</td>
<td>6.9</td>
</tr>
<tr>
<td>L ventral striatum</td>
<td>-</td>
<td>-15 -5 -6</td>
<td>-8.4</td>
<td>-2.7</td>
<td>-3.5</td>
<td>5.2</td>
</tr>
</tbody>
</table>

**Table 5-3**  Group activity differences during post-error slowing (TD-ADHD)

MTG middle temporal gyrus; pHPC, parahippocampus.
5.5 Discussion

This is the first study to separate error detection from post-error slowing in adolescents with and without ADHD. We found that TD adolescents replicated previous findings in healthy adults (Chevrier & Schachar, 2010), with deactivations in the midbrain, HPC, dorsal striatum and ventral ACC during error detection, and deactivation in ventral striatum and OFC during post-error slowing. By contrast, ADHD adolescents did not fully deactivate the dorsal striatum and ventral ACC during error detection, and activated instead of deactivated the OFC and ventral striatum during post-error slowing. Previous studies have concluded that ADHD may result from abnormal reinforcement processing over the course of development (Jucaite et al., 2005; Liston, Cohen, Teslovich, Levenson, & Casey, 2011; Umemoto et al., 2014). The current results add detail to previous theories of altered reinforcement processing in ADHD during moment-to-moment feedback. Abnormal signaling in dopamine pathways after errors would correspond to an opposite process of reinforcement in ADHD compared to TD adolescents, which might result in the opposite preparatory activity we observed in these subjects when engaging top-down control (Bhaijiwala et al., 2014). The development of executive functions and their associated cortical resources are a direct consequence of their use and reinforcement over the course of experience. By providing a novel view of reinforcement learning activities in the context of an inhibition task, the results in this study provide new perspectives on how differences in executive functions associated with ADHD could emerge as a result of a more fundamental disturbance to central reinforcement mechanisms.
The striatal deactivations we observed in TD adolescents on errors have been referred to as the “most solid reward response in the brain” (Schultz, 2015). In TD adolescents, appropriate reinforcement signals in ascending and descending dopamine pathways were accompanied by activation in IFG and parietal areas, and deactivation of other task-related areas. Parietal regions and the IFG are two of the most strongly implicated brain structures in the efficient performance of inhibitory control. Reinforcement of activity in two of the most implicated structures in inhibitory control and deactivation of other regions is consistent with a signal-to-noise enhancing function of dopamine, which suppresses weaker activity and enhances stronger activity. This is thought to constrain neural processing to the most relevant details during performance of goal-directed behavior and over the course of learning (Kroener, Chandler, Phillips, & Seamans, 2009).

The right IFG deactivated instead of activated in ADHD adolescents during post-error adjustment, which is consistent with previous observations of diminished activity in IFG (Vasic et al., 2012) and parietal regions (Rubia, Halari, Mohammad, et al., 2011) on errors. Our results suggest that previous findings of diminished IFG activity on errors actually reflect an opposite reinforcement of this region during prediction error processing. Opposite reinforcement of right IFG on errors could result in opposite signal-to-noise enhancing effects (i.e. reducing instead of enhancing signal-to-noise) in ADHD compared to TD adolescents, thereby suppressing instead of highlighting appropriate networks during preparation. This is consistent with a previous analysis of these ADHD subjects that revealed opposite preparatory activity (deactivation instead of
activation) in right IFG during preparation, but normal activity during successful inhibition (Bhaijiwala et al., 2014).

Deactivation of OFC and ventral striatum in TD adolescents on post-error slowing is consistent with the role of these structures in encoding prediction error magnitude (Brian Knutson et al., 2005; O’Doherty, 2004; O’Doherty et al., 2004). Prediction error magnitude should be proportional to the amount of adjustment in an instrumental task, such as the SST, in the same way that it is proportional to the adjustment of value associations on explicit reward task errors (Gallistel & Gibbon, 2000). Activation instead of deactivation of OFC and ventral striatum in ADHD subjects during post-error slowing is consistent with previous findings that activities in these structures do not encode error magnitude in ADHD subjects as in healthy controls (Wilbertz et al., 2012).

The pattern of error-related activity in structures that project to and receive from midbrain dopamine neurons found in healthy adults and replicated here in TD adolescents has multiple functional effects in distributed circuits involved in goal-directed behavior. Deactivation of midbrain dopamine neurons, like we saw on error detection in TD adolescents, has several effects on their targets, like dorsal striatum, ACC and HPC. Deactivation of the dorsal striatum would interrupt the reinforcement of ongoing behavioral processing (Middleton & Strick, 2000; Middleton & Strick, 2002). Holroyd and Coles (Holroyd & Coles, 2002a) proposed that error-related dopamine signals affect the recruitment of prefrontal task-controllers by the ACC. Therefore, these drops in prefrontal activity on errors could interrupt ongoing processing and reset information in working memory (Braver & Cohen, 1999). Deactivation in the HPC
could similarly serve to interrupt ongoing representations of motivational context. Interruption of processing in the ACC and HPC on error detection could facilitate a shift in attention from ongoing behavioral processing to post-error adjustment processing. At a cellular level, a phasic drop in dopamine input increases the gain of pyramidal neurons, making them more susceptible to inputs from more distal dendritic terminals, which facilitates a search for new adaptive strategies (Servan-Schreiber et al., 1990; Yang & Seamans, 1996).

The blunted deactivation of the midbrain and its primary targets we observed during error detection in ADHD adolescents could hinder the ability to set the conditions necessary for the implementation of post-error adjustments. Failure to engage the appropriate conditions for adjustment at the initial error detection stage could result in several difficulties with responding to feedback such as: difficulty interrupting ongoing processing, leading to perseverative behavior; difficulty resetting information in working memory, leading to maladaptive integration of new feedback-based information; and a failure at a neural level to be susceptible to new adaptive strategies. Difficulties with stopping ongoing behavioral processing and becoming susceptible to new information are reminiscent of impulsive and inattentive symptoms in ADHD.

Altered error processing in ADHD might affect fronto-posterior resetting of the theta rhythm. Theta reset occurs immediately after the detection of errors, precisely when the structures we observed during error detection are known to deactivate. The HPC and ACC, both of which are sensitive to the effects of phasic dopamine activity and both of which deactivated on error detection, are known to be in phase-locked mutual coherence
in the theta band (~7Hz), with HPC leading the ACC. This relationship likely serves to frame prefrontal recruitment of resources by the ACC according to situational context information from the HPC.

Feedback information of any kind, such as errors that are detected in the SST, causes a reset of the mutual coherence between HPC and the ACC (Siapas, Lubenov, & Wilson, 2005). This reset is followed (~150 ms later) by a brief (~ 70 ms duration) increase in spike-timing dependent plasticity while post-error adjustment activity is taking place (Hyman, Zilli, Paley, & Hasselmo, 2010; Hyman, Hasselmo, & Seamans, 2011). During this period of enhanced plasticity, prefrontal neurons that have not previously been entrained into mutual coherence by the HPC can be recruited, and are associated with new functional networks and new behaviors (McCartney, Johnson, Weil, & Givens, 2004). ADHD subjects have been shown to have reduced functional connectivity in widespread cortical networks (Castellanos, Kelly, & Milham, 2009; Castellanos & Proal, 2012; Kerstin Konrad & Eickhoff, 2010; Murias, Swanson, & Srinivasan, 2007; Murias, Webb, Greenson, & Dawson, 2007; Sato, Hoexter, Castellanos, & Rohde, 2012). Thus, the decreased deactivation of the ACC and HPC we observed in ADHD subjects on error detection could interfere with the plasticity effects caused by theta reset on errors, making it more difficult for adjustment-related processing to have persistent effects in these subjects. Difficulty with being susceptible to and integrating new information into future prediction and behavior could underlie attention and learning difficulties encountered in ADHD.
Deactivation of ventral striatum and caudal OFC on negative prediction errors, such as stop task errors, is known to encode prediction error magnitude (Knutson et al., 2005; O’Doherty et al., 2004). Prediction error magnitude is defined as the difference in value or utility between actual and expected outcomes. In theoretical models of machine and animal learning, prediction error magnitude determines the amount of adjustment that occurs after errors are detected (Gallistel & Gibbon, 2000). Deactivation of structures that encode error magnitude in TD adolescents during post-error slowing reflects the fact that errors of greater subjective magnitude are followed by greater adjustment, which in the SST corresponds to greater response slowing. However, activity in these structures was reversed in the ADHD adolescents scanned here, with greater post-error slowing being associated with activation instead of deactivation in OFC and ventral striatum.

In healthy reinforcement learning, deactivation of the OFC and ventral striatum after prediction errors drops the threshold for the passage of information through the dorsal striatum, while adjustment-related processing is taking place (Horvitz, 2002). Consistent with an interpretation of increased processing in the dorsal striatum during post-error slowing, TD adolescents showed increased activity during post-error slowing in the same part of the dorsal striatum that deactivated on error detection. Subcortical reinforcement of ongoing adjustment processing, provided by increased throughput in the dorsal striatum, would coincide with a phasic increase in dopamine related to altered behavior (Niki & Watanabe, 1979; Schultz et al., 1993), thereby enhancing long-term potentiation and depression at their target synapses (Kandel et al., 1986). In healthy individuals, the combined effect of these processes on errors can lead to stable long-term changes in functional networks and behavior.
The opposite patterns of activity we observed in ADHD compared to TD adolescents in structures that convey reinforcement learning signals during post-error slowing could result in adjustment-related processing in task-related networks being reinforced in the wrong way. This is consistent with the lack of potency of behavioral interventions in ADHD (Wolraich et al., 2011). Given an opposite reinforcement response to feedback, attempts at providing ‘therapeutic’ course-corrections based on the properties of healthy reinforcement mechanisms, could potentially lead these individuals farther “off-course”. Therefore, in addition to momentary difficulties dealing with immediate feedback, ADHD adolescents also manifest the long-term effects of a history of opposite reinforcement when trying to prepare these networks for action. The opposite pattern of reinforcement activity in ADHD adolescents reported here is very likely the cause of opposite preparatory activity when engaging top-down control (Bhaijiwala et al., 2014). The opposite reinforcement of networks involved in top-down control would be expected to have generalized effects on executive function as are observed in ADHD, without being caused by an inherent deficit in the prefrontal networks that carry out executive function.

Despite the opposite patterns of reinforcement activities during post-error slowing in ADHD compared to TD adolescents, there was no significant difference between post-error slowing in these two groups. The lack of apparent behavioral differences combined with the presence of striking activity differences in reinforcement learning pathways suggests that slowing after errors might be the result of different mechanisms in ADHD compared to TD subjects. Both error detection and post-error slowing maps showed extensive activations in the ADHD group that were either absent or in the opposite
direction in TD adolescents. Rather than the error serving to interrupt ongoing processing, it instead elevated the intensity of ongoing processing in the ADHD group. These activities are consistent with continued attempts to stop after the error has been detected and adjustment processing has begun. Even in the absence of dysfunctional reinforcement, such an interference effect would cause a lower level of preparation on the next trial similar to the reduced performance that has been observed in the first few trials of interference tasks before task-set stabilizes (Jensen, 1965; Stroop, 1935b). Therefore, our findings suggest how apparently post-error slowing could be the result of entirely different mechanisms. In TD adolescents, post-error slowing can be guided the impact of dopamine-driven reinforcement mechanisms. By contrast, post-error slowing in ADHD adolescents might not be the result of a mechanism per se. Rather, post-error slowing in ADHD subjects could reflect the combined interference between ongoing behavioral processing that is not properly interrupted when errors are detected, with post-error adjustment processing under the influence of an opposite process of reinforcement, thereby reducing the level of preparation and possibly necessitating a re-establishment of the task set on the following trial.

Several studies have presented compelling arguments that ADHD symptoms could arise due to altered dopamine signaling over the course of development. Findings of decreased binding potential for the dopamine transporter in the midbrain in ADHD suggest that dopamine signaling is fundamentally altered in ADHD (Jucaite et al., 2005). Hauser et al. used fMRI during a reversal learning task with many reward prediction errors and found impaired medial prefrontal activity during prediction errors (Hauser et al., 2014). Using competing learning models to fit the behavioral and imaging data and
comparing model parameters between groups, they suggest that altered reward prediction errors in the medial prefrontal cortex drive impaired learning and decision making in ADHD.

A recent study examined the error positivity, an ERP component proposed to index the impact of dopamine reward signals in the ACC, during performance of a maze task (Umemoto et al., 2014). They found that error positivity was greater for money than for points in ADHD subjects but not in control subjects, and concluded that “disruption of the ACC-dopamine interface may underlie the impairments in motivational control observed in childhood ADHD.” Further, a review of functional imaging studies in ADHD (Liston et al., 2011) proposed that existing evidence collectively suggests that “the core symptoms of ADHD might derive from dysregulated modulation of cortical plasticity in the developing brain, resulting in altered patterns of corticocortical connectivity that might persist into adulthood.”

The opposite reinforcement response to errors we observed in ADHD compared to TD adolescents might explain, rather than result from, decreased cortical thickness and contraction of striatum that occurs in ADHD over the course of development (Shaw et al., 2006, 2014). Inhibitory control and concomitant activation of right IFG continue to mature well into adulthood (Bedard et al., 2002; Velanova et al., 2009). Therefore, the fact that inhibitory control is one of the most pronounced deficits in ADHD, and IFG shows pronounced cortical thinning, could be the result of an opposite reinforcement process over the course of development. Functional brain networks that experience
abnormal reinforcement over longer periods of development would be expected to show more pronounced effects.

The current results could also explain lower cerebellar volume in ADHD (Durston, 2003) particularly in the cerebellar vermis. The cerebellar vermis is associated with prefrontal functions that typically develop later such as inhibitory control, and its development is related to the consolidation of prefrontal processing that becomes more automated over the course of learning. Therefore, reduced volume of cerebellar vermis in ADHD might reflect the altered developmental trajectory of executive functions like inhibitory control, which appears to resist the automaticity seen in TD controls.

A recent study (Hart et al., 2014) applied a machine learning algorithm to fMRI activity from subjects performing the SST to discriminate those with and without ADHD and found that the regions that were most predictive of ADHD were in parts of the brain that develop earlier (ventromedial fronto-limbic), whereas regions most predictive of normal controls were in regions that develop later (lateral prefrontal, striatum and temporal-parietal). It is reasonable that ADHD subjects, who experience abnormal preparation and feedback related activity in prefrontal regions that mature later, would rely on networks that developed earlier and activate more normally during stop task performance.

The current finding of normal activation in dorsal ACC in ADHD adolescents is surprising given previous reports of abnormal error-related ERP components thought to reflect the impact of dopamine signals in the dorsal ACC (Liotti et al., 2005; Umemoto et al., 2014). Our results suggest that the response in dorsal ACC is distinct from the impact of dopamine deactivations in the ventral ACC, which is the primary recipient of
midbrain dopamine neurons in the prefrontal cortex (Goldman-Rakic, 1995). fMRI studies of error processing have shown that the dorsal ACC acts as a Bayesian ideal observer (Ide et al., 2013) during prediction and evaluation of outcomes, and is therefore crucial to the identification of task-set and corresponding prefrontal resources required.

The dorsal ACC is both anatomically and functionally connected to prefrontal regions that carry out inhibition (Shenhav, Botvinick, & Cohen, 2013), and activity in dorsal ACC on errors is associated with conscious awareness of errors (Orr & Hester, 2012). These features indicate that the dorsal ACC has some functional control over prefrontal resources that are required given the currently relevant task-set, and provides conscious insight into the occurrence of errors. Therefore it appears that dorsal ACC activities are involved in allocating prefrontal resources to task-set related operations during anticipation and evaluation of prediction errors, whereas ventral ACC activities are involved in reinforcement-learning signals related to prediction error magnitude.

The current finding of normal activity in dorsal ACC and abnormal activity in ventral ACC suggests that mechanisms underlying model-based prediction and evaluation of task-set related information might be intact in ADHD, but that the reinforcement processes conveyed by the midbrain dopamine system have undesired consequences on how task-related networks are reinforced based on prediction error magnitude. This is consistent with our previous observations of ADHD subjects recruiting the same prefrontal resources during preparation as TD adolescents, but rather than activating, they instead deactivated these regions (Bhaijiwala et al., 2014). These results indicate that ADHD deficits might not be an issue of poor insight leading to maladaptive
cognition and behavior. Rather, ADHD might arise from the impact of dysfunctional reinforcement of task-related processing at moments of accurate insight, which results in dysfunctional preparation of relevant task-related structures when they are needed.

Current developmental theories of ADHD argue that factors involving stress and lack of appropriate stimulation at early stages of development lead to physiological and epigenetic changes that cause resources usually used for growth to be diverted into greater stress sensitivity (Miller et al., 2012; Russell, Sagvolden, & Johansen, 2005; Sagvolden et al., 2005; Sagvolden & Johansen, 2012; van der Kooij & Glennon, 2007). These changes are proposed to increase self-reliance and rapid responses to threatening stimuli in potentially hostile or unstable environments. The current findings of ADHD adolescents prematurely deactivating ventral striatum and OFC during error detection could reflect a presumptive error magnitude response, possibly facilitating more rapid reaction to alarming stimuli. Under certain circumstances or in certain environments, it is more favorable to react quickly than to tally details for future reference. In fight-or-flight scenarios, an advantage of 50-100 ms can make the difference between life and death. In day to day life, altered reward processing of salient stimuli could lead to the discovery of novel perspectives, ideas and behaviors that are outside of our top-down models of the world (Cardo et al., 2010; Glover, 2011; J. Williams & Taylor, 2006). However, the ability to react quickly and the ability to observe and integrate relevant information are at odds in terms of our physiology and neural wiring. Therefore, an increased ability to react quickly in uncertain conditions would be expected to come at the cost of fluidly adapting to more predictable environments as in ADHD. Although additional methods would be required to confirm a presumptive error magnitude
encoding in ADHD, our results are consistent with a view of ADHD traits being associated with a neurophysiology that is more optimized to respond quickly in uncertain conditions than to respond appropriately in more predictable situations.

5.6 Conclusions

Our results demonstrate that adolescents with ADHD show reduced deactivation in midbrain dopamine targets when errors are detected. Reduced deactivation in midbrain dopamine targets when errors occur might cause problems with interrupting ongoing processing, and inducing appropriate plasticity conditions during adjustment processing that follows. Accordingly, ADHD subjects showed opposite activity compared to TD adolescents (activation instead of deactivation) in structures that modulate dopamine activity during adjustment processing. It appears that instead of errors serving to enhance the tuning of task-set related operations as in TD adolescents, errors instead invoke dysfunctional reinforcement activities, which could interrupt rather than enhance an established task-set.

The results from this study could not have been found using existing cognitive models which are restricted to the logic that process $X$ requires computations from region (or network of regions) $Y$. These models have not previously accommodated an interpretation of deactivations beyond the lack of process $X$. Allowing for a phase of error detection activity distinct from error magnitude activity revealed that error processing is indeed composed of multiple phases of activity, that these phases of
activity can be in opposite directions, and can involve functional deactivations that represent more than the simple lack of some process $X$.

Abnormal reinforcement processing that we observed in ADHD adolescents could prevent errors from leading to lasting changes in appropriate prefrontal networks. In contrast to opposite activity in reinforcement pathways that convey error magnitude but not task-related information, ADHD adolescents in this study exhibited normal activity in dorsal ACC, which is involved in recruiting prefrontal resources based on task-specific demands. When combined with our previous observations of opposite recruitment of these resources (i.e. preparatory deactivation instead of activation in IFG) compared to TD adolescents (Bhaijiwala et al., 2014), an overall picture emerges in which ADHD subjects are targeting appropriate brain networks, but experiencing abnormal reinforcement of these networks. This maladaptive reinforcement process might underlie deficits in executive functions like inhibitory control that are carried out by these networks, and the concomitantly reduced cortical thickness in these networks.

This work contributes to the growing literature suggesting a dysfunctional modulation of cortical plasticity in ADHD, and provides new perspectives on the underlying neural deficits. The ability to separate error detection from post-error slowing could benefit the investigation of developmental trajectories and the effects of therapeutic interventions in reinforcement learning pathways and task-related brain networks.
<table>
<thead>
<tr>
<th>Structure</th>
<th>BA</th>
<th>Location</th>
<th>BOLD signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>L precentral gyrus</td>
<td>4</td>
<td>-27 -26 66</td>
<td>-8.8</td>
</tr>
<tr>
<td>L postcentral gyrus</td>
<td>7</td>
<td>-14 -52 67</td>
<td>-6.2</td>
</tr>
<tr>
<td>R postcentral gyrus</td>
<td>1/3</td>
<td>36 -32 63</td>
<td>-7.3</td>
</tr>
<tr>
<td>R SFG</td>
<td>6</td>
<td>21 -8 69</td>
<td>-4.6</td>
</tr>
<tr>
<td>R SFG (dACC)</td>
<td>6</td>
<td>2 20 61</td>
<td>13.4</td>
</tr>
<tr>
<td>R SFG (dACC)</td>
<td>6</td>
<td>8 9 66</td>
<td>9.8</td>
</tr>
<tr>
<td>R SFG (dACC)</td>
<td>6</td>
<td>4 20 38</td>
<td>9.6</td>
</tr>
<tr>
<td>L ventral ACC (ACC1)</td>
<td></td>
<td>-18 -13 33</td>
<td>-1.8</td>
</tr>
<tr>
<td>L ventral ACC (ACC2)</td>
<td></td>
<td>-14 6 29</td>
<td>-2.0</td>
</tr>
<tr>
<td>L ventral ACC (ACC3)</td>
<td></td>
<td>-17 23 22</td>
<td>-2.1</td>
</tr>
<tr>
<td>R SFG</td>
<td>8</td>
<td>40 20 48</td>
<td>4.3</td>
</tr>
<tr>
<td>R SFG</td>
<td>10</td>
<td>30 49 27</td>
<td>8.9</td>
</tr>
<tr>
<td>L SFG</td>
<td>9</td>
<td>-26 50 29</td>
<td>5.6</td>
</tr>
<tr>
<td>R IFG/insula</td>
<td>45/13</td>
<td>53 15 1</td>
<td>6.9</td>
</tr>
<tr>
<td>L IFG/insula</td>
<td>47/13</td>
<td>-48 19 1</td>
<td>5.6</td>
</tr>
<tr>
<td>R medial frontal gyrus</td>
<td>6</td>
<td>3 32 33</td>
<td>8.4</td>
</tr>
<tr>
<td>R Precuneus</td>
<td>7</td>
<td>2 -73 50</td>
<td>14.4</td>
</tr>
<tr>
<td>R Precuneus</td>
<td>19</td>
<td>1 -79 39</td>
<td>14.4</td>
</tr>
<tr>
<td>R cuneus</td>
<td>19</td>
<td>1 -85 27</td>
<td>11.9</td>
</tr>
<tr>
<td>R lingual gyrus</td>
<td>18</td>
<td>1 -80 1</td>
<td>14.9</td>
</tr>
<tr>
<td>R fusiform gyrus</td>
<td>19</td>
<td>26 -82 -15</td>
<td>13.4</td>
</tr>
<tr>
<td>R inferior parietal</td>
<td>40</td>
<td>49 -48 48</td>
<td>7.3</td>
</tr>
<tr>
<td>L inferior parietal</td>
<td>40</td>
<td>-49 -49 43</td>
<td>4.8</td>
</tr>
<tr>
<td>R MTG</td>
<td></td>
<td>58 -25 -7</td>
<td>5.7</td>
</tr>
<tr>
<td>R IFG</td>
<td>47</td>
<td>26 10 -19</td>
<td>3.6</td>
</tr>
<tr>
<td>R dorsal striatum</td>
<td>-</td>
<td>20 12 0</td>
<td>-2.4</td>
</tr>
<tr>
<td>R thalamus, anterior</td>
<td>-</td>
<td>7 -10 18</td>
<td>3.2</td>
</tr>
<tr>
<td>L cerebellar tonsil</td>
<td>-</td>
<td>-24 -58 -33</td>
<td>-8.2</td>
</tr>
</tbody>
</table>

**Supplementary table 5.1.** Whole-brain significant activity estimates during error detection in TD adolescents. dACC, dorsal anterior cingulate cortex.
<table>
<thead>
<tr>
<th>Structure</th>
<th>BA</th>
<th>Location</th>
<th>BOLD signal</th>
</tr>
</thead>
<tbody>
<tr>
<td>L precentral gyrus</td>
<td>6</td>
<td>-46 -6 24</td>
<td>4.6</td>
</tr>
<tr>
<td>R precentral gyrus</td>
<td>4</td>
<td>35 -24 53</td>
<td>3.9</td>
</tr>
<tr>
<td>L MFG</td>
<td>6</td>
<td>-20 5 50</td>
<td>-4.9</td>
</tr>
<tr>
<td>R MFG/cingulate</td>
<td>24</td>
<td>13 -4 50</td>
<td>4.1</td>
</tr>
<tr>
<td>R insula</td>
<td>13</td>
<td>29 4 21</td>
<td>3.6</td>
</tr>
<tr>
<td>L caudate tail</td>
<td>-</td>
<td>-16 -32 23</td>
<td>-4.0</td>
</tr>
<tr>
<td>L SFG</td>
<td>9</td>
<td>-23 50 29</td>
<td>-6.5</td>
</tr>
<tr>
<td>R MTG</td>
<td>22</td>
<td>51 -41 5</td>
<td>4.1</td>
</tr>
<tr>
<td>R pHPC</td>
<td>30</td>
<td>17 -39 4</td>
<td>4.5</td>
</tr>
<tr>
<td>L insula</td>
<td>13</td>
<td>-40 -7 15</td>
<td>4.4</td>
</tr>
<tr>
<td>R IFG</td>
<td>47</td>
<td>49 30 -3</td>
<td>6.5</td>
</tr>
<tr>
<td>R STG</td>
<td>38</td>
<td>47 14 -13</td>
<td>7.8</td>
</tr>
<tr>
<td>L MTG</td>
<td>21</td>
<td>-60 -33 -10</td>
<td>-6.4</td>
</tr>
<tr>
<td>L STG</td>
<td>38</td>
<td>-42 6 -21</td>
<td>5.2</td>
</tr>
<tr>
<td>R culmen of cerebellum</td>
<td>-</td>
<td>36 -47 -23</td>
<td>7.8</td>
</tr>
<tr>
<td>R vermis of cerebellum</td>
<td>-</td>
<td>3 -65 -31</td>
<td>5.2</td>
</tr>
</tbody>
</table>

**Supplementary table 5.2.** Whole-brain significant activity estimates during post-error slowing in TD adolescents.
**Structure** | **BA** | **Location** | **BOLD signal**
--- | --- | --- | ---
L postcentral | 3 | -19 -32 65 | -5.7
L precentral | 4 | -15 -23 70 | -4.6
R precentral | 4 | 27 -23 61 | -3.7
L precuneus | 7 | -5 -55 61 | -6.0
R SPL | 7 | 28 -47 61 | -5.6
R precentral | 4 | 42 -13 57 | -6.5
L MFG | 8 | -25 24 44 | -4.5
L cingulate | 31 | -4 -27 39 | -2.8
L precuneus | 31 | -38 -75 38 | -6.9
L ventral putamen | - | -17 9 -1 | -2.7
L ventral striatum | - | -15 0 4 | -2.4
L genual ACC | 24 | -5 23 -2 | -2.7
R genual ACC | 24 | 4 29 1 | -4.5
Bi ACC | 32 | 0 51 -3 | -11.1
Bi medial frontal | 10 | 2 65 2 | -14.1
Bi medial frontal | 10/11 | 2 56 4 | -11.4
L pHPC | 36 | -38 -21 -16 | -4.5
R pHPC | 20/36 | 44 -22 -16 | -3.0
R fusiform gyrus | 37 | 42 -49 -18 | -5.1
R cerebellum | - | 34 -44 -29 | -4.1
R/Bi dACC/medial frontal | 8/32 | 5 19 47 | 7.11
R cuneus | 19 | 5 -78 36 | 12.0
R lingual gyrus | 17 | 11 -92 0 | 14.9
R IPL | 40 | 46 -47 41 | 5.3
R SPL | 7 | 29 -68 47 | 4.5
L supramarginal gyrus | 40 | -56 -50 28 | 6.6
R cingulate | 23 | 4 -24 31 | 3.9
R SFG | 10 | 30 51 25 | 11.4
L SFG/MFG | 10 | -26 60 19 | 12.1
R IFG | 44 | 51 15 19 | 6.0
L IFG | 44 | -51 14 13 | 6.5
R thalamus/caudate | - | 11 4 15 | 4.2
L HPC | 28 | -19 -29 9 | 5.0

**Supplementary table 5.3.** Whole-brain corrected activity estimates during error detection in ADHD adolescents. Although the number of deactivation loci were greater than the number of activation loci, the extent and intensity of activations were greater than that of deactivations. Deactivations were not present in ventral ACC of dorsal striatum, but were present in sensory-motor areas, ventral striatum and genual ACC and medial frontal cortex. SPL, superior parietal lobe; IPL, inferior parietal lobe.
<table>
<thead>
<tr>
<th>Structure</th>
<th>BA</th>
<th>Location</th>
<th>BOLD response</th>
</tr>
</thead>
<tbody>
<tr>
<td>R IFG</td>
<td>44/13</td>
<td>44 13 12</td>
<td>-6.4</td>
</tr>
<tr>
<td>L paracentral</td>
<td>6</td>
<td>-1 -29 71</td>
<td>14.2</td>
</tr>
<tr>
<td>L postcentral</td>
<td>3</td>
<td>-19 -32 66</td>
<td>10.3</td>
</tr>
<tr>
<td>L occipital</td>
<td>18</td>
<td>-33 -81 -8</td>
<td>18.0</td>
</tr>
<tr>
<td>Bi precuneus</td>
<td></td>
<td>1 -62 51</td>
<td>19.4</td>
</tr>
<tr>
<td>Bi lingual gyrus</td>
<td>18</td>
<td>3 -83 -8</td>
<td>25.3</td>
</tr>
<tr>
<td>R MFG</td>
<td>6</td>
<td>31 0 57</td>
<td>5.3</td>
</tr>
<tr>
<td>R medial frontal</td>
<td>6</td>
<td>2 2 49</td>
<td>6.4</td>
</tr>
<tr>
<td>R MFG</td>
<td>8</td>
<td>37 26 38</td>
<td>6.7</td>
</tr>
<tr>
<td>R MFG</td>
<td>9</td>
<td>49 20 32</td>
<td>7.4</td>
</tr>
<tr>
<td>L ACC</td>
<td>24/32</td>
<td>7 16 33</td>
<td>3.9</td>
</tr>
<tr>
<td>R ACC</td>
<td>32/24</td>
<td>2 10 32</td>
<td>4.6</td>
</tr>
<tr>
<td>R medial frontal</td>
<td>8/9</td>
<td>7 52 38</td>
<td>10.4</td>
</tr>
<tr>
<td>Bi medial frontal</td>
<td>9</td>
<td>1 48 21</td>
<td>9.0</td>
</tr>
<tr>
<td>Bi medial frontal</td>
<td>11</td>
<td>2 58 13</td>
<td>8.6</td>
</tr>
<tr>
<td>Bi thalamus, md</td>
<td></td>
<td>2 -15 19</td>
<td>7.8</td>
</tr>
<tr>
<td>L GPi</td>
<td></td>
<td>-18 -3 9</td>
<td>2.2</td>
</tr>
<tr>
<td>L thalamus, va</td>
<td></td>
<td>2 -2 6</td>
<td>4.1</td>
</tr>
<tr>
<td>R thalamus, va</td>
<td></td>
<td>5 -2 7</td>
<td>3.2</td>
</tr>
<tr>
<td>L IFG</td>
<td>47</td>
<td>33 23 -16</td>
<td>11.5</td>
</tr>
<tr>
<td>R IFG</td>
<td>47</td>
<td>38 16 -17</td>
<td>7.2</td>
</tr>
<tr>
<td>L STG</td>
<td>38</td>
<td>-36 10 -31</td>
<td>6.3</td>
</tr>
<tr>
<td>R STG</td>
<td>38</td>
<td>38 17 -29</td>
<td>7.4</td>
</tr>
<tr>
<td>Bi PCC</td>
<td>29</td>
<td>2 -41 9</td>
<td>10.9</td>
</tr>
<tr>
<td>L STG</td>
<td>41</td>
<td>-47 -31 8</td>
<td>3.9</td>
</tr>
<tr>
<td>L MTG</td>
<td>39</td>
<td>-47 -53 8</td>
<td>3.9</td>
</tr>
<tr>
<td>L STG</td>
<td>22</td>
<td>-61 -35 18</td>
<td>8.9</td>
</tr>
<tr>
<td>L insula</td>
<td>13</td>
<td>-38 -23 20</td>
<td>2.8</td>
</tr>
<tr>
<td>L pHPC</td>
<td>36/20</td>
<td>-39 -20 -15</td>
<td>9.4</td>
</tr>
<tr>
<td>R pHPC</td>
<td>28</td>
<td>22 -19 -20</td>
<td>6.4</td>
</tr>
<tr>
<td>R pHPC</td>
<td>36/20</td>
<td>37 -22 -22</td>
<td>7.4</td>
</tr>
<tr>
<td>R cerebellum, uvula/</td>
<td></td>
<td>5 -72 -35</td>
<td>9.3</td>
</tr>
<tr>
<td>Semi-lunar lobule</td>
<td></td>
<td>-12 -72 -39</td>
<td>9.7</td>
</tr>
<tr>
<td>L cerebellum, uvula/</td>
<td></td>
<td>-12 -72 -39</td>
<td>9.7</td>
</tr>
<tr>
<td>Semi-lunar lobule</td>
<td></td>
<td>-12 -72 -39</td>
<td>9.7</td>
</tr>
</tbody>
</table>

**Supplementary table 5.4.** Whole-brain corrected activities during post-error slowing in ADHD adolescents. The only deactivation present in ADHD adolescents during post-error slowing was in right IFG. This deactivation overlapped a region that was significantly activated in TD adolescents, indicating an opposite reinforcement of right IFG in these groups.
Supplementary figure 5.1. Error detection activities in a) HPC and b) midbrain (locations from Table 5.1) showing significant deactivation in TD but not ADHD adolescents. Error bars denote standard error.
**Supplementary figure 5.3.** Post-error slowing activity in a) ventral striatum and b) OFC (locations given in Table 5.1) showing deactivation in TD adolescents and activation in ADHD adolescents. Error bars denote standard error.
6 Conclusion

6.1 Summary of results

Inhibitory control and error processing are executive functions that are important to
goal-directed behavior, both of which are abnormal in ADHD. Previous studies have
imaged inhibitory control based on the predictions of various cognitive models. These
approaches have modeled inhibition as a single stage of activity. However, inhibition
involves multiple stages of preparation before stop signals appear, and the detection of
and adjustment to errors after failures of inhibition. The aim of this thesis was to present
methods for separating these phases of activity in healthy adults using fMRI, and to
apply these methods to the study of adolescents with and without ADHD.

6.1.1 Warning- and response-phase activities in the SST

The first study in this thesis involved the separation of two phases of preparatory activity
(warning-phase and response-phase) in the SST (Chevrier et al., 2015). Both of these
phases precede the appearance of stop signals, and both should contribute to the trade-
off between going and stopping on every trial. The novel approach successfully
separated warning-phase from response-phase activity, with warning-phases activating
posterior networks and response-phases activating fronto-posterior and subcortical
networks. Results showed that preparation to stop begins at the very onset of the earliest
warning stimuli and continues up to the point of response execution or inhibition.
Whole-brain correction for multiple comparisons revealed patterns of warning- and
response-phase activities that are consistent with previous observations using ERP but
which have not previously been dissociated with fMRI, namely: left followed by right
parietal activity, deactivation followed by activation of primary motor cortex, posterior
followed by fronto-posterior activity, and cortical followed by cortical-subcortical activity.

Posterior networks that activated during warning-phases are consistent with signaling the need for top-down control, and fronto-posterior and subcortical networks that activated during response-phases are consistent with the implementation of top-down control. Regression analyses showed that both of these phases contained activities that have distinct effects on the speed and accuracy of going and of stopping. The ability to separate previously hidden factors that influence inhibitory control during preparatory periods is of potential use in the development of models of preparation and inhibitory control, and the study of disorders marked by poor top-down control.

6.1.2 Error detection and post-error slowing

The second study involved separating two phases of error processing activity (error detection and post-error slowing) on failed stop trials. Error detection deactivated the midbrain and its principal dopaminergic targets, the dorsal striatum, ventral ACC and hippocampus, whereas post-error slowing deactivated the reciprocal pathway that modulates midbrain dopamine activity and is known to encode prediction error magnitude, namely the ventral striatum and caudal orbitofrontal cortex.

The current finding of error detection and post-error slowing activities in ascending and descending dopamine pathways supports a reinforcement learning perspective of stop task errors. Consistent with post-error adjustments being proportional to prediction error magnitude, post-error slowing deactivated structures that are known to encode
prediction error magnitude, but have previously only been isolated using reward and reinforcement learning tasks.

In addition to the current finding of activity related to error magnitude that has been observed in other tasks, the inclusion of an error detection process revealed patterns of activity in an ascending dopamine pathway, which have been suppressed in previous imaging studies that combined deactivations during error detection with activations during error magnitude processing. Separating error detection, which appears to set the conditions for subsequent adjustment-related processing, from adjustment-related processing itself, provides new insights into the neural mechanisms involved in reinforcement learning that would not have been accessible to approaches based on the predictions of any previous imaging model of reinforcement learning.

6.1.3 Error detection and post-error slowing in adolescent ADHD

The third study involved the separation of error detection and post-error slowing in fourteen typically developing (TD) adolescents and fourteen adolescents with ADHD. We observed the same deactivations during error detection and post-error slowing in TD adolescents as in healthy adults. By contrast, ADHD adolescents only partially deactivated the ascending dopamine pathway when errors were detected, and activated the reciprocal pathway that deactivated in healthy adults and TD adolescents during post-error slowing.

The application of this approach to adolescents with and without ADHD also provided insights into the nature of dysfunctional reinforcement signaling that have not been accessible to the predictions of cognitive models, where interpretations of activity
differences are limited to under-activation reflecting deficient function, and over-activation reflecting compensation for deficient function. The ability to achieve a whole-brain view of how task-related networks come into play and are reinforced based on feedback presented here provides a method of reaching beyond a phrenology that relates observed activity differences directly onto our cognitive model based descriptions of function and dysfunction. During error detection, the failure of ADHD adolescents to fully deactivate the ascending dopamine pathway could interfere with the ability to interrupt ongoing processing and behavior, and become susceptible to feedback-related information, potentially leading to problems with attention and the integration of new information. This shift in operating conditions from ongoing behavioral processing to error based adjustment is not part of conventional reinforcement learning models that do not address how mechanisms arise from their physical substrates, but is of primal importance in self-contained brains that have nothing but their physical substrate. However, accounting for neural activity that might accomplish this shift has not previously been incorporated into previous imaging models of error processing.

Intact functioning of the dorsal striatum is known to be necessary for the expression of newly learned behavior (Atallah et al., 2007). But once the new behavior has been expressed with an intact dorsal striatum, it is no longer needed for the new behavior. This might help explain the ability of ADHD subjects to sustain attention and learn tasks that they find enjoyable, whereas less enjoyable tasks are more resistant to attention and learning. The observations of reduced deactivation of dorsal striatum in ADHD adolescents during error detection, and decreased activation when adjustment processing was taking place, could suppress the expression of new behaviors based on the error.
Therefore, new insights from errors might be locked out of instrumental behaviors, including self-narratives, having psychological and social impacts.

During post-error slowing, an opposite pattern of activity in structures that modulate midbrain dopamine activity in ADHD subjects could result in dysfunctional reinforcement of task-related networks, and the failure of moments of insight to induce long-term changes in neural connectivity and function. Such a dysfunctional reinforcement process could help explain recent observations of opposite preparatory activity in task-related networks in ADHD adolescents, the delayed maturation of function and cortical thickness, and the limited effectiveness of behavioral interventions based on the known properties of healthy reinforcement learning systems.

6.2 Speculations on ADHD, dopamine and the suspension of disbelief

The results in this study show patterns of activity in dopamine pathways that are clearly different in adolescents with and without ADHD. However, definitive diagnostic criteria for ADHD remain elusive. Despite strong evidence of heritability, the study of genetics and neurotransmitter concentrations have not yet identified unambiguous differences between individuals diagnosed with ADHD and healthy controls (Del Campo et al., 2013). In this section I will touch on environmental factors during development that contribute to altered dopamine signaling, and explore a new perspective on the psychology of ADHD. Such a perspective may help provide a framework to better
reconcile social and psychological impacts with neurophysiological differences underlying executive function deficits.

The current results suggest an atypical impact of model-free reinforcement signals, but an intact system for model-based targeting of neural resources. In contrast to the diminished deactivation of ventral ACC in ADHD compared to TD adolescents during error detection, ADHD adolescents activated dorsal ACC similarly to TD adolescents. Ventral ACC is a direct target of midbrain dopamine neurons that convey reward prediction errors, whereas dorsal ACC activates according to an ideal Bayesian observer, and reflects both conscious insight into the ongoing task set as well as neural insight into the cortical resources required for that task set. We also observed the recruitment of appropriate networks in the wrong way (e.g. deactivation instead of activation) during both preparatory and adjustment processing. In combination, these findings suggest that the adolescents with ADHD scanned here might possess greater neural insight regarding the current task set and neural resources required than previously thought. However, the immediate reinforcement response that occurred in TD adolescents was opposite in ADHD. These signals are known to occur in the first couple hundred milliseconds after an error, which precedes any conscious deliberation.

The pre-conscious nature of reinforcement signals that are altered in ADHD could potentially be a frustrating experience; an integral part of the gestalt we experience when we make mistakes appears to be altered in ADHD. Such a disruption in immediate perception could lead to apparently distorted narratives, which could cause problems with social interaction, but which might not reflect the mechanisms of the disorder.
It is common for children with ADHD to account for inappropriate behavior with phrases like “I don’t know,” or “It wasn’t me”. Individuals with ADHD certainly have problems integrating new feedback-based information into future insight and planning. However, the connection between problems of narrative-based insight in ADHD and the underlying neurobiology might be outside of current formulations of insight-based learning. Current models of insight learning have addressed relationships between procedural and representational aspects of learning models that differentially affect performance (Öllinger, Jones, & Knoblich, 2008). Our results suggest that insight-based problems in ADHD could instead result from the interaction between task-related (i.e. model-based) and reinforcement (model-free) processing. Better understanding the connections between altered narratives and altered neurophysiology would undeniably improve social acceptance, as well as the study and treatment of ADHD.

There are anecdotal examples from other disorders with far more distorted narratives that are not treated as problems of insight because the neural underpinnings are better understood. For example, in hemispatial neglect that occurs after a significant percentage of right hemisphere strokes (Heilman & Valenstein, 1979; Parton, Malhotra, & Husain, 2004), patients can experience symptoms like not feeling that their left arm is their own. Rather than becoming upset and curious about their supposed foreign arm, these patients can make remarks such as “the nurse switched it with another patient” with little alarm. Another example is patients who either because of stroke or surgery have severed neural connections between regions involved in facial recognition (fusiform face area) and emotional response (amygdala), resulting in what is referred to as Capgras delusion (Christodoulou, Margariti, Kontaxakis, & Christodoulou, 2009).
These patients recognize the people they know, but do not get the emotional response they are accustomed to. This does not lead to curiosity about the difference in their perception, but can lead to the perception of others as imposter replacements, a perception that can be highly resistant to rational explanations (Ramachandran, 1998). Both of these examples show how a pre-consciously altered perception can lead to distorted narratives. Certainly these states have psychological impacts and effects on social interaction, but they are not studied or treated as fundamental problems of insight. In these cases, the disrupted neural circuits and course of treatment is relatively clear. Similarly with ADHD, the opposite pattern of reinforcement experienced with errors would occur too fast for conscious introspection, potentially causing complex interactions with narratives and social cognition. It would undeniably be of value to distinguish ADHD symptoms that are the direct result of biological mechanisms from downstream psychological consequences.

Given that genetic factors involving dopamine have not been identified that would explain the physiological differences observed here, what else might lead to the development of abnormal phasic dopamine responses on prediction errors in ADHD? Exposure to stress in early development is a risk factor for ADHD symptoms (Glover, 2011). It has been argued that hyperactivity and risk-taking behaviors could be adaptive in uncertain and hostile environments, could increase the ability to obtain mates, and promote novel hunting and foraging behaviors (Cardo et al., 2010; Glover, 2011; J. Williams & Taylor, 2006). Environmental factors, particularly those involving elevated or extended periods of stress at early stages of development, may contribute to altered dopamine signaling in ADHD.
An animal model for ADHD is the spontaneously hypertensive rat (SHR), which was originally developed for the study of hypertensive heart disease, but which also shows signs of ADHD (Miller et al., 2012; Sagvolden & Johansen, 2012). SHR dams exhibit ADHD symptoms as well as anxiety and low licking and grooming of their pups. A lack of contact and grooming at early stages of development signals to these pups that they will need to be more self-reliant in regulating their safety, temperature and nutritional state. This reduced exposure to nurturing behavior is associated with elevated stress and anxiety as well as epigenetic changes that have negative consequences on neural development and subsequent stress-sensitivity (Weaver et al., 2004). These factors combine to cause resources that are normally converted into growth to be diverted into stress, anxiety and immediate survival. The cross fostering of SHR pups with high licking and grooming dams is protective against ADHD symptoms (Masterpasqua, 2009). Therefore, it would appear that altered dopamine signaling in SHR rats is at least partly a learned response from early exposure to stress.

Stress followed by relaxation induces physiological responses that underlie placebo responses, but excessive stress leads to learned helplessness effects (Stefano, Fricchione, Slingsby, & Benson, 2001). In humans, these mechanisms are also involved in pleasure, addiction, attachment, sex and love (Esch & Stefano, 2004). Further, placebo-induced physiology and belief modulate dopamine release (Fricchione & Stefano, 2005; Stefano et al., 2001). Therefore dopamine constrains goal-directed behavior, but integrates visceral and hormonal states with our neural gating of attention and cognition based on our wants and beliefs.
In addition to its important roles in reward and reinforcement learning, dopamine is also necessary and sufficient for incentive motivational salience, or wanting (Berridge, 2007). In other words, dopamine is responsible for the saliency or “brightness” of our wants, our aversions and whatever constitutes important, sudden or surprising information or events. Dopamine is involved in using saliency to sustain attention and guide instrumental behaviors in the pursuit of goals. Therefore, the disruption of reinforcement mechanisms in ADHD might affect the way that we form desires, which sustain our attention and frame the context in which executive control takes place.

The protective influence of high licking and grooming dams on the development of ADHD symptoms in SHR pups can also be achieved with mechanical stimulation. Therefore, the preference for grooming-like sensations in rats could be an innate mechanism for the development of attention to important stimuli, similar to the innate preference for viewing faces over other similar stimuli in human babies. In addition to signaling safety, the development of attention to sensation without the need for reactive behavior would seem to be a primary component of the protection afforded by licking and grooming in rats.

Sensation precedes cognition in development. Before we learn about details, we first learn to pay attention to details. Exposure to sensation in safe conditions could provide the opportunity to turn the process of attending to important stimuli into a habit. Initial exposure to grooming in rats could serve to focus attention on sensory stimulation in the context of safety, allowing higher order neural processing to take place without engaging in reactive behavior. This early learning could induce appropriate autonomic
responses that allow for higher-order neural processing to take place, facilitating more optimal decisions.

Properly timed environmental exposure is crucial to the development of many higher faculties – if we are not exposed to speech at an early age we can never develop a capacity for language. Similarly, our capacity for attention might result from early sensory stimuli that are paired with cues indicating sufficient levels of safety and protection to suspend disbelief long enough for higher executive functions to develop – we learn to observe from exposure to being observed. The absence of such exposure at crucial periods of development could undermine the ability to attend to reward-predicting cues and long-term secondary reinforcers in favor of reacting as quickly as possible to potentially threatening stimuli. In stressful conditions that require high levels of vigilance and rapid fight or flight responses, fewer resources are committed and less time is available for executive functioning to take place (Glover, 2011).

The tendency to deliberate and the tendency to react quickly are at odds in terms of our neural wiring and physiology. Early exposure to stress could disrupt the ability to engage appropriate stress and relaxation responses to environmental and social cues. In rats, decreased anxiety is evident in increased exploratory behavior (Becker, Abraham, Kindler, Helmeke, & Braun, 2007). In humans, in addition to exploratory behavior, the ability to relax facilitates a state of suspended disbelief that is required as we construct our stories about ourselves and the world. As soon as a child begins talking it is evident that they do not perceive a world of cognitive constructs but of stories. Our stories are a necessary substrate on which cognition and learning take place. Even in adulthood our
relationships with others are not framed by cognitive principles as much as by our
stories.

Our ability to perceive the world in terms of stories depends on our innate capacity for
fantasy. Fantasy is not an end product of cognitive development but rather precedes and
frames all cognition and goal-directed behavior. In order to have goal-directed behavior,
it is first necessary to construct frameworks in which we are capable of desiring (Žižek,
1989). The process by which an object becomes imbued with some property that
transforms it into an object of desire, is not a property of the object but of the subject
(Kant, 1781), it “happens to us.” It is therefore likely that the development of desires and
cognitive mapping share neural mechanisms, involving dopamine neurons and the basal
ganglia, neither of which directly participate in conscious states. The significance of
networks that participate in conscious states is that these networks are all capable of
observing one another by establishing functional interactions and exchanging
information. However, striatal and dopaminergic networks that convey a sense of value
and agency do not participate in consciousness because they do engage in this kind of
functional connectivity, and cannot be “observed” by networks that do participate in
conscious states.

Fantasy and the suspension of disbelief are not purely individual matters either, but are
socially-charged and conveyed on us by group ideologies, which are essentially
ensembles of the fantasies of the group. Fantasy and ideology are ironically the most
concrete level of human interaction. It is undeniable that the most important ecosystem
for human survival is the social environment. Therefore, it is of primal importance to
balance our inner narratives with the appropriate ideology in our social interactions. It appears that we accomplish this balance with the same reward-based learning circuitry that allows animals to learn from and adapt to their environment.

It might appear that speculations on fantasy and ideology are far removed from the study of neural activity in individuals. However, it warrants consideration for two unassailable reasons. Firstly, the brain and its reward systems have evolved to learn from and adapt to the environment. In humans, social ideology is undeniably the most important landscape that must be navigated in order to maximize reward and the propagation of genes. Secondly, almost every part of the brain, and every corresponding perception, decision and behavior is modulated by reward context. Although reward is often operationalized in terms of choice preferences, reaction times and orienting responses, these measures hardly cover the scope of our internal wants and fears. In psychoanalytic terms, the ubiquity of the presence of reward-based modulation at every level of perception, cognition and behavior, results from the fact we perceive the disembodied ‘other’ of our fantasies gazing upon us. No perception, or instrumental behavior, can escape the gaze of the object of our ideology. The effects of the perception of the gaze of the other upon us are present in every measurable behavior, and in any measurement of neural activity.

This is not to imply that our neural mechanisms have metaphysical causes that originate in an actual disembodied other, but that the seemingly metaphysical other is the consequence of real neural mechanisms that have evolved to perform concrete tasks. The evolution of neural systems that help us pay attention and scrutinize our perceptions
appears to have reached a point where we could no longer suppress the scrutiny of the origin of our own wants and perception of agency in the world at large. This reflective level of perceptual and cognitive scrutiny, combined with the neural mechanisms that have evolved to generate frameworks in order to guide goal-directed behavior, led to the psychological construct of the "other." The ideological other provides evolutionary advantages by maintaining vigilance and reinforcement amidst highly demanding tasks in the pursuit of complex rewards.

The emergence of the perception of the other has in turn led to the emergence of the psychological construct of an “objective observer.” The concept of an absolute frame of reference is itself a fantasy, as it is not attainable for a real living being. Regardless, it is exactly this fantasy that is necessary for the highest and most sober of human pursuits. It is because we have a concept of an objective observer that humans are able to abstract, do secondary theory of mind, generate stories, contemplate philosophy, develop societies, do mathematics and discover and apply scientific methods. The theoretical, absolute frame of perceptual reference afforded by the concept of an objective observer has also given rise to uniquely human constructs of the self. This unique sense of self is entangled with our perceived relation to the other.

For humans, our relationship with the other is central to what we would perceive to be our mental health, which requires perceiving the other at a healthy distance. Mental health is not marked by lack of fantasy, but by not taking ourselves and the events in our lives too seriously. Various factors in human life can potentially sensitise the normally
healthy distance and quality of relationship we maintain with the fantasies that arise out of our perception of the other.

The preparatory and reinforcement activities identified in the studies presented here provide crude snapshots of moments when circuits that don’t participate in consciousness, impart a sense of value (dopamine) and agency (striatum) on circuits that do (neocortex). The development of the perception of the other would arise from the accumulation of these kinds of moments, such as the orientation of attention to stimuli and the automatic and preconscious distribution of value we perceive in them, such as during the preparatory activities identified in Chapter 3; and during the presentation of feedback and reward, such as during error detection and post-error adjustment in Chapter 4. These activities are demonstrably different in the ADHD adolescents studied in Chapter 5. Therefore, one must wonder what the activity differences observed here could tell us about how the other of ideology might differ in ADHD: how might these physiological correspond to more subjective psychological differences?

The psychological construct of the other, unlike any real being, is of something that sees and affects everything but cannot be seen or affected itself, which appears to exist both inside ourselves and in the world at large. The history of one’s perception of the other must have strong psychological impacts on the trust that is required to engage in any kind of framework. This sense of trust is necessary for any person to willingly enter into any behavior, contract or relationship. The results presented here indicate that these ADHD adolescents possessed insight regarding the nature of the error and requisite adjustments, indicated by normal activity in dorsal ACC and the targeting of appropriate
cortical networks. However, the subcortical reinforcement of these insights that serves to automatically improve behavior in their peers does not seem to have the same effect in adolescents diagnosed with ADHD. Psychologically, such a history of reinforcement would seem to inevitably generate an implicit sense of distrust in the other, the all-pervasive agent that conveys order and value on the world at large. This sense of distrust could result in the failure to suspend disbelief and “get on with” instrumental tasks that is commonly observed in ADHD.

The biological role of relaxation in the suspension of disbelief is to allow time for executive processing to take place without engaging reactive behavior, affording the pursuit of more complex rewards. Conversely, the psychological role of our narratives and fantasies, in addition to framing and contextualizing the known, is to avert stress and anxiety by obstructing our view of the unknown – it is very stressful and anxiety-inducing to focus on how much we do not understand. The prevalence of smoking and its association with social detachment in adolescent ADHD (Groenman et al., 2013; Lichtenstein et al., 2012) could reflect the need to reduce anxiety by engaging in an instrumental behavior (with effects on acetylcholine receptors, dopamine and noradrenaline) in the absence of a normally protective narrative. Further, the benefits of exercise in ADHD subjects (Archer & Kostrzewa, 2012; Grassmann, Alves, Santos-Galdróz, & Galdróz, 2014; Wigal, Emmerson, Gehricke, & Galassetti, 2012) could be related to the benefits of exercise on stress relaxation mechanisms (Esch & Stefano, 2010), which might facilitate more appropriate visceral responses to social stimuli and the establishment of more stable and adaptive narratives.
It would seem that early learning of stress relaxation responses are involved in something like the suspension of disbelief that occurs when children take on their earliest stories about themselves and the world. A similar kind of suspension of disbelief is also required to take on any instrumental conditioning, including instruction in a school environment. The proto-hypothesis here is that many factors, such as early exposure to elevated stress, and their effects on dopamine release, might serve to alter the threshold of this kind of suspension of disbelief required to accept premises and motivate the learning and performance of corresponding instrumental behaviors.

An altered threshold for reward processing in general is apparent in ADHD; individuals with ADHD appear to activate reward systems and adjust to feedback information more normally when incentive motivation is increased by providing monetary—instead of point-based rewards (Umemoto et al., 2014). Therefore, rather than just having difficulties accomplishing their goals, individuals with ADHD might have deeper difficulties constructing the frameworks in which their desires can be formulated and expressed.

Suspension of disbelief is a literary concept that has historically arisen at times of increased skepticism about supernatural phenomena, when writers of fiction like Samuel Coleridge and Horace were trying to forge new fantasies for disenchanted cultures. A similar but more modern concept is that of “acting under a description” developed by Elizabeth Anscombe (Anscombe, 1957; Müller, 2001). Acting under a description is a conception of human intentionality that addresses some of the distance between measurable behaviors and intentional actions. In short, when asked what we are doing,
we do not describe our motoric acts like “I am pushing my arms upwards,” but our intentions like “I am opening the window.” To the extent that we are volitional beings, we can only understand ourselves by understanding the descriptions under which we act. This is the level at which we directly experience mental health and quality of life.

We are not defined by the series of events we have experienced as much as the interpretations under which we have acted. Conversely, our descriptions of ourselves depend on the descriptions available to our society, which change across cultures and over the course of history. Ian Hacking extended the work on acting under a description (Hacking, 1998), using multiple personality disorder (MPD) as an example. Before widespread media attention to MPD, there were only ten documented cases in the preceding 50 years, but in the following decade over six thousand cases were believed to be diagnosed. The appearance of a new way of describing oneself and one’s relation to past experiences provided by MPD caused an explosion of incidence, which has been referred to as a “semantic contagion.” The goal of this example was not to question whether MPD was real, but to question our notion of reality in the tradition of Wittgenstein, by asking “is MPD a real what?” (Hacking, 1998) The issue here is not whether ADHD diagnosis and prevalence are a semantic contagion, but whether ADHD might be characterized by differences in the ability to smoothly direct the descriptions under which we act, which might underlie observed deficits of executive function. It is difficult to pursue a goal if the stories that frame goal-directed behavior are not stable.

It is possible that differences in dopamine signaling in ADHD reflect a different threshold for acting under a description. This is not to imply that individuals with
ADHD do not understand themselves, but it does suggest that the search for descriptions under which to act does not diminish with age as in people without ADHD. As with other age-inappropriate behaviors, the search for defining the self seems to persist further into adulthood, as is apparent from a recent paper on ADHD adults entitled “Do I need to become someone else?” (Schrevel, Dedding, van Aken, & Broerse, 2014) This study determined that adults with ADHD found problems directly associated with symptoms to be less distressing than problems of low self-image, feelings of lack of acceptance and being under-valued. It is possible that dysfunctional reward processing could cause problems with responding to normal levels of praise and criticism. Conversely, altered reward processing could prevent the kind of feeling required to even enter into frameworks in which we act under a collective description, which could be problematic in any micro-culture such as a place of work.

It can be argued that certain forms of social instability associated with ADHD, such as in relationships or employment, can also be adaptive (Cardo et al., 2010). If the descriptions under which one is acting are unpleasant or unfulfilling then it can be highly adaptive for an individual to be less committed to them. Further, every ideology has problems, and societal fitness requires individuals whose views of the world are less constrained by the ideologies around them. Social construct theories argue that ADHD diagnosis is an arbitrary and socially constructed threshold to describe behaviors outside of the social norm (Parens & Johnston, 2009). Regardless of the existence of ADHD as an “objective disease,” the fact remains that individuals diagnosed with ADHD are at increased risk of difficulties integrating into society in ways that are detrimental to their
quality of life, which warrant efforts to better understand and provide therapeutic interventions when problems become unmanageable.

The study of the development of wants has historically been the domain of advertising, market research and propaganda. Market research is defined by the manufacturing of wants, engaging the suspension of disbelief and providing descriptions under which to act in accordance with a given mandate. This mercantile relationship with our wants is not restricted to industry, but is a way that we describe ourselves. When we get the feeling that we should not go along with any plan or idea, we “don’t buy into it.” This is the relationship between ‘credit’ and ‘credibility,’ both of which refer back to incentive motivational salience. The proposal here is that altered reward signaling in ADHD could alter both the physiological thresholds for “buying into” various descriptions under which we can collectively act. This physiological interpretation could also help understand subjective psychological states in ADHD.

In addition to studying how to achieve our goals by focusing on deficient cognitive processing, we could consider how to gain more insight into our narratives and ideologies, and construct better descriptions under which we agree to act. Agreeing to the description under which we are acting is the biggest determinant of the mental and physical resources we will happily commit. A better understanding of the mechanisms by which we develop wants and agree to act under descriptions could open a wide array of new descriptions under which to act.

In the case of ADHD, understanding the underlying mechanisms associated with behavioral differences displaces stigma and increases social acceptance. The idea
proposed here is that executive control deficits in ADHD might be the downstream result of altered stress responses that are learned at early stages of development. Stress relaxation mechanisms are involved in the interaction between belief and the integration of cognitive and visceral states, and in the development and expression of wants. These mechanisms help frame our desires the same way that our desires frame goal-directed behaviors that require executive control. The kind of dopamine signaling that appears to be disrupted in ADHD would seem to be an important hub through which desires, based on the descriptions under which we believe we are acting, frame the context in which executive function integrates with physiology. A more comprehensive understanding of the connections between biological and psychological factors could benefit our understanding of and potential therapeutic interventions for ADHD and other neuropsychiatric disorders.

6.3 Future work

The work presented here provides some proof of principle with regard to separating arbitrarily brief within-trial activities in the SST. It is possible to separate within-trial activities as long as they do not always co-occur, unfold in a known temporal sequence when they do co-occur, and trials are presented with the appropriate temporal jittering of inter-trial intervals (Ollinger et al., 2001). In addition to confirming the validity of these constraints on separating within-trial activities, we also confirmed that the angles between within-trial regressors (e.g. between warning- and response-phase regressors in Chapter 3, and between error detection and post-error slowing regressors in Chapters 4
and 5) were substantially greater than 20° in order to confirm that issues of multicollinearity would not preclude their separation (O’Brien, 2007). The patterns of activity during preparatory periods and error processing revealed by the current approach were highly consistent with what would be expected from single cell recording and ERP approaches with higher temporal resolution, but lower spatial sensitivity. The novel whole-brain view of these processes provides new insights into preparatory and reinforcement processing in healthy adults and adolescents, and into the differences between adolescents with and without ADHD. Therefore, the methods described here can be extended with further analyses, and promise further novel insights if adapted for the study of other tasks, disorders, and drug manipulations.

Firstly, a follow-up study could perform correlation analyses on the fMRI results from ADHD adolescents presented here in order to examine whether the level of deactivation in dorsal striatum and ventral ACC predict differences in behavior (speed of go responses, stop accuracy, SSRT and post-error slowing), diagnostic measures (hyperactivity, inattention) or post-error slowing activities in ventral striatum and OFC. Similar analyses could determine whether activity in ventral striatum and OFC during post-error slowing are related to behavioral or diagnostic measures, or affect the level of preparatory activity in task-related networks. Further, given the opposite patterns of activity in ADHD compared to TD adolescents during preparation and post-error slowing, a discriminant analysis could be performed to determine whether preparatory and error related activities predict diagnostic status.
Previous studies have reported normalized brain activation in IFG, ACC and in reward and reinforcement learning pathways in ADHD subjects on stimulant medication (An et al., 2013; Rubia, Halari, Mohammad, et al., 2011; Rubia, Halari, Cubillo, et al., 2011; Rubia, Halari, et al., 2009; Shafritz et al., 2004). However, these claims are often the result of medication inducing a lack of significant difference of activity in a narrow ROI. Whole brain separation of the within-trial activities that combine to form these net differences would give a better view of the effects that stimulant medications have on the brain. More specifically, the methods developed here could determine whether MPH normalizes preparatory activity in IFG or increases the level of activity during stop-phases, whether it normalizes the level of deactivation in dorsal striatum and ventral ACC on error detection, and whether it causes deactivation instead of activation of ventral striatum and OFC during post-error slowing.

The methods described here can be used to study a wide array of neuropsychiatric disorders marked by deficient top-down control and/or abnormal dopamine function, such as obsessive compulsive disorder (Woolley et al., 2008), traumatic brain injury (Konrad, Gauggel, Manz, & Scholl, 2000; Sinopoli, Schachar, & Dennis, 2011), Parkinson’s disease (Cameron et al., 2012; Praamstra et al., 1996) and schizophrenia (Braver & Cohen, 1999). Many of these disorders have similar inhibitory control deficits that are not distinguishable based on behavioral data or from previous imaging approaches. Given that several important brain structures exhibit both increases and decreases in activity in a single trial, separating these phases of activity with fMRI promises to identify new subtle differences between disorders with apparently similar behavioral deficits. Further, most of these disorders are believed to be downstream
results of long term plasticity processes involving the actions of neuromodulatory networks like dopamine at moments like prediction errors. Separating error detection from post-error adjustment (error magnitude) activities in ascending and descending dopamine pathways promises unique insights into the mechanisms involved in various neurospysciatric disorders.

Separating within-trial activities could also help identify component processes specific to different strategies in selective inhibition tasks, which include ignore trials that resemble stop trials, but which should not be inhibited (Bissett & Logan, 2014). It has been postulated that different strategies can be used when performing selective stop tasks, and that these strategies can be distinguished based on their response time distributions. Separating preparatory from inhibitory phases of activity could help determine whether differences in reaction time distributions correspond to distinct neural signatures that reflect real differences in strategy.

Separating within-trial activities could offer improvements to the imaging of other response tasks that involve brief preparatory periods such as Stroop and switch tasks (Carter & van Veen, 2007; Ruge, Jamadar, Zimmermann, & Karayanidis, 2013), working memory tasks (D’Esposito, Postle, Ballard, & Lease, 1999), tasks of emotional processing (Fossati et al., 2003), or instrumental tasks with high error rates. Although a great deal has been done with parametric manipulations of factors that influence behavior and neural activity during the performance of these tasks, less attention has been paid to obtaining whole-brain views of rapidly unfolding component processes. The results in this study show that it is important to separate these phases of activity.
where possible to correctly identify the networks involved in various stages of processing. These methods could also be applied to classic reward tasks, which have not previously been imaged with models that allow for an error detection process that is distinct from error magnitude. This would allow for comparing similarities and differences between prediction errors on reward and instrumental learning tasks. The ability to separate error detection from other aspects of prediction errors could also aid in the development of new models of how reinforcement learning takes place in the brain.

Separating within-trial activities with fMRI could be aided by the addition of ERP approaches using electroencephalography (EEG) and magnetoencephalography (MEG). Preparatory activities that predicted individual differences in the speed and accuracy of going and stopping could be investigated with greater temporal precision using adaptive spatial filter (beamformer) techniques that can estimate activity at arbitrary locations (Quraan & Cheyne, 2010). Similarly, using a beamformer approach in the putative dopamine networks identified here during performance of a classic reward task could provide new insights into reinforcement learning. For example, such an approach could be used to test current postulates that positive prediction errors are encoded in the intensity of dopamine activity, whereas negative prediction errors are encoded by the duration of their suppression of activity (Bayer & Glimcher, 2005).
7 References

Ackermann, H., Daum, I., Schugens, M. M., & Grodd, W. Impaired procedural learning after damage to the left supplementary motor area (SMA)., 60 Journal of neurology, neurosurgery, and psychiatry 94–97 (1996). doi:10.1136/jnnp.61.5.555


Band, G. P., & van Boxtel, G. J. (1999). Inhibitory motor control in stop paradigms:


Bellgrove, M. A., Hawi, Z., Gill, M., & Robertson, I. H. (2006). The cognitive genetics of attention deficit hyperactivity disorder (ADHD): sustained attention as a


doi:10.1093/cercor/bhu050


motor preparation and stimulus anticipation in a time estimation task. 

doi:10.1016/0013-4694(88)90132-0


doi:10.1093/cercor/bhq239


doi:20/22/8443 [pii]


doi:10.1016/j.neuroimage.2012.01.057

Cardo, E., Nevot, A., Redondo, M., Melero, A., de Azua, B., García-De la Banda, G., &


ascending dopaminergic systems of the midbrain: A moveable electrode mapping study. *Brain Research, 185*(1), 1–15. doi:10.1016/0006-8993(80)90666-6


doi:10.5114/aoms.2010.14269


188


Ollinger, J. M., Corbetta, M., & Shulman, G. L. (2001). *Separating processes within a*
doi:10.1006/nimg.2000.0710

doi:10.1027/1618-3169.55.4.269


doi:10.1146/annurev.neuro.25.112701.142937


needs of adults with ADHD. *Health Expectations*, n/a–n/a. doi:10.1111/hex.12328


doi:10.1152/jn.00238.2004

doi:10.1097/01.chi.0000140452.51205.8d


doi:10.1111/j.1469-7610.2007.01851.x


Wigal, S. B. (2009). Efficacy and safety limitations of attention-deficit hyperactivity
disorder pharmacotherapy in children and adults. *CNS Drugs.*
doi:10.2165/00023210-200923000-00004

doi:10.1177/1087054712454192


Woolley, J., Heyman, I., Brammer, M., Frampton, I., McGuire, P. K., & Rubia, K.


