Action control in uncertain environments

Peter Smittenaar

A dissertation submitted in partial fulfilment of the

requirements for the degree of Doctor of Philosophy

Wellcome Trust Centre for Neuroimaging

Institute of Neurology

UCL

February 2015

I, Peter Smittenaar, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Peter Smittenaar

Abstract

A long-standing dichotomy in neuroscience pits automatic or reflexive drivers of behaviour against deliberate or reflective processes. In this thesis I explore how this concept applies to two stages of action control: decision-making and response inhibition.

The first part of this thesis examines the decision-making process itself during which actions need to be selected that maximise rewards. Decisions arise through influences from model-free stimulus-response associations as well as model-based, goal-directed thought. Using a task that quantifies their respective contributions, I describe three studies that manipulate the balance of control between these two systems. I find that a pharmacological manipulation with levodopa increases model-based control without affecting model-free function; disruption of dorsolateral prefrontal cortex via magnetic stimulation disrupts model-based control; and direct current stimulation to the same prefrontal region has no effect on decision-making. I then examine how the intricate anatomy of frontostriatal circuits subserves reinforcement learning using functional, structural and diffusion magnetic resonance imaging (MRI).

A second stage of action control discussed in this thesis is post-decision monitoring and adjustment of action. Specifically, I develop a response inhibition task that dissociates reactive, bottom-up inhibitory control from proactive, top-down forms of inhibition. Using functional MRI I show that, unlike the strong neural segregation in decision-making systems, neural mechanisms of reactive and proactive response inhibition overlap to a great extent in their frontostriatal circuitry. This leads to the hypothesis that neural decline, for

example in the context of ageing, might affect reactive and proactive control similarly. I test this in a large population study administered through a smartphone app. This shows that, against my prediction, reactive control reliably declines with age but proactive control shows no such decline. Furthermore, in line with data on gender differences in age-related neural degradation, reactive control in men declines faster with age than that of women.

Table of contents

Acknowledgements

I’ve learned more than I can remember during my time at the FIL, and I have many people to thank for that. Ray has supported me throughout and I’ve tried my best to learn from him what I could about all things science and life. Thank you for the Berlin trips and boat rides and, above all, the advice and encouragement—especially at the end when it was most useful. Klaus Wunderlich supervised me through my rotation project, which was heading for failure until he steered it straight. Marc Guitart-Masip taught me the ways of SPM and was more than generous with his time. Nick Wright came up with the idea of using TMS to study model-based control—he has since lamented not doing it himself. Analysing the 2-step data greatly benefitted from Tom FitzGerald’s expertise in all things computational. Feeling I hadn’t exploited the 2-step task quite enough yet, George Prichard and Joern Diedrichsen lent me their expertise in tDCS to do a third (and final!) study on this task. My basal ganglia imaging project was inspired by chatting to many people including Robb Rutledge, Laurence Hunt, Peter Dayan, Tim Behrens, Joel Winston and Zeb Kurth-Nelson—thank you all for your wisdom. Siawoosh Mohammadi and Martina Callaghan were instrumental in getting the impossibly complicated sequences I wished for up and running. For the final chapter about response inhibition I thank the Great Brain Experiment team: I remember us laughing at the notion of ever going over 1k users! Ric Davis provided flexible IT support and Glyn Lewis explained big data statistics when an ANOVA wouldn’t do anymore. During my MSc in Nijmegen Roshan Cools was the first to introduce me to dopamine and decision-making—her enthusiasm and advice made me pursue it further. All the work in this thesis depended on the outstanding support

from the Dolan lab; Jan Glensman and her imaging team; Ric Davis and his IT crowd; David Bradbury, Peter Aston and all the others in admin and technical support; Karl Friston, Guillaume Flandin and other staff who for decades now have patiently explained SPM and neuroimaging to people like myself; and Marina Anderson who helped me navigate the bureaucracy and with my gardening. The Wellcome Trust generously funded these studies.

The past years haven't all been work, thankfully. Weekly squash with first Antoine Lutti and then Marcos Economides and Robb kept my ego in check after playing table tennis with Zeb and Laurence. Marijn Martens was always close on chat and provided a laugh and a place to sleep every time I went back to Nijmegen. I've met many people first through dodgeball, then through frisbee and board games—I hope to keep playing for many more years with all of you. There are too many other colleagues-turned-friends to thank here, but I should mention Helen, Eleanor "Asian Mayhem" Loh, Raphael, Mona, Archy, James, Spas, James, Peter Z, Giles, Suz, George, Ash, Marc, Matej, Laura, Erman and Elisabeth.

I also thank my family, Els, Pieter and Richard, who have supported me throughout. Without Richard leading the way I don't think I would have considered a PhD in the UK. I'm happy you opted for London over Singapore—it's been great having you around.

Dearest Rebecca, I'm going to take the high road and not joke about all your oddities. All I can say is that the best thing about this PhD is that it brought me you.

1 Introduction

1.1 Conceptual overview

This thesis addresses how the human brain supports reward learning, decision-making and action control. It builds on a large body of work describing behavioural models of action control and the functional neuroanatomy of decision-making when rewards and punishments are at stake. In this chapter I will provide an outline of the work, define the key terms that will be used throughout this thesis and present an overview of the chapters.

Human decisions are shaped by countless factors. We often deliberate on our choices, agonizing over the possible consequences of our actions by weighing the risks against the gains. This process plays out over seconds, minutes or many days. Thankfully we do not need to invest such mental effort for each action we take, as the limited capacity of our brain suggests we would grind to a halt just making breakfast in the morning. Instead we automate many of our decisions, making it unnecessary or even impossible for deliberate thought to intervene. A prominent example is addiction, whereby a once deliberate choice to take drugs, try gambling or go shopping becomes so engrained in our decision-making machinery that no amount of consideration of likely negative consequences prevent these maladaptive behaviours from expressing themselves.

The first aim of the work in this thesis is to better understand what neural factors determine the extent to which humans use more complex or simple strategies in our decisions. I will explore how to directly manipulate the use of these strategies through dopamine and brain stimulation. Using neuroimaging I will ask how these value-based decisions are implemented in subcortical structures such as the basal ganglia. The second aim is to understand how, after a decision has been made, we exert rapid self-control to alter these decisions in response to changing circumstances. Most of this work focuses on the role of preparation in the execution of rapid self-control.

1.2 Definitions

1.2.1 Rewards, values, models and decisions

Many terms in the field of learning and decision-making have intuitive meanings; nevertheless we should define them more precisely. A fundamental concept is that of *reward*, which is operationalised as the 'intrinsic desirability of a state' (Sutton and Barto, 1998). More broadly it is whatever an organism tries to maximise over the long run, and can be further classified: unconditioned reinforcers are desirable in and of themselves possible by virtue of the engine of evolution, manifest in the desirability of water, food and sex; conditioned reinforcers are desirable only by virtue of their association with other reinforcers, as in the case of money which can buy all three rewards mentioned above. The maximization of reward is achieved by calculating *values* at each decision point. The value function describes, for each available action or state, how much reward it will yield in the long run. For example, the immediate reward of being in an airport might be considered low, but the value of that same state is high if airports predict holidays and conferences in the near future. Typically, we try to understand an organism's value function by examining how it is expressed in choice.

The field of reinforcement learning is, to a large extent, concerned with efficient ways of calculating values. Two such ways, model-based and model-free, play a prominent role throughout this thesis, and are discussed in-depth in section 2.2. It is worth briefly discussing the notion of a model: it refers to any representation that mimics the behaviour of the environment, for example a set of rooms in a building and the way they are connected. These models can be used for planning and calculating the value function on-line. This differs critically from what has been defined as model-free algorithms, which lack such a model of the environment and use more primitive methods of approximating value.

At this point we have discussed rewards, values and models. What are 'decisions' in this framework? Decisions are often taken to involve some conscious, deliberative effort by the organism. But decades of psychological research has shown that many actions are reflexive, model-free or habitual, i.e. driven without any deliberation. Here I consider any action derived from a value function to be a decision, therefore including both model-free and model-based actions. Although the definition of the terms as presented here comes from the field of reinforcement learning, these notions pervade psychology, cognitive neuroscience and economics.

1.2.2 Self-control

We can think of decision-making as a fallible process that needs both time and, now and then, post-decision adjustment. A failure to do so leads to what is varyingly called impulsivity or a lack of self-control. These are multifactorial concepts (Evenden, 1999), though always defined in the context of poor actions leading to undesirable outcomes. In chapters 9 and 10 I will specifically consider post-decision inhibitory self-control, or more plainly, the ability to prevent an action as it is about to be executed. Although this type of self-control should be considered distinct from self-control at the time of choice, its impairment can be observed just the same in, for example, addiction (Ersche et al., 2012). One promising avenue in the study of self-control is how we *prepare* for situations that will challenge our ability to inhibit our actions. For example, a recovering drug addict might resolve not to approach a dealer on the street, rather than rely on their immediate ability to stop themselves in case the situation arises. Such *proactive control*, then, is intimately linked with goal-directed choice, and is similarly thought to rely on working memory, maintenance of future goals and top-down control originating in frontal cortex (Aron, 2011; Braver, 2012; Schall and Godlove, 2012).

1.3 Outline of thesis

I will start by reviewing the literature on reinforcement learning and the central role it occupies in the psychology and neuroscience of reward learning and decision-making. In particular, I will discuss recent advances in understanding how multiple reinforcement learning systems in the brain trade off and compete with one another. This will be followed by an overview of the literature on inhibitory self-control, focusing on notions of proactive and selective inhibition and their neural correlates. Chapter 3 provides a background on the methodology used in the subsequent chapters.

The empirical work in this thesis is divided into two parts. Chapters 5, 6 and 7 present work on model-free and model-based reinforcement learning. In chapter 5 I employed a systemic manipulation of dopamine levels. This is known to affect both model-free and model-based control separately, but I provide novel insights into its effects when both types of control are allowed to compete. In an effort to pin down the anatomy of this trade-off, chapters 6 and 7 examine prefrontal roles in reinforcement learning by applying a transient functional lesion or supposed gain-of-function through neurostimulation, respectively. Together, these three chapters provide novel insights into direct alterations of decision-making strategies. I then ask how action values and rewards from reinforcement learning models are represented in the anatomy of the basal ganglia and its recurrent loops with the cortex, using a combination of structural, diffusion-weighted and functional imaging.

The second half of the empirical work centres on a paradigm for investigating the role of preparation in selective inhibitory control of action. Chapter 9 presents a novel characterization of behaviour on this task, before examining how preparation is implemented in neural structures known to be involved in outright inhibition. In brief, the task allows simultaneous measurement of the speed and selectivity of inhibition, and I ask how this trade-off is reflected in neural structures. Chapter 10 then applies this same paradigm on a much larger scale by means of a smartphone experiment. This allowed us to map the demographics of proactive self-control.

Finally, the discussion (chapter 11) I will discuss the implications of this work, drawing together insights from the chapters to examine the link between decision-making and self-control.

2 Literature review

2.1 Overview

The work described in this thesis encompasses learning, decision-making and action control. In this chapter I start with an overview of reinforcement learning, describing its history in animal learning as well as in artificial intelligence, the underlying algorithms and its neural implementations. A distinction will be made between different solutions to the problem of reward maximization, and how an organism might arbitrate between distinct strategies.

In a continually changing environment adaptive behaviour does not end with a value-based decision. Examining adjustments to ongoing actions provides a window into prefrontal and subcortical control mechanisms that are, fundamentally, rapid decision-making systems. I will review the concepts and models underlying the field of inhibitory self-control in section 2.3, and more recent work in the field examining how preparation and expectation shape self-control. I end each section by explaining how outstanding questions in the field are addressed by the work that makes up the core of this thesis.

2.2 Reinforcement learning

2.2.1 Multiple solutions to the same problem

If the long-term goal is survival and reproduction, the apparently trivial decisions we make throughout the day are what determine success. Understanding the building blocks of such adaptive behaviour in a complex and uncertain environment can be guided by models from artificial intelligence, decision frameworks in economics, and heuristics, biases and cognitive strategies in psychology. As we shall see in this review of the literature, the class of models I focus on reside in reinforcement learning. Its algorithms not only accurately describe value-based learning and choice in humans; but also provide suggestions for their efficient implementation in neural systems.

Dual process theories, based on the notion that a problem can be solved in multiple ways, are ubiquitous in psychology and artificial intelligence. Within decision-making, the first of two such processes has been called System 1 (Kahneman, 2011), unconscious, habitual (Dickinson, 1985), direct (Sutton and Barto, 1998), or model-free (Daw et al., 2005); the second process referred to as System 2 is conscious, goal-directed, indirect, or model-based, respectively. Here I adopt the nomenclature of model-free and model-based control, as the algorithms I implement are borrowed from reinforcement learning theory rather than psychology or economics. In Figure 2.1 I present characteristics of these two modes of control. A major trade-off concerns statistical efficiency and computational power. A model-based system can use sparse data to make predictions about never-seen-before situations, but at a cost of computationally expensive forward planning and calculation. In contrast, a model-free system can only rely on previous experience without extrapolation to novel situations, and in doing so is computationally lean and fast. Critically, this suggests that an organism needs to determine what controller to employ for any given problem, and this in turn depends on the statistics of the environment (Simon and Daw, 2011). After discussing model-free and model-based reinforcement learning in sections 2.2.2 and 2.2.3, I will turn to the question of trade-off between these control strategies in section 2.2.4.

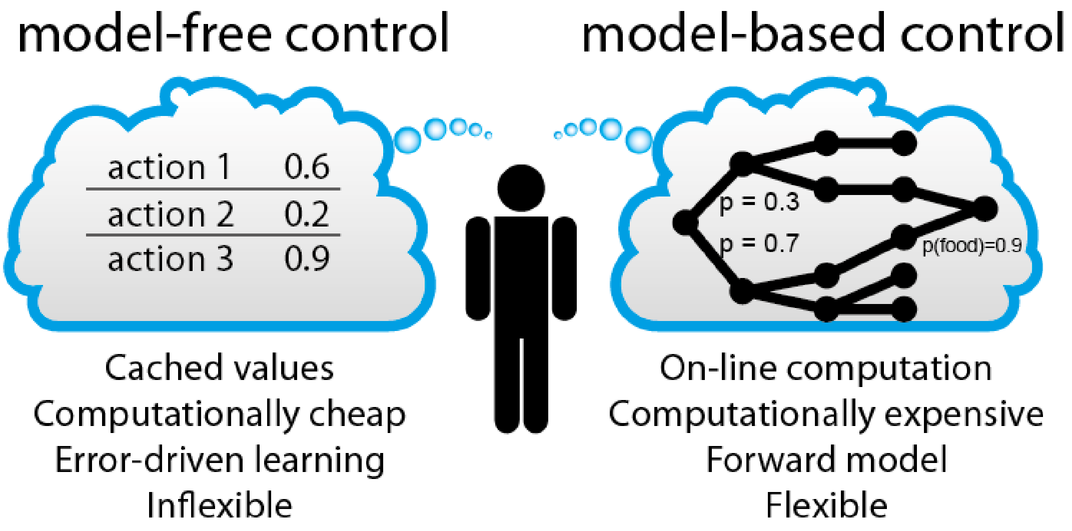


Figure 2.1: Model-free and model-based decision strategies and their characteristics. A similar set of contrasting properties can be used for many other dual-process theories of decision-making.

2.2.2 Model-free RL

2.2.2.1 Model-free control in animal and human psychology

Attempts to understand animal behaviour started in earnest with Thorndike in the late 19th century (Thorndike, 1898). He introduced the intuitive Law of Effect, noting that actions that are followed by pleasant consequences are likely to be repeated, whereas behaviour followed by unpleasant feedback is likely to be avoided (Thorndike, 1911). He thus placed concepts of action, reward and learning within the same context. This behaviourist perspective involving error-driven learning was further pursued by Skinner using operant conditioning paradigms (Skinner, 1938). It is important to establish at this stage that this thesis is concerned with instrumental learning—that is, learning what actions to perform and what actions to avoid. This is distinct from classical conditioning illustrated by Ivan Pavlov's work (Pavlov, 1906), whereby associations are learned between unconditioned reinforcers (e.g. a bell) and conditioned reinforcers (e.g. food), with no action by the animal itself. Models of conditioning, such as the Rescorla-Wagner update rule (Rescorla and Wagner, 1972; Wagner and Rescorla, 1972), partially share the mechanics of error-driven updating with some instrumental learning models discussed in this thesis (though Rescorla has updated his view and considered Pavlovian learning to be equivalent to learning a model of the environment; Rescorla, 1988). Nonetheless, I will restrict my discussion to action learning and refer to published works for more background on conditioning (Pavlov and Anrep, 1960; Rescorla and Wagner, 1972; Pavlov, 2003; Gazzaniga, 2004).

More directly relevant to a current understanding of model-free reinforcement learning is the work by Anthony Dickinson and colleagues. They developed the gold standard for assessing model-free (‘habitual’) behaviour in the form of the devaluation paradigm (Dickinson et al., 1983; Dickinson, 1985; Balleine and Dickinson, 1998). In this paradigm an animal is trained to press a lever to obtain food, and the food is subsequently devalued by lacing it with lithium or satiating the animal on that specific food. When placed back into the operant chamber, the rat will continue or stop pressing the lever depending on whether the learning phase was long or short, respectively. Given that the test phase is in extinction, i.e. without feedback, in order to refrain from obtaining the now devalued food the rat must have a representation of the consequences of its action. This is termed goal-directed or model-based control. In contrast, once the lever pressing has been engrained as a model-free habit the mere presentation of the lever stimulus triggers a response without consideration of its consequences. This stimulus-response behaviour is also called habitual or, as described throughout this thesis, model-free control. As we will see in section

2.2.2.4.3 this paradigm has been widely used to map the neural substrates of model-free control.

Although the devaluation paradigm has been applied to humans to test for model-free control (Valentin et al., 2007; Tricomi et al., 2009), in chapters 5, 6 and 7 I used a different assay that captures any expression of model-free control rather than habitual actions that have been ‘stamped in’ during over extended training sessions (Wise, 2004).

2.2.2.2 Model-free control in artificial intelligence

Most would agree that a robot pre-programmed to execute a set of tasks or movements, such as those found in 19th-century factories, is not intelligent. Slightly more complex are those machines whose actions depend, through pre-set rules, on measurement of the environment—a thermostat or a movement-activated lamp, for example. Yet more interesting are entities that can adapt and learn from their environment, where behaviour is not fixed but is now programmed to adapt. The field of artificial intelligence, and in particular reinforcement learning, has endeavoured to build such algorithms for the past sixty years. The goal is to find efficient ways of choosing actions that maximise reward. In the introduction to this thesis I introduced concepts like reward and value functions, and here I will briefly elaborate on specific examples of such algorithms. It has proven useful to think of (artificial) behaviour in terms of Markov Decision Processes (MDPs; Bellman, 1956; Howard, 1960; Markov, 1971). In this framework the agent transitions through discrete states, and probabilistic transitions are governed by the choices of the agent. Rewards are available in some states, and the goal of the agent is to choose actions so as to maximise rewards in the long run. This framework can encompass both model-free agents, as described shortly, as well as model-based agents (Figure 2.2A). Although MDPs are widely used, there is a mismatch between its discrete states and the continuous nature of the world. An alternative way of evaluating these problems is through an estimation of some value quantity at each time step, whereby changes in this quantity represent good or bad actions (Minsky, 1954). This is captured in temporal difference learning algorithms, a firmly established method for modelling reinforcement learning agents (Witten, 1977; Sutton and Barto, 1981; Barto and Sutton, 1982). Expectations are built into this framework by value transferring from the inherently rewarding stimuli to predictors of those stimuli, as the predictors themselves come to increase the current estimate of long-run reward.

One framework I describe here is that of Q-learning (Watkins, 1989; Watkins and Dayan, 1992), which brings together various aspects of MDPs and temporal difference models to estimate state-action values (Q-values). Critically, Q-learning does not require a model of the environment unlike methods such as dynamic programming (section 2.2.3.2); it can update estimates incrementally without having to wait for a sequence to be finished, as is the case for Monte Carlo methods; it is often referred to as an off-policy method, meaning that it is guaranteed to acquire the optimal policy even when allowed to explore and choose suboptimal actions (Watkins and Dayan, 1992). This distinguished it from u Sarsa methods which are on-policy (Rummery and Niranjan, 1994). Put simply, an optimal policy can be learned faster under Q-learning despite the presence of exploration (Sutton and Barto, 1998). Nonetheless, in chapter 4 onwards I also used Sarsa, a slight modification on Q-learning that is only relevant to multi-step problems. It has been suggested that learning signals in animals resemble Sarsa rather than Q-learning approaches (Morris et al., 2006).

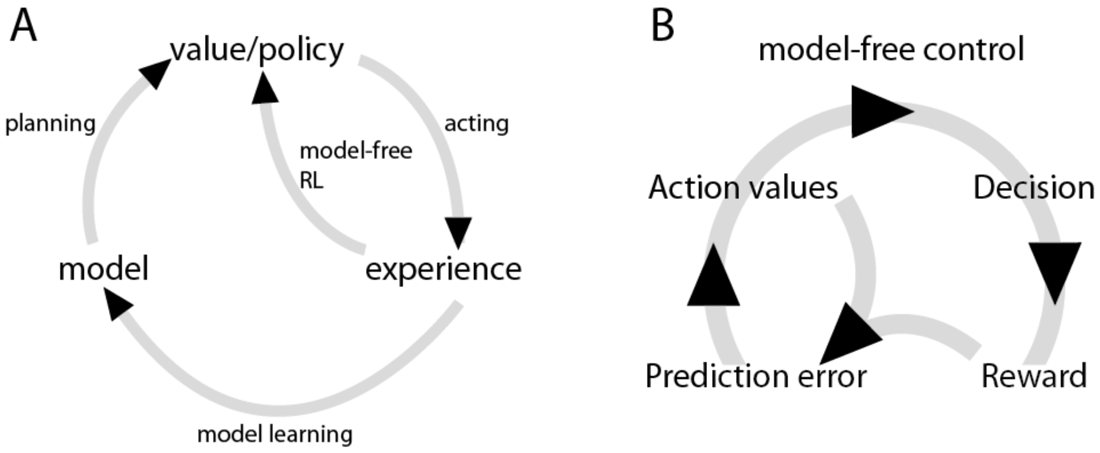


Figure 2.2: Schematics of two reinforcement learning strategies. (A) A model-free agent takes the experience or feedback from actions and directly updates the policy without an intervening model. Model-based control has an added complexity whereby a policy arises from an evaluation of a model, which itself has been learned through experience. (B) Examining model-free learning more closely, action values are updated through a reward prediction error. Figure A is based on figure 9.2 in Sutton and Barto (1981).

Having discussed developments in animal learning theory as well as in artificial intelligence, these two strands of research converged with work in the 1990s showing dopamine signals can be described in terms of reinforcement learning signals (Houk et al., 1995; Montague et al., 1996; Schultz et al., 1997). Finally, Daw et al. (2005) explicitly framed animal and human temporal difference learning as model-free control. In the next two sections I will describe the algorithm of Q-learning as used throughout this thesis, followed by an overview of the neural correlates of a model-free system in the rodent and primate brain.

2.2.2.3 Algorithms

I will briefly describe the basics of Q-learning, an algorithm that describes how an agent without an explicit model of its environment might learn what actions to take and what actions to avoid (i.e. learn an optimal policy). The equations are adapted from Sutton and Barto (1998) and described in their one-step form, that is, without an eligibility trace that allows for action values more than a single action back in the past to be updated. The implementation of a two-step eligibility trace is described in section 4.4.

Q-learning attempts to learn the value of state-action pairs in an MDP environment. The MDP consists of states $s \in S$ where actions $a \in A(s)$ are available. The agent tries to obtain Q^*, which is the optimal action-value function. This function is approximated by, at each time t, computing the following (Equation 6.6 in Sutton and Barto, 1998):

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha \left[r_{t+1} + \gamma \max_a Q(s_{t+1}, a) - Q(s_t, a_t) \right]$$

Where r_t is the immediate reward at time t, γ is a fixed discount factor for future value, and $\max_a Q(s_{t+1}, a)$ represents the action value of the best action in the state the agent ends up in after action a_t from s_t. Critically, it does not depend on what action is *actually* chosen in $s_{(t+1)}$. This is the only difference between Q-learning and Sarsa, which rather than the argmax uses the actual chosen option on the next state (in boldface, and as used in chapter 4; Rummery and Niranjan, 1994; Sutton and Barto, 1998):

$$Q(s_t, a_t) \leftarrow Q(s_t, a_t) + \alpha[r_{t+1} + \gamma \boldsymbol{Q(s_{t+1}, a_{t+1})} - Q(s_t, a_t)]$$

Central to the function of this agent is the learning rate α, which describes the weighting function of past experiences, and the notion of reward prediction error

(RPE), which is the term multiplied by α. A low α means that the Q-values are updated slowly, such that even events far in the past still influence the current estimate. A learning rate of 1 means only the current event is used to estimate the Q-value through the RPE. Although beyond the scope of this discussion, in static or variably volatile environments a fixed learning rate is suboptimal (Yu and Dayan, 2005; Courville et al., 2006; Simon and Daw, 2011). Variable learning rates have been implemented in Bayesian frameworks and here correlates were also observed in the brain (Behrens et al., 2007a; Behrens et al., 2008). As I did not manipulate volatility in the learning experiments I used a more conventional fixed learning rate approach (Daw, 2011).

Once the state-action values are known, an action is selected through some policy, for example by selecting the best action except on a random set of trials where an agent chooses randomly with probability ε (hence the policy's name—ε-greedy; e.g. Daw et al., 2006). Throughout this thesis I used a *softmax rule*, which assigns a probability to each of n actions in a state:

$$p(a_i) = \frac{e^{\beta * Q_{a_i}}}{\sum_{b=1}^{n} e^{\beta * Q_{a_b}}}$$

where β is the inverse temperature. A low inverse temperature means all choice options are almost equiprobable, irrespective of their Q-value. High inverse temperatures pushes choices towards the option with the highest Q-value, irrespective of how small the difference might be. As such, this parameter is linked to the exploration/exploitation trade-off (Daw et al., 2006), though equally it captures the unpredictability (noise) in the agent's choices. I added various parameters to these models in chapters 4-8 though these merely add a bias, extra learning rate or eligibility trace, without affecting the fundamental characteristics of a temporal difference model. I will now turn to the remarkable similarities between physiological signals and components of the algorithms described above, which is the real reason reinforcement learning has played such a central role in the neuroscience of learning and decision-making for the past 20 years.

2.2.2.4 Neural correlates of model-free learning

2.2.2.4.1 Dopamine signalling a prediction error

No paper more clearly signalled the fusion of empirical neuroscience and computational neuroscience than Schultz et al. (1997), who showed that the firing of cells in the dopaminergic midbrain resembles an update signal from temporal difference models. Previous work had already shown that dopamine neurons fire in response not only to rewards but also to stimuli predicting rewards (Romo and Schultz, 1990; Schultz et al., 1993) as well as to unexpected rewards (Mirenowicz and Schultz, 1994). In parallel, work from Dayan, Houk et al. (1995) and Montague et al. (1996) developed theoretical frameworks which led to the critical insight that DA neuron firing is most parsimoniously explained in terms of TD learning (Schultz et al., 1997). Corroborating results have been found using various techniques, including fMRI of the midbrain (D'Ardenne et al., 2008; Duzel et al., 2009), human single-unit recordings (Zaghloul et al., 2009) and, perhaps most critically, in a rodent model whereby individual neurons in the midbrain were identified as dopaminergic or GABAergic (Cohen et al., 2012). In this pivotal study it was shown that dopamine cells indeed signal a reward prediction error, whereas GABAergic interneurons provide the expected value signal (Figure 2.3; Cohen et al., 2012).

Further evidence from optogenetic stimulation showed this pattern had causal properties in that phasic firing of DA neurons led to behavioural conditioning (Tsai et al., 2009). Lastly, it seems the RPE signalled by dopamine is derived from a Sarsa model rather than Q-learning (Morris et al., 2006). Over the past 20 years, then, the dopamine system has been firmly embedded in model-free error-driven learning. However, we will see in chapter 11 that this signal might also reflect updates from more complex value systems such as model-based control (Daw et al., 2011).

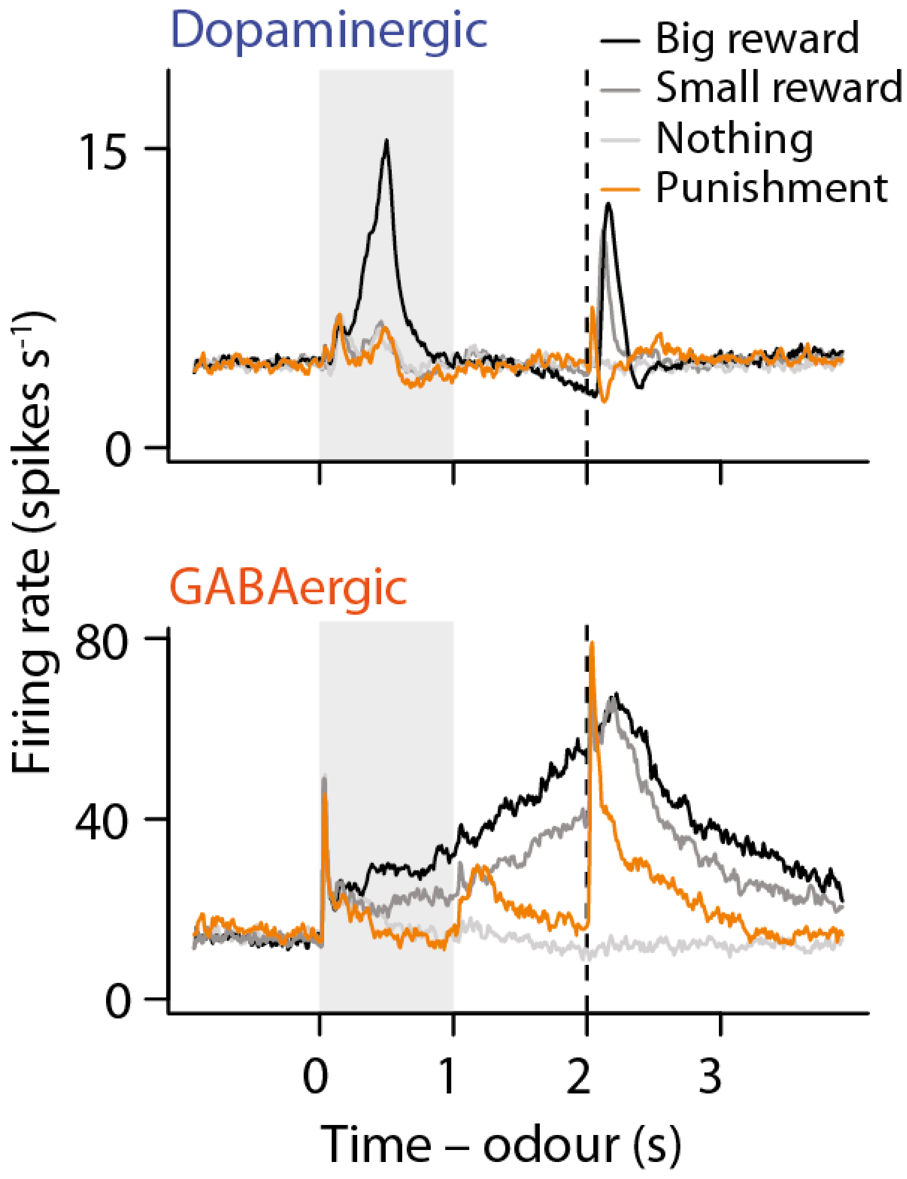


Figure 2.3: Firing rates for identified dopaminergic and GABAergic cells in the ventral tegmental area. Dopaminergic cells increase their firing rate in response to an odour cue that has been learned to signal a big reward, in line with a theory for dopamine signalling a temporal difference reward prediction error. GABAergic cells had a markedly different response profile—a sustained rather than transient pattern of firing that scaled with expected value. This suggests these units provide the subtractive component to reward prediction error. Figures reproduced from Cohen et al. (2012).

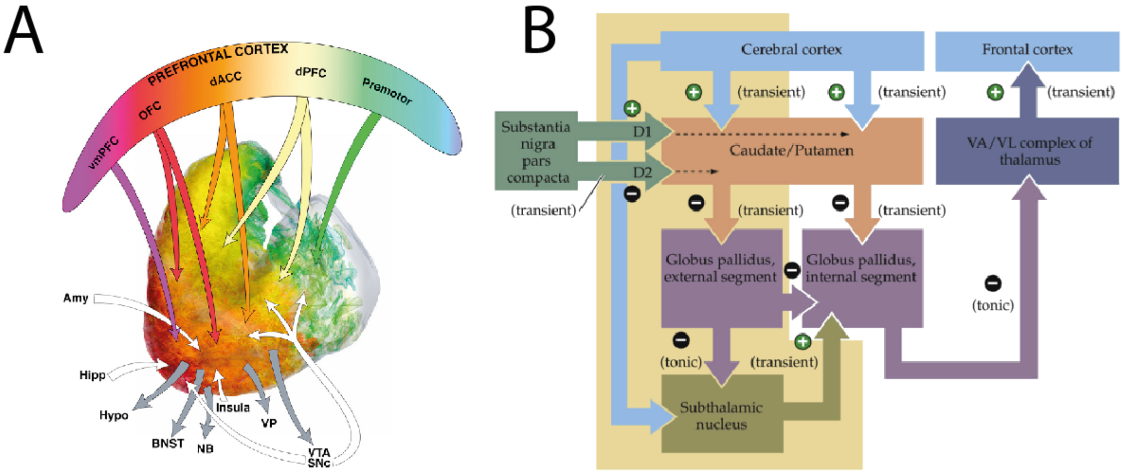
2.2.2.4.2 The anatomy of the basal ganglia

If we accept that dopaminergic cells in the midbrain signal a reward prediction error to update action values, where might these values themselves be updated and stored? Ascending dopaminergic projections arise along roughly eight different pathways, of which the mesolimbic and nigrostriatal projections are the strongest (Steiner and Tseng, 2010). Both these pathways project to the striatum: the mesolimbic projections enter the nucleus accumbens, the most rostroventral part of the striatum; the nigrostriatal pathway terminates all across the putamen and caudate nucleus (Fallon and Moore, 1978; Beckstead et al., 1993; Haber et al., 2000).

The anatomy and structural connectivity of the striatum continue to provide inspiration for empiricists and theorists. In chapter 8 I explore the relationship between this anatomy and how it relates to function, and in chapter 9 I study its function in proactive inhibitory control, so I will take some time here to make clear some of the anatomical features of the basal ganglia.

Starting at the macro-anatomical level, almost all of cortex projects to the striatum, which is the primary input region of the basal ganglia (Alexander et al., 1986; Alexander and Crutcher, 1990; DeLong, 1990; Haber et al., 2000; Haber, 2003). These pathways then continue through the pallidum, subthalamic nucleus, substantia nigra and thalamus before reaching cortex again, giving rise to the term cortico-basal ganglia-thalamo-cortical ‘loop’ (Alexander et al., 1986). This was initially speculated to subserve motor control by funnelling diverse inputs directly into motor cortex via the thalamus (Kemp and Powell, 1971). Further work revealed that despite a strong convergence of inputs, cortical topography is maintained throughout, such that thalamo-cortical projections reach most parts of cortex that provided the initial inputs. These pathways have been divided in many different ways, each with its own naming convention and grouping criteria (Alexander et al., 1986; Alexander and Crutcher, 1990; Haber, 2003), a testament to the inherently continuous nature of any topographically organised network (Figure 2.4A; Haber and Behrens, 2014). Crudely, these pathways encompass limbic, associative and motor loops, covering the entire frontal cortex as well as motor regions (Haber, 2010).

A microscopic level analysis of the striatum shows glutamatergic projections from cortex arriving on dendrites of GABAergic medium spiny projection neurons (MSNs; Somogyi et al., 1981), which make up about 95% of cells in the striatum (Kemp and Powell, 1971). They are distributed homogeneously throughout the striatum and come in two equally numerous types: so-called ‘direct pathway’ MSNs project to the internal segment of the globus pallidus (GPi) or substantia nigra (SN), whereas ‘indirect pathway’ MSNs project to the external segment of the globus pallidus (GPe; Loopuijt and Van der Kooy, 1985; Kawaguchi et al., 1990). Alternative naming for these two pathways are striatonigral and striatopallidal or Go and Nogo, respectively (Figure 2.4B). Although there are reports that the two pathways do not receive identical inputs from cortex (Lei et al., 2004; Reiner et al., 2010; Wall et al., 2013), their dominant feature reflects each loop consisting a direct and indirect component that receives similar input from cortex. Activation of the direct pathway leads to GABAergic inhibition of the GPi, which in turns *dis*inhibits the thalamus and cortex. Conversely, activation of the indirect pathway leads to a triple negative by adding inhibitory projections from the GPe, leading to inhibition of cortex (Figure 2.4B). This architecture in principle allows for interesting computations and functions, such as selection of cortical representations and a gating of actions, as I discuss in the next section.



Multiple, partially overlapping loops

Direct and indirect pathways

Figure 2.4: Levels of organization in the basal ganglia. (A) The entire frontal cortex has topographically organised projections to the striatum. In this schematic it is emphasised that projections are partially overlapping. (B) The input from cortex is processed through a direct and indirect pathway. Activity in these pathways has an excitatory and inhibitory effect on cortex, respectively. Note that dopaminergic input has opposite effects on these two pathways. Figure A is reproduced from Haber and Behrens (2014), Figure B from Purves et al. (2001).

As introduced at the start of this section, nigrostriatal projections can alter synaptic transmission throughout the striatum. Dopamine acts as a neuromodulator and is primarily released around the neck of dendritic spines (Freund et al., 1984). This dopamine diffuses to affect D1 receptors up to 2 µm and D2 receptors up to 7 µm from the synapse before being removed from the extracellular space via dopamine reuptake (Cragg and Rice, 2004). The difference in effective radius is due to the higher affinity of D2 compared to D1 receptors (Richfield et al., 1989). Critically, these two receptors have opposing effects on MSN excitability (Hartman and Civelli, 1997) and (although initially contested, Surmeier et al., 1993; Aizman et al., 2000; Gerfen and Bolam, 2010) are now known to neatly differentiate direct from indirect pathway MSNs (Gerfen et al., 1990). That is, direct pathway MSNs express depolarizing D1 receptors, and indirect pathway MSNs express hyperpolarizing D2 receptors (Figure 2.4B). Dopamine can thus modulate the balance of activity between these pathways by activating the direct and inhibiting the indirect pathway (Gerfen and Surmeier, 2011).

To summarise, the basal ganglia receives strongly converging inputs from cortex and is structured along parallel, possibly interacting, loops. It has direct and indirect pathways passing through pallidum, subthalamic nucleus and thalamus to modulate limbic, associative and motor representations in cortex.

I now discuss studies in animal models and humans examining the function of the basal ganglia system in model-free reinforcement learning.

2.2.2.4.3 Function of the basal ganglia in model-free reinforcement learning

Although the anatomy and microcircuitry of the basal ganglia have been carefully mapped, it has proven more difficult to study the function of each of its components. This is partly due to extensive overlap of the direct and indirect pathway, and their indiscriminability in classical electrophysiological recordings. In this section I will briefly discuss foundational lesion work before highlighting electrophysiological recordings and recent optogenetics studies of the basal ganglia circuitry in the context of choice and learning. Human work on the characteristics of the striatum during reinforcement learning is plentiful, but it is often as crude as lesion studies in terms of understanding fine microcircuitry. Nonetheless key insights from animal studies have helped inform the role of the basal ganglia in higher cognitive function. Lastly I will highlight some of the more influential computational models that place reinforcement learning in frontostriatal circuits as they usefully highlight the relationships between key components of the network.

2.2.2.4.3.1 Animal work on habits in the basal ganglia

In section 2.2.2.1 I briefly discussed the work of Anthony Dickinson and colleagues on habitual behaviour. Habits are also known as stimulus-response (S-R) associations that have been ‘stamped in’ through reinforcement (Thorndike, 1898; Landauer, 1969). Early work suggested a critical role for the basal ganglia in habit formation (e.g. Mishkin et al., 1984; Salmon and Butters, 1995), and a strong body of work in rodents now shows that such stamping in occurs in the dorsolateral striatum, also known as the putamen in primates. Using a devaluation test (Dickinson et al., 1983), it was shown that muscimol-induced deactivation or destructive lesions of the dorsolateral striatum made the animal perpetually sensitive to devaluation, suggesting its effect was to impair an ability to form habits (Featherstone and McDonald, 2004; Yin et al., 2004; Yin et al., 2005; Yin and Knowlton, 2006). Other work suggested that the formation of these model-free S-R associations relies heavily on dopamine (Wise and Bozarth, 1987; Faure et al., 2005; Wickens et al., 2007), which is known to modulate the plasticity in corticostriatal synapses (Kötter, 1994; Calabresi et al., 2007). Together, this work provides some clues regarding the role of dopamine and corticostriatal loops in habits, a quintessential form of model-free control.

2.2.2.4.3.2 Animal work on model-free RL in striatum

I now return to the question posed at the start of this section, namely how does the striatal system encode cached, model-free, action values that are updated by dopaminergic reward prediction errors as described in section 2.2.2.4.1. Electrophysiological recordings showed that in the period before a value-based choice, up to a third of striatal neurons code an action-specific value derived from a reinforcement model (Samejima et al., 2005; Samejima and Doya, 2007; Lau and Glimcher, 2008; Kim et al., 2009; Nakamura et al., 2012). Reward prediction errors, despite their being signalled through diffuse dopaminergic projections (Fallon and Moore, 1978), also show action-specificity in the dorsal striatum (Stalnaker et al., 2012). In an oculomotor task neurons in the caudate could be segregated into two groups that separately code action and outcome (Lau and Glimcher, 2007). In a delay discounting task neurons in the caudate coded temporally discounted action values (Cai et al., 2011b). Taken together, these data suggest the dorsal striatum exhibits both action coding and action value coding, providing a necessary neural substrate for dopaminergic prediction errors to update action values after feedback (Samejima and Doya, 2007; O'Doherty, 2014). However, none of these studies were able to determine whether the neurons were part of the direct or indirect pathway, leaving critical questions about the mechanics of this microcircuitry unanswered.

Novel techniques such as optogenetic perturbation of MSNs have provided long sought-after evidence for theories on direct and indirect pathway function. First, stimulation of direct and indirect MSNs led to behavioural activation and inhibition, respectively (Kravitz et al., 2010; Kravitz et al., 2012), and coordinated co-activation of both pathways is necessary for basic movements (Cui et al., 2013; Jin et al., 2014). Critical to model-free reinforcement learning is the evidence that optogenetic stimulation of direct pathway MSNs acts as persistent, long-lasting reinforcement, whereas stimulation of indirect pathway MSNs acts as transient punishment (Kravitz et al., 2012). These manipulations worked even in the presence of D1 and D2 receptor antagonists, suggesting dopamine-independent activation of these pathways is sufficient to generate, in this specific instance, place preference learning (Kravitz et al., 2012; Paton and Louie, 2012).

In summary, this work supports the view that reinforcement learning in rodents is subserved by dopaminergic prediction errors modulating action values stored in corticostriatal pathways. When faced with a stimulus, the direct pathway action channel of the basal ganglia most strongly associated with the stimulus (through previous reinforcement) will be the strongest candidates for expression. Next I will consider what evidence we have in humans for such a mechanism of model-free RL.

2.2.2.4.3.3 Human striatum in model-free RL

Only with the advent of optogenetics and viral Ca^{2+} indicators could neuroscience begin to directly measure the direct and indirect pathway (Kravitz and Kreitzer, 2011; Cui et al., 2013). In humans, then, it is unclear how to understand striatal function given that we currently lack the tools to study these same pathways in great detail. An outstanding challenge is to functionally and structurally delineate the presence of a direct and indirect pathway in humans. Functional MRI has been the dominant approach to studying striatal function, mostly driven by necessity as non-invasive electrophysiological methods cannot (yet) measure signals deep in the brain. FMRI studies have shown model-free reinforcement learning signals across the striatum. An early influential paper suggested the caudate is the ‘actor’, representing action values during instrumental choice and RPEs during instrumental (but not Pavlovian) learning (O'Doherty et al., 2004). The dorsal striatum seems to only care for value when *actions* are involved (Tricomi et al., 2004; Guitart-Masip et al., 2011; Guitart-Masip et al., 2014). Many studies since have described model-free action and learning signals in human fMRI (for reviews see Balleine and O'Doherty, 2010; Dolan and Dayan, 2013; Haber and Behrens, 2014). A recurring theme in this thesis is the dichotomy between model-based and model-free values. However, the studies cited above did not employ tasks that can dissociate value signals from these two controllers. Recent work has shown that when these two controllers are contrasted, model-free values are observed specifically in the putamen and not caudate (Wunderlich et al., 2012b; Lee et al., 2014), corroborating animal work that finds dorsolateral, but not dorsomedial, striatum to be critical for model-free (habit) learning (see section 2.2.2.1). Together, these studies show the dorsal striatum is a prime candidate for both the selection of actions as well as learning about these actions based on feedback (Jessup and O'Doherty, 2011).

2.2.2.4.3.4 Basal ganglia models of model-free RL

The exquisite architecture of the basal ganglia as described above has led to many models of its function relating to (value-based) action selection and action learning (Houk and Wise, 1995; Mink, 1996; Doya, 1999; Gurney et al., 2001; Frank, 2005; Bogacz and Gurney, 2007; Hong and Hikosaka, 2011), working memory (Frank and O'Reilly, 2006; Hazy et al., 2006) and incentive salience (Berridge, 2007), as well as many other functions such as speech (e.g. Civier et al., 2013). The computational function most commonly ascribed to the basal ganglia can be summarised as ‘selection’—be it items in working memory or actions being represented in cortex. As such, the striatum can be thought of as being in a loop necessary for selection, where learning, feedback and selection all co-occur in the same brain region (Jessup and O'Doherty, 2011).

As an exemplar model of instrumental learning I use that of Frank et al. (2004) (Figure 2.5). At the heart of this model are a number of action channels (e.g. for right- vs left-handed response) that are duplicated for a direct and indirect pathway. Their relative activity determines which action reaches threshold in premotor cortex for execution. The direct and indirect pathways are activated through inputs from cortex, serving as the ‘stimulus’ that triggers a response in S-R learning. On the first trial an action is randomly selected (through noise) and the resulting positive or negative feedback leads to an increase or decrease in dopamine release from the midbrain. This critical step induces long-term potentiation (LTP) in the direct pathway for positive feedback, and in the indirect pathway for negative feedback. As such, the direct pathway action channel for an action that led to reward will undergo LTP, leading to stronger activation with the next occurrence of the same stimulus (‘stamping in’; Thorndike, 1898). Conversely, an action that leads to low rewards will have its *indirect* pathway channel strengthened and is less likely to be selected on the next occurrence of the stimulus. Over multiple trials, then, the network learns S-R associations that are most likely to lead to reward. This simple but elegant coalescence of stimuli, actions, selection and learning reflects much of our thinking about basal ganglia function in health as well as disease (Tekin and Cummings, 2002; Maia and Frank, 2011; Dichter et al., 2012). Nonetheless, novel techniques for measuring and manipulating direct and indirect pathways are continuously inspiring altered models (Calabresi et al., 2014). As we will see in chapter 5 another potentially fruitful pursuit could be to understand the role of striatal pathways and dopamine in model-based control (Daw et al., 2011).

Having discussed the algorithms, anatomy and functional correlates of model-free reinforcement learning, I now turn to a (shorter) treatise of model-based control before addressing the relationship between both controllers.

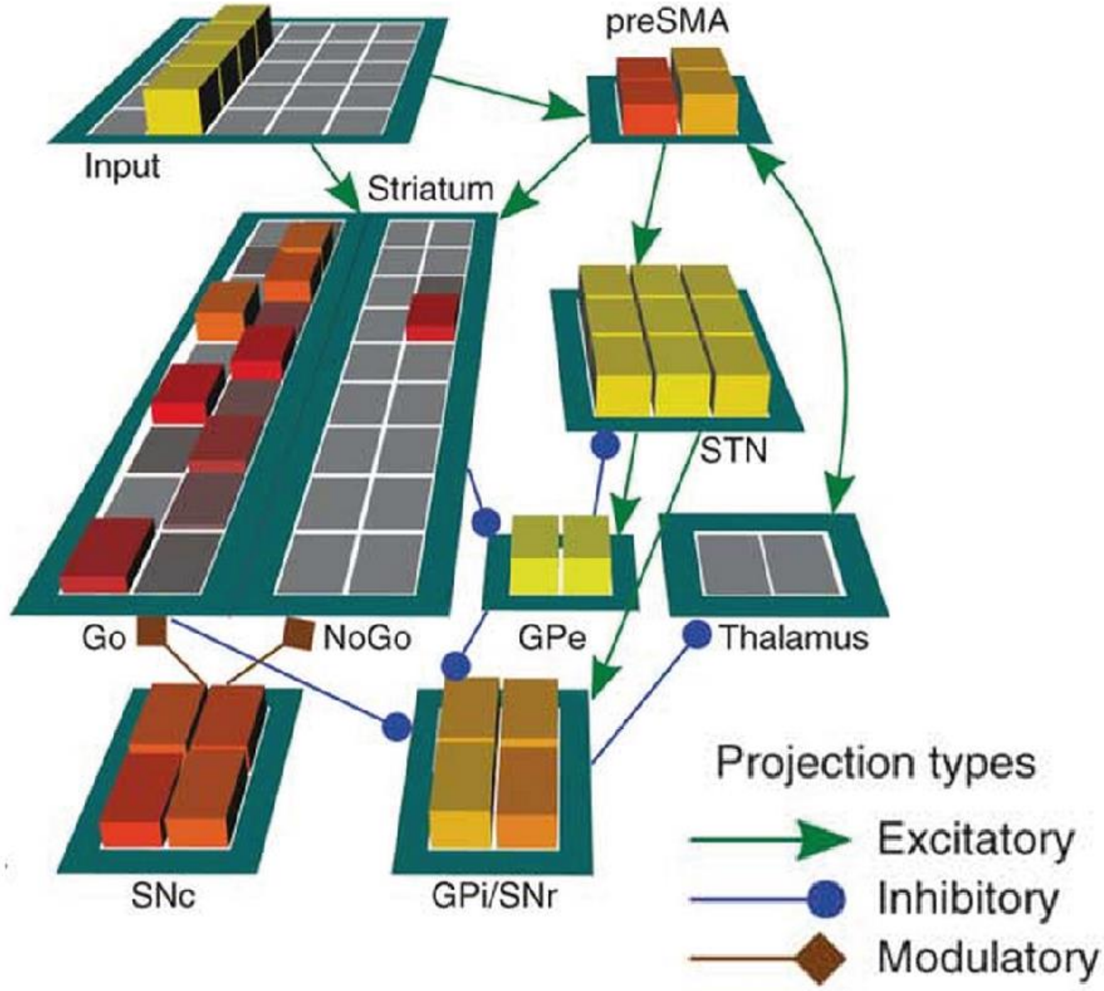


Figure 2.5: An exemplar computational model of cortico-basal ganglia function. Action channels are represented as columns in each region. Once activity in preSMA reaches a certain threshold an action is triggered. To reach this threshold, activity propagates from other cortical areas ('Input') through the direct and indirect pathways of the basal ganglia. These pathways are termed Go and NoGo here, respectively. Dopaminergic modulatory signals from the substantia nigra pars compacta (SNc) govern the relative activity and learning in the two pathways. The figure is reproduced from Maia and Frank (2011).

2.2.3 Model-based RL

2.2.3.1 Brief history of model-based RL—animal learning and psychology

Whereas model-free S-R learning started in the final years of the 19th century with Thorndike's work, it took three decades to mount an offensive against a purely S-R account of animal behaviour. Tolman's work suggested a fundamentally different view of behavioural control—one in which cognitive maps of the environment are learned as S-S associations and guide decisions through a mental search of these maps (Tolman, 1932; Tolman, 1948). Such cognitive maps have since been shown to exist in cell assemblies in the hippocampus (Keefe and Nadel, 1978). This idea of prospection and representations of future decisions and outcomes was operationalised by Dickinson and colleagues as goal-directed control—the antipode of habitual control (Adams and Dickinson, 1981; Dickinson et al., 1983; Balleine and Dickinson, 1998). As explained in section 2.2.2.1, an animal that does not pursue actions known to lead to a devalued outcome must have a response-outcome (R-O) association whose desirability is assessed at the time of choice. Conceptually identical devaluation paradigms were used in humans to measure goal-directed control successfully (Valentin et al., 2007).

However, outcome devaluation leaves the human neuroscientist with precious few trials to study behaviour and its neural correlates, as humans are able to rapidly adjust to novel situations. Spurred along by the fact that the definition and operationalization of goal-directed control had become thoroughly entrenched in the devaluation paradigm, a new terminology was introduced to capture more broadly decision strategies that rely on prospection, models of the environment and mental simulation: model-based control (Doya, 1999; Doya et al., 2002; Daw et al., 2005). This shift in nomenclature has turned the spotlight onto the computations underlying this type of control, such as the complexity of performing mental searches in environments more complex than Skinner boxes (Huys et al., 2012). In-depth reviews on the history of goal-directed and model-based control can be found elsewhere (Balleine and Dickinson, 1998; Rangel et al., 2008; Doll et al., 2012; Dolan and Dayan, 2013). Before describing some of the recent work on the behavioural and neural correlates of model-based control in section 2.2.3.4, I will first touch upon the origin of model-based control in artificial intelligence and the basic algorithms that can describe such an agent.

2.2.3.2 *Brief history of model-based RL—artificial intelligence*

Whereas model-free control need only keep track of cached state-action values, model-based control requires on-the-fly calculation of optimal decisions based, in the most extreme case, on an evaluation of all possible future states. As in model-free RL (section 2.2.2), this challenge was approached as a Markov Decision Process (MDP) with discrete states, actions and probabilistic transitions between states. Nonetheless, planning even a few steps ahead in a relatively contained situation, such as chess for example, leads to a combinatorial explosion of possible states and actions that would require evaluation—the 'curse of dimensionality' (Bellman, 1956). A computational framework to formally address this forward search was first described by Bellman (1956) in what is now called the Bellman equation (section 2.2.3.3). It generally requires complete knowledge of the system such that it can be evaluated all the way through to an end state. In the following decades much work went into finding shortcuts and efficient ways of approximating the optimal solution, an endeavour particularly interesting given that the brain's model of the world is inherently incomplete and neural resources are a valuable commodity. For an in-depth review of this work I refer the reader to established works (Bryson Jr, 1996; Sutton and Barto, 1998) .

2.2.3.3 *Algorithms*

The Bellman optimality equation for the optimal policy Q^* is as follows (after equation 4.2 in Sutton and Barto, 1998)

$$Q^*(s,a) = \sum_{s'} P^a_{ss'}[R^a_{ss'} + \gamma \max_{a'} Q^*(s',a')]$$

where $P^a_{ss'}$ denotes the transition probability going from s to s' given action a, $R^a_{ss'}$ the immediate reward available if that transition indeed happens, and γ denotes a discount factor for future value. Its formulation is remarkably simple: the value of an action equals the sum over the value of each possible consequent state multiplied by their probability of occurring, with each consequent state evaluated by searching all *its* consequent states. In this sense it is equal to the expected or Pascalian value as used in behavioural economics, for $\gamma = 1$. Note also that the algorithm assumes that in each future state the agent will choose the best available action, and therefore does not account for possible lapses or exploratory choices. Lastly, this type of model is a *distribution model*, in that it assesses the entire distribution of possible outcomes. We now turn to the brain and discuss some of the neural instantiations of model-based control.

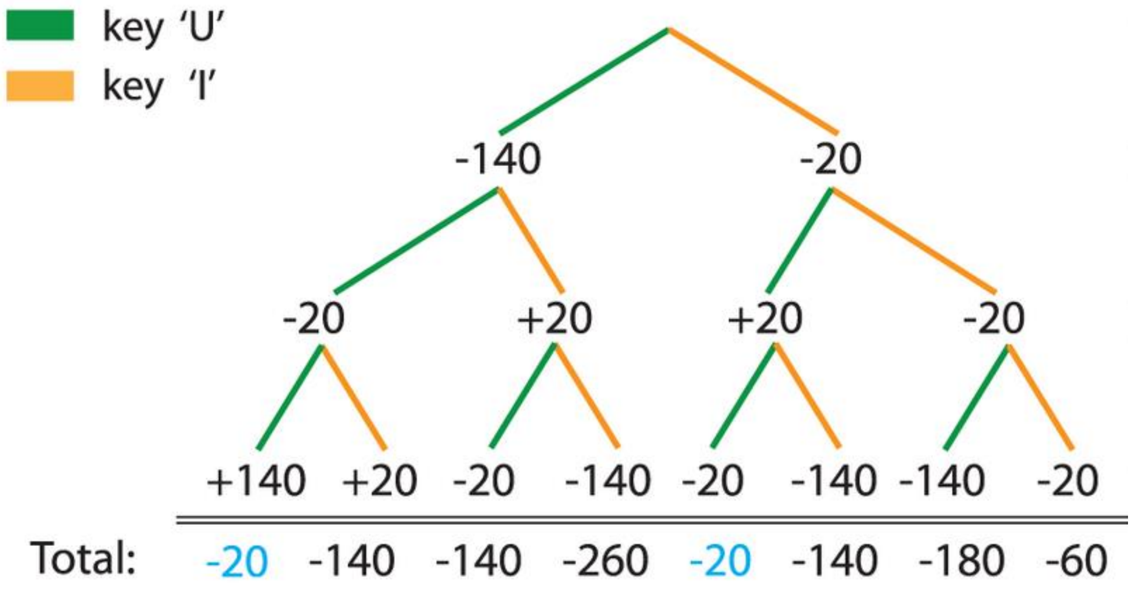


Figure 2.6: Example model-based tree search for deterministic transitions. The decision-maker is currently at the top (or root) of the tree and is prospecting to calculate the value of 8 possible outcomes. Model-based planning assumes perfect knowledge regarding the values and transitions at each state. This figure is reproduced from Huys et al. (2012), who observed that planners prefer routes that do not involve a large loss given equal final outcomes. This suggests that humans have developed methods of pruning the decision problem to save computational expense.

2.2.3.4 Neural mechanisms of MB RL, ubiquity of substrates

The complexity of computation in a model-based system is reflected in its neural underpinnings. Unlike the relatively straightforward neural instantiation of model-free RL in the striatum, model-based control is much more loosely defined and components of its computations have been found across cortex, striatum and hippocampus. Indeed, one might couch many things, including planning, processing of fictive feedback, learning a model, and revaluation as model-based control of some sort simply because it cannot be done model-free. In this short review I focus primarily on studies related to value-based choice in a learning environment. Others have dealt with related topics such as decisions in the framework of neuroeconomics (Rangel et al., 2008; Glimcher and Fehr, 2013), working memory (Curtis and Lee, 2010; Baddeley, 2012) and planning (Owen, 2005; Tanji et al., 2007).

Along with studies showing a critical role for dorsolateral striatum in model-free control (section 2.2.2.4.3.1) it was shown in animal experiments that dorsomedial striatum and prelimbic cortex are critical for goal-directed control (Balleine and Dickinson, 1998; Corbit and Balleine, 2003; Killcross and Coutureau, 2003), in particular its acquisition (Ostlund and Balleine, 2005). Prefrontal and striatal correlates have also been found in humans on numerous occasions. For example, activity in the ventral orbitofrontal cortex is reduced for devalued compared to non-devalued stimuli, suggesting this region represents prospective rather than cached values (Valentin et al., 2007). In a planning task without learning it was found that the caudate nucleus represents both end-state and intermediate values, as would be expected from a model-based system (Figure 2.7A; Wunderlich et al., 2012b). Others have focused on how we might build models of our environment to use in planning: Gläscher et al. (2010) translated the latent learning experiment by Tolman and Honzik (1930) to an fMRI study whereby the participant is exposed to an environment in the absence of reward, triggering latent learning of the transition probabilities $P_{ss'}^{a}$ from section 2.2.3.3. Upon the introduction of reward the participant then uses this model to obtain rewards in an efficient manner. Critically, during the latent learning period participants showed ‘state prediction errors’ that could facilitate model learning in the dorsolateral prefrontal cortex (Figure 2.7B; Gläscher et al., 2010). Other prefrontal regions are also implicated in learning model-based associations, including orbitofrontal cortex for cognitive maps (Wilson et al.,

2014) and stimulus-outcome associations (Klein-Flugge et al., 2013). Indeed, much of cortex and sub-cortical structures are involved in model-based control one way or another, a phenomenon described as ‘the ubiquity of model-based RL’ (Doll et al., 2012). The latter authors noted that even regions traditionally presumed to be purely involved in model-free control appear to show model-based influences, hinting at interactions between these two systems. This is the topic of the next section, where we begin to understand how two seemingly disparate controllers might successfully cohabitate in the brain to generate adaptive behaviour.

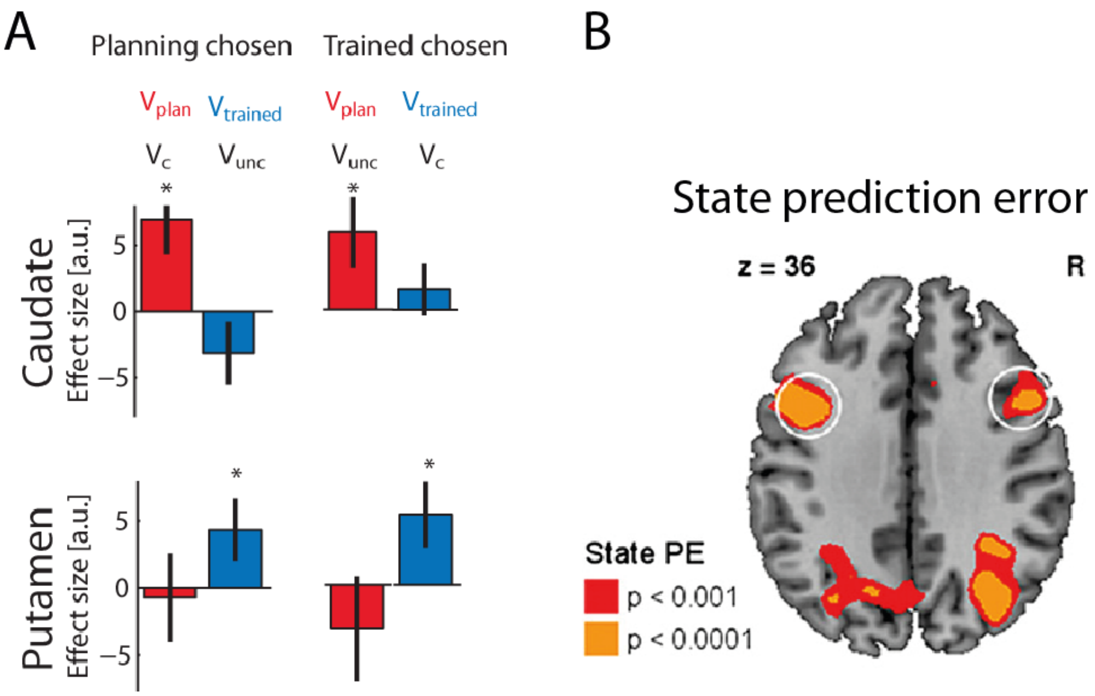


Figure 2.7: Neural correlates of model-based components. (A) In a decision task that involved both planned and extensively trained values, the caudate represents the planned values, whereas the putamen represents extensively trained values. This is independent of what values was eventually chosen, as would be expected from two systems that compete with one another. The functional anatomy is in line with rodent work showing goal-directed function in dorsomedial striatum and habitual function in dorsolateral striatum. (B) Learning a model of the world can conceivably occur through state prediction errors,

which signal unexpected state-state transitions. This figure shows BOLD activity scaling with state prediction errors in dorsolateral prefrontal cortex. In chapter 6 I transiently disrupt this region of the brain using transcranial magnetic stimulation to determine its necessary role in model-based control. Figure A is reproduced from Wunderlich et al. (2012b), Figure B is reproduced from Gläscher et al. (2010).

2.2.4 Balance between MF and MB

2.2.4.1 Rationale of two systems

Why invest in both a model-based and model-free controller if they are designed to achieve the same goal—that is, to maximise rewards? The current line of thinking is that both systems come with their own strengths and weaknesses, such that together they can deal with the statistics and dynamics of our environment. The model-free system, although very efficient in terms of only having to store individual cached values, requires many repetitions (or repeated experience) to approach the true value function. This in itself can be expensive or even life-threatening if the action to be learned is, say, whether to run towards or away from a lion. However, model-free learning can be highly efficient for predictable, repetitive tasks or for motor skill learning—for example, when forming habits. Conversely, the model-based system requires a huge amount of resources in terms of attention, working memory and time but it is statistically efficient and generates decisions from a limited number of samples. Together, this is the computational efficiency versus statistical efficiency trade-off (Sutton and Barto, 1998; Daw et al., 2005; Dolan and Dayan, 2013). In simulation and human behaviour Simon and Daw (2011) showed that the statistics of the environment—the rate of change of the probability ('volatility') and noisiness of reward—differentially favour model-based or model-free control. For example, in a noisy, low-volatility environment a model-free controller performs well as its incremental learning leads to a smoothing out of the noise. Conversely, a low noise, high-volatility environment favours a model-based system as it can more efficiently track rapid changes in the environment without fear of over-fitting the noise (Simon and Daw, 2011). Taken together, organisms with only a single controller might fare well in a restricted set of environments, but will not thrive in the real world, in which the environmental statistics vary widely.

2.2.4.2 Deciding how to decide

If we accept that there are multiple control structures, then what 'controls the controller'? I should note at this point that the shorthand of 'two systems' I have been using so far is merely a convenient way of talking about the different forces that act on behaviour. In fact the reality is that these two systems are embedded in the same brain, and can be seen to describe the extremes of behavioural control that in reality is more likely to be on a continuum (Dolan and Dayan, 2013). In the general discussion I will outline some recent ideas about how these types of control might be intertwined, but for convenience assume two distinct systems and ask how their relative levels of control might be governed.

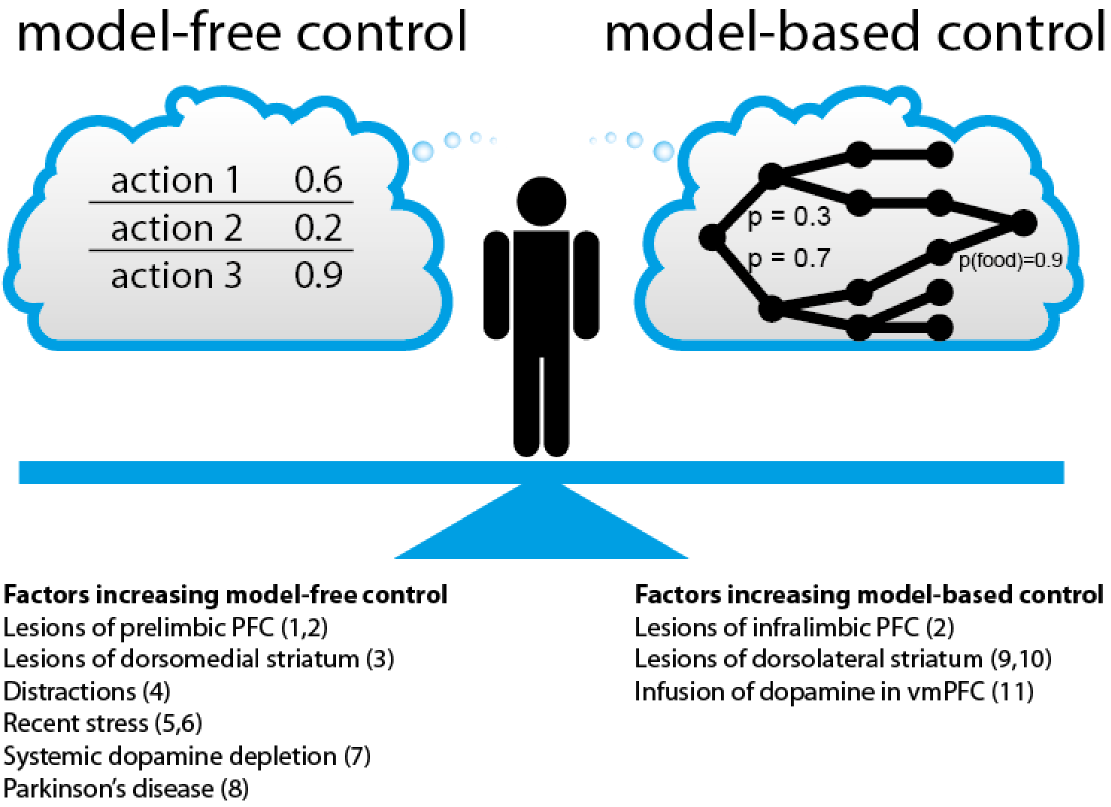
he prevailing hypothesis is that, in true Bayesian fashion, each system exerts control over behaviour dependent on its precision or uncertainty (Daw et al., 2005). Uncertainty in the model-free system arises from the inaccurate, lagged cached values, whereas uncertainty in the model-based system is a product of the computational complexity. A study examining the neural correlates of planned versus cached values found that during choice, the medial prefrontal cortex is functionally coupled with the caudate nucleus—representing planned values—and the putamen—representing model-free values (Wunderlich et al., 2012b). The medial PFC, then, represented the chosen value irrespective of what controller was used to drive that decision. These results suggest two systems that are engaged in parallel. A more detailed study on the competition itself tested directly the notion that uncertainty determines the relative influence of these values (Lee et al., 2014). The latter authors manipulated uncertainty in a model-based system by making the transition probabilities more or less deterministic. In their model, parallel model-based and model-free controllers reported on their reliability through state- and reward prediction errors, respectively (cf. Gläscher et al., 2010). The relative reliabilities then governed the weighting of values contributing to a single integrated Q value. Both these reliability signals, and their maximum, were observed in inferior lateral prefrontal cortex and frontopolar cortex, in line with a role for this region as an arbitrator. Furthermore, they observed that when the arbitrator favoured model-based control, the frontal regions were more negatively coupled to striatal regions known to be involved in model-free control (Lee et al., 2014). This should not be mistaken as evidence for an inhibition of the model-free system by a frontal arbitrator—nor is this argued by Lee et al. (2014)—as their connectivity analysis is ambiguous with respect to directionality and sign. Nonetheless, it provides the first thorough test of arbitration through uncertainty, and future work could help understand the origin of the reliability signals themselves.

An alternative account for the competition between systems has been put forward by Keramati et al. (2011). They suggest that the additional time required for a model-based calculation presents an opportunity cost, that is, time that could have been spent on gathering more rewards. A decision to engage in this calculation then depends on the added value it provides over a model-free prediction. For example, if the model-free system is already confident about the best option, then there is no need to engage in an expensive forward search. This provides an intuitive yet quantitative approach to understanding what factors might drive an investment of mental effort into a problem. Future studies that test the effects of reward rate (and thus opportunity cost) on the trade-off between model-based and model-free choice could provide a test for this model.

2.2.4.3 Shifting the balance of control

It has been suggested that an imbalance in control between a model-based and model-free controller might be implicated in various disorders such as Parkinson’s disease (Redgrave et al., 2010; de Wit et al., 2011); addictions (Everitt and Robbins, 2005) including alcohol dependence (Sebold et al., 2014), food and methamphetamine (Voon et al., 2014); and obsessive-compulsive disorder (Voon et al., 2014). Finding ways of manipulating this balance provides both insights into healthy function and potential avenues for treatment. Figure 2.8 summarises some of the work that has manipulated the extent of control by each system, including through lesions, drugs, disease or cognitive manipulations. Critically, this work shows that disabling one system can reveal behavioural influences of the other system that would otherwise be hidden. As noted before it suggests these systems work in parallel, with model-free learning occurring in the background even if a model-based system is currently in control (Wassum et al., 2009). Conversely, even when a model-free system has the reins a model-based system can swoop in if the situation calls for it

(Isoda and Hikosaka, 2007, 2011). The work in chapters 5-7 further explores how the balance in control can be manipulated using non-invasive methods applicable in healthy humans.



Factors increasing model-free control
Lesions of prelimbic PFC (1,2)
Lesions of dorsomedial striatum (3)
Distractions (4)
Recent stress (5,6)
Systemic dopamine depletion (7)
Parkinson's disease (8)

Factors increasing model-based control
Lesions of infralimbic PFC (2)
Lesions of dorsolateral striatum (9,10)
Infusion of dopamine in vmPFC (11)

Figure 2.8: A non-exhaustive list of work that tilted the balance between model-based and model-free control one way or another. [1] Tran-Tu-Yen et al. (2009), [2] Killcross and Coutureau (2003), [3] Yin et al. (2005), [4] Otto et al. (2013), [5] Schwabe and Wolf (2009), [6] Schwabe and Wolf (2011), [7] de Wit et al. (2012b), [8] de Wit et al. (2011), [9] Yin et al. (2004), [10] Balleine and O'Doherty (2010), [11] Hitchcott et al. (2007)

2.2.5 Thesis work addressing reinforcement learning

The first part of this thesis focuses on reinforcement learning behaviour and its neural correlates. This section has highlighted the importance of multiple driving forces of behaviour, and in a series of experiments I will describe how we can and cannot shift their balance through neurostimulation and pharmacology (chapters 5 to 7). I then ask how the value representations in corticostriatal loops, as described in section 2.2.2.4.3.3, can be derived from anatomical measurement of corticostriatal white matter connectivity in healthy humans (chapter 8).

2.3 Response inhibition

2.3.1 General overview

In the previous section I touched on various decision-making strategies an animal might take to maximise its rewards. This approach assumes a fixed decision point after which consequences unfold irrevocably. In reality, an abstract decision is followed by motor preparation, action initiation and continuous monitoring and adjustment. Given an uncertain environment, this allows for radical changes to behaviour when conditions suddenly change. For example, the onset of a green light at a crossing might evoke a decision to start walking. But the sound of police sirens fast approaching can just at the last moment trigger a complete reprogramming of the action, even if it was already initiated.

In the second part of this thesis I explore how the brain rapidly inhibits actions when required by changes in the environment. I will specifically address the role of uncertainty and prior expectation on the behavioural and neural expression of inhibition. Deficiencies in inhibitory control have been implicated in neurological and psychiatric disorders (Verbruggen and Logan, 2008; Aron, 2011), perhaps most famously in attention-deficit/hyperactivity disorder (Barkley, 1997) and addiction (Ersche et al., 2012).

Inhibitory control is part of a broader field of self-control and impulsivity, which has been studied across many disciplines. Indeed, impulsivity is a catch-all for a wide range of behavioural phenomena in economics, psychology and psychiatry, each with its own tasks and models. For authoritative reviews I refer the reader elsewhere (Logan et al., 1997; Evenden, 1999; Whiteside and Lynam, 2001; Madden and Bickel, 2010; Bari and Robbins, 2013; Moeller et al., 2014). In this section I go into some background regarding one facet of impulsivity, namely the ability to withhold one’s action after an intention has been formed, and how proactive control for this type of inhibition alters its execution.

2.3.2 Reactive response inhibition

2.3.2.1 The stop-signal task

Response inhibition tasks are among the most common in cognitive science, stretching back as far as the go/no-go task developed by Donders (1868). In this task, most trials consisted of a go cue requiring a button press; the other trials contain a no-go cue indicating nothing should be done. The simple expectation that a response is required will lead to errors of commission on no-go trials, indicating a failure of inhibition (Bari and Robbins, 2013). A more challenging version of this task is the stop-signal task (Lappin and Eriksen, 1966). As popularised by Logan et al. (1984), every trial contains a go cue, but on a subset of trials this cue is quickly followed by a stop cue—the psychologist’s equivalent of a police siren just before stepping out onto the road. By adjusting the delay between go and stop signal (stop-signal delay, SSD) to approximate successful inhibition on 50% of trials, a simple calculation yields the stop-signal reaction time (SSRT, Figure 2.9; Logan, 1994; Band et al., 2003; Verbruggen and Logan, 2009b). This measure is used to reflect the number of milliseconds it takes to stop an action after onset of the stop signal; much like a reaction time indicates the number of milliseconds it takes to respond to the onset of a cue. Its intuitive interpretation, ease of administration and calculation has made it a staple in inhibition research. In chapters 9 and 10

I go into more detail regarding some basic predictions made by this model, and show that data from the modified stop-signal tasks satisfies these predictions.

An aim in this thesis is to understand how an inhibitory ability changes with uncertainty about the environment. I manipulated this by adding components of selectivity and preparation to the task, which are attributes of response inhibition that have come to the fore only in the past decade (Vink et al., 2005; Coxon et al., 2007; Chikazoe et al., 2009). I will review recent behavioural work on proactive, selective inhibition and discuss its potential neural correlates.

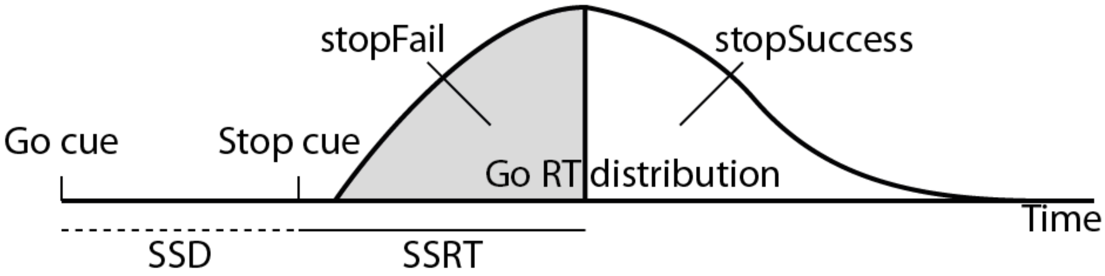


Figure 2.9: Rationale behind the stop-signal reaction time. The independent horse race model assumes the inhibitory process is of fixed duration; if it is initiated early enough (short SSD) and the go response happens to be slow on that trial (long RT), the inhibition can catch up with the go process and prevent the action from occurring. I will call this a 'stopSuccess' trial. In contrast, if the inhibitory process is started late or the go response happens to be fast, the action will be executed before it can be inhibited, resulting in a 'stopFail' trial. The duration of the SSRT is estimated by subtracting the mean SSD from the RT at the intersection of stopSuccess and stopFail trials, set such that the area-under-curve for stopSuccess is equivalent to the actual proportion of stopSuccess trials.

2.3.3 Extending the framework: proactive and selective inhibition

2.3.3.1 Selectivity of inhibition

The classic stop-signal task involves only a single button press. This leaves little room to study whether inhibition is an action-specific process or a ‘global’ stop signal that temporarily shuts down all motor action—or in the example of stepping out onto the road, does sudden inhibition of walking also inhibit unrelated actions such as speaking or typing on a phone? Strictly selective action inhibition is an ability to stop a single action without interfering with other ongoing actions (Aron, 2011). Unfortunately selectivity comes at a cost: it is slower compared to stopping all responses (Coxon et al., 2007) and creates interference with ongoing actions (Coxon et al., 2007, 2009; Greenhouse et al., 2012). One explanation for the interference effect is that upon inhibition, the subthalamic nucleus (STN) drives the entire motor loop of the GPi, in turn driving widespread inhibition across motor cortex (Schall and Godlove, 2012; Schmidt et al., 2013). Some of the most convincing evidence for this notion of global inhibition comes from studies showing that inhibition of a single finger response reduces the excitability of leg areas of motor cortex (as measured by motor-evoked potentials; Badry et al., 2009; Greenhouse et al., 2012). Intriguingly, another study showed *reduced* leg suppression when the participant is warned about which specific response might require inhibition, suggesting that preparation might play a critical role in selective targeting of actions in motor cortex (Majid et al., 2012). I therefore now turn to the topic of proactive inhibition, or the role of expectation and preparation in action control.

2.3.3.2 Proactive inhibition

Prediction and expectation play a central role in neuroscience (e.g. Rao and Ballard, 1999; Friston et al., 2006), and certainly so in inhibitory action control. Many studies have manipulated expectations by changing the stop-signal probability, most commonly by cueing people about the relative probability a stop-signal might occur on a given trial (Chikazoe et al., 2009; Verbruggen and Logan, 2009a; Jahfari et al., 2010; Swann et al., 2011; Zandbelt et al., 2012). This manipulation reliably leads to a slowing of go RT for higher stop probabilities, suggesting a strategic adjustment to increase chances of stopping successfully. The evidence for effects of stop-signal probability on the actual speed of the inhibitory process after correcting for this slowing—the SSRT—is more ambiguous: although one study observed faster SSRT in the high-probability condition (Chikazoe et al., 2009), this has not held up in further studies (Jahfari et al., 2012; Zandbelt et al., 2012). A component of the slowing of responses seems to be suppression of the corresponding motor representation in motor cortex, measured as a reduction in motor-evoked potential magnitude after cueing and before action execution (Claffey et al., 2010; Cai et al., 2011a). This in itself seems to be driven by proactive recruitment of the entire fronto-basal ganglia network as well as parietal cortex, a set of regions also involved in outright response inhibition (Chikazoe et al., 2009; Zandbelt and Vink, 2010; Jahfari et al., 2011; Jahfari et al., 2012; Zandbelt et al., 2012).

2.3.3.3 Preparing for selective inhibition

From the previous sections it should be clear that the response inhibition field has slowly but surely inched towards more ecologically interesting forms of self-control. It is not often that a drastic, unprepared inhibition of all motor action is required, and it could be argued that more subtle forms of self-control are more closely related to disorders of inhibition (Aron, 2011; Schall and Godlove, 2012). For example, withholding oneself from reaching out to the cookie jar upon passing by surely does not involve widespread and global inhibition—if it did we would freeze in place on every occasion. More likely this falls in a scenario whereby, even before seeing the cookie jar, there is selective suppression of the specific action of reaching out for the jar. When considered this way, the process of proactive control is not dissimilar to decision making, as both involve the control over actions so as to optimise the long-run benefit. Perhaps unsurprisingly, the first forays into this topic have suggested that proactive inhibitory control engages many structures we know from decision making: a frontostriatal circuit that involves both the associative and motor loops (Aron, 2011; Majid et al., 2013). In particular the action channels in the basal ganglia described in section 2.2.2.4.2 are hypothesised to serve as an efficient substrate for selective inhibition, compared to global inhibition via the cortico-subthalamic nucleus hyperdirect pathway (Aron, 2011).

2.3.4 Thesis work addressing open questions in inhibitory action control

In chapters 9 and 10 I use a selective stop-signal task and manipulate the amount of information that is available for proactive control. This allows the examination at the behavioural level how preparation affects both the speed and selectivity of inhibition; at the neural level, I can ask what prefrontal and sub-cortical brain regions mediate the improvements in behaviour seen with preparation. In a second study I modify the same task to work on a smartphone, allowing the collection of data from tens of thousands of participants. The key question here is to understand how ageing, which has a particularly pronounced detrimental effect in the frontal cortex, affects the ability to exert reactive and proactive control.

3 Methods

3.1 Physics of MRI

3.1.1 Protons in a magnetic field

The single proton present in a hydrogen atom has a quantum property called spin (Figure 3.1). Moving any such proton into a magnetic field, such as the Earth's field or the field of a magnetic resonance imaging (MRI) scanner, causes a proportion of the spins to align to the field in one of two states with splitting energy ΔE: a low energy spin-up state (parallel to the field), or a high energy spin-down state (anti-parallel to the field). Depending on the strength of the field B_0 and the magnetic moment of the atom, a slight majority of spins will be in the spin-up state. In addition to the net magnetization parallel to the field the hydrogen protons 'wobble' or precess around the field direction at a speed termed the Larmor frequency ω. These concepts relate to each other following according to the following equations:

$$\Delta E = \hbar \frac{\omega}{2\pi}$$

$$\omega = \frac{\gamma}{2\pi} B_0$$

Where ħ is the reduced Planck constant (in J s) and $\frac{\gamma}{2\pi}$ the gyromagnetic ratio (in Hz T^{-1}) determined by the composition of the nucleus. Fortunately, the abundant ^{1}H protons have a relatively large gyromagnetic ratio of 42.576 MHz T^{-1}, leading to a slight majority of 50.000013% of protons aligning parallel to the field at 3 T and 37 °C (Huettel et al., 2004).

$$\frac{P_{parallel}}{P_{anti-parallel}} = e^{\frac{\Delta E}{k_B T}}$$

It is this majority that allows us to perform MRI, as I will describe shortly. It is also this principle that has driven people to use stronger B_0 fields so as to increase the ratio of spin-up to spin-down protons. All studies reported in this thesis used 3 T.

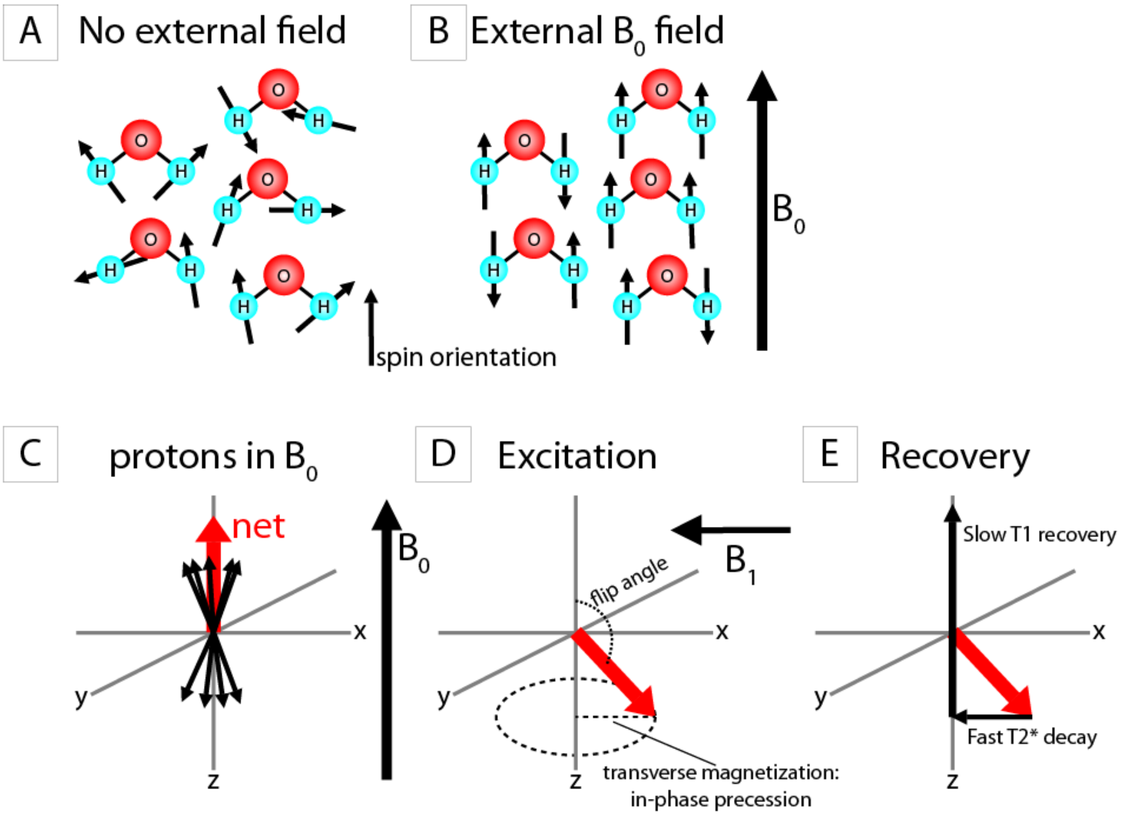


Figure 3.1: Magnetization of H^+ in H_2O as a function of magnetic fields. (A) In the absence of a strong magnetic field the spin of protons is randomly oriented. (B) The application of an external field B_0, as applied in MRI to protons in the brain, causes spins to align spin-up or spin-down. (C) As a slight majority of protons aligns itself spin-up, but none of the precessions are in phase, the net magnetization of protons is aligned straight along the B_0 field. (D) The additional energy delivered through a B_1 pulse at the Larmor frequency flips some spin-up protons in spin-down state, removing or even inverting the longitudinal magnetization. At the same time, this pulse brings the precession into phase, yielding transverse magnetization. This can be picked up by sensors around the head. (E) The transverse magnetization is quickly lost due to field

inhomogeneities and spin-spin interactions, whereas the longitudinal magnetization decays only slowly due to spin-lattice interactions.

3.1.2 Manipulating net magnetization

This $P_{parallel} > P_{anti-parallel}$ state can be described by a net magnetization vector along the z-axis in three-dimensional Cartesian space, where the z-axis is aligned with the main magnetic field B_0 of the system (Figure 3.1). The net magnetization vector can be manipulated though radio-frequency (RF, also called B_1 field) pulses that match the Larmor frequency and applied orthogonal to the B_0 field. The pulse firstly brings the spins into phase, leading to transverse magnetization in the x-y plane, and secondly reduces the longitudinal magnetization along the z-axis by exciting spins into the anti-parallel state. The duration of the RF pulse determines the flip angle of this vector away from the Z-axis, which in all experiments reported here was 90 degrees. The signal that is measured by magnetic resonance imaging is either the transverse magnetization, which decays in a matter of tens milliseconds following the RF pulse with a time constant termed T2*, or the longitudinal (z-axis) magnetization which recovers in a matter of hundreds of milliseconds with a time constant termed T1. Crucially, the T2* signal depends on the speed of dephasing of the spins. In fMRI for example it is possible to observe slower dephasing (i.e. an increase in T2* signal) due to a decrease in concentration of paramagnetic deoxyhaemoglobin (further discussed in section 3.2.1). In contrast, the T1 signal is determined by the spin-lattice interactions and can for example be used to distinguish tissue types such as grey and white matter. By adjusting the echo time (TE; readout time following the RF pulse) and repetition time (TR; time between two RF pulses on the same voxel) the signal becomes dominated by T1 and T2* contributions.

3.1.3 Building a 3-dimensional image

How can these physical concepts be exploited to obtain an image of the human brain? Whereas nuclear magnetic resonance (NMR) has been used since late 1940 to measure non-spatial properties of molecules in a solution, it was in 1976 that Paul Lauterbur and Sir Peter Mansfield independently realised that a sample, such as the brain, could be spatially dissected into slices and ultimately voxels (3-dimensional volumes arranged in a grid) using magnetic gradients on top of the static B_0 field (Figure 3.2A; Lauterbur, 1973; Mansfield, 1977). When the gradients are applied along the direction of the B_0 field, for example, the Larmor frequency of protons in superior parts of the brain (e.g. motor cortex) will be different from those in inferior parts of the brain (e.g. temporal lobe). A narrow-band RF pulse will then only excite the slice of the brain that matches the frequency of the RF pulse (Figure 3.2B). After excitation of the slice, an additional ‘frequency-encoding’ gradient can shift the Larmor frequency along one dimension of the slice, such that a Fourier decomposition of the signal is equivalent to a spatial decomposition. A further phase-encoding gradient allows for the second dimension of the slice to be isolated (Figure 3.2C). As such, a 3D volume is usually constructed from sequentially acquired 2D slices. In chapters 8 and 9 I used a more recent form of imaging for my functional acquisitions, acquiring data across 3 dimensions simultaneously (Pykett et al., 1982; Papanikolaou and Karampekios, 2008). The benefit of this approach is faster acquisition of high-resolution data and a higher signal-to-noise ratio per unit of time (Lutti et al., 2013). These sequences were implemented on Siemens Magnetom TIM Trio hardware (Siemens Healthcare, Erlangen, Germany)

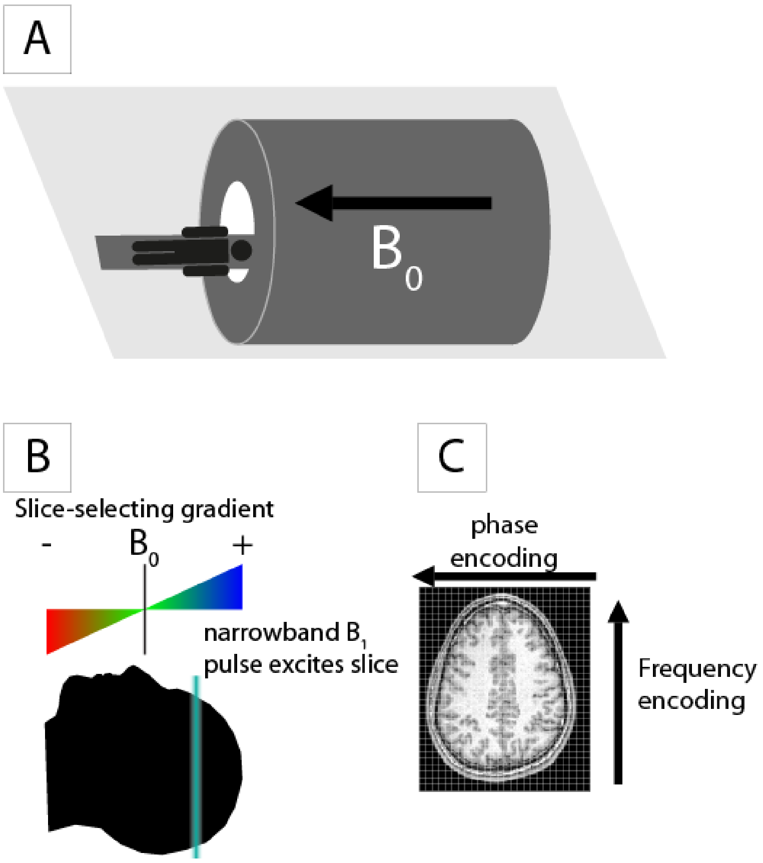


Figure 3.2: Using MRI to build a 3D volume of tissue. (A) An MRI scanner contains a permanent B_0 field. As tissue is moved into the scanner, the protons align into spin-up and spin-down states. (B) The Larmor frequency is approximately homogenous across the scanner, such that a B_1 pulse would excite all the protons. A slice-selecting gradient is applied at the time of the B_1 pulse to briefly differentiate the Larmor frequencies across the gradient. Only protons within the slab of tissue with the precession frequency of B_1 are excited. (C) Signal from within the selected slab of tissue is further divided into voxels using frequency and phase encoding. Note that in 3D imaging this process is slightly different, as slices are not measures sequentially but rather simultaneously.

3.2 Functional MRI

3.2.1 Basis of the BOLD signal

Functional magnetic resonance imaging (fMRI) provides a measure of neural activity in the brain. The three main currencies of neural activity are action potentials passed down axons of neurons, post-synaptic potentials across dendrites of neurons, and molecular signals that bind the former two across synapses. The fMRI signal is most closely associated with activity at the synapse and in particular the resulting post-synaptic potentials, which generate changes in blood flow measurable in fMRI. I will discuss first the effect of neural activity on blood flow, and secondly how such a change in flow is measured in fMRI.

The brain is permeated by capillaries and arterioles that supply blood strictly based on demand: neurons and supporting astrocytes regulate local blood flow according to the energetic demands of the cells (Attwell et al., 2010). Specifically, over 80% of change in blood flow originates from pericytes contracting and relaxing around capillaries in response to chemical signals from nearby neural tissue (Hall et al., 2014). These signals are mediated by ions and small molecules such as nitric oxide, K^+, adenosine, CO_2 and arachidonic acid metabolites, all by-products of metabolism around the synapse, as well as glial-mediated feedforward signals that increase blood flow before metabolites reach the vascular system (Attwell and Iadecola, 2002; Haydon and Carmignoto, 2006; Iadecola and Nedergaard, 2007), and can result from glutamatergic signalling in both cortex and sub-cortex (Sloan et al., 2010). Taken together, neural activity leads to an ‘opening of the gates’ which floods the local tissue with oxygenated blood, washing away deoxyhaemoglobin ($Hb\text{-}dO_2$).

Simultaneous recordings have shown that the dominant driver of such changes in blood flow is post-synaptic activity (i.e. incoming information) rather than action potentials (i.e. outgoing information) (Mathiesen et al., 1998; Logothetis et al., 2001).

Why are changes in blood flow in response to neural activity important? An increase in blood flow results in a decrease in Hb-dO_2 and an increase in Hb-O_2. It is the former that has an effect on the T2* signal observable in fMRI. The unbound Fe^{2+} in Hb-dO_2 is paramagnetic, causing strong local distortions in the B_0 field, and serves as a natural contrast agent (Ogawa et al., 1990; Ogawa et al., 1992). Any H^+ protons in H_2O near the Hb-dO_2 will thus precess at a different frequency than other protons, causing them to rapidly go out of phase. This speeds up the loss of transverse magnetization, i.e. T2* is shortened. Together, this suggests that neural activity, through an increase in blood flow and decrease in Hb-dO_2, will remove these local field distortions and thus increase the T2*-weighted signal. Hence the signal is named Blood Oxygen Level-Dependent (BOLD). It is worth noting that in response to neural activity there are two additional processes that both push Hb-dO_2 concentrations up: a relaxation of pericytes causes an increase in blood volume (increasing the quantity of Hb-dO_2), and an increase in oxygen use leads to more rapid dissociation of O_2 from Hb-dO_2. However, both these components are weaker than the effect of blood flow changes, leading to a net increase in T2* signal in response to neural activity.

3.2.2 Haemodynamic response function

I thus measure neural activity indirectly through the BOLD signal, so it is pertinent to understand how the two are linked in time and amplitude

(Logothetis, 2008). If I were to flash a bright light in the eyes of a participant being scanned in fMRI, causing strong glutamatergic firing in primary visual cortex 25 to 30 ms after (Schroeder et al., 1998), the idealised BOLD signal in V1 would evolve as described in Figure 3.3A (for empirical BOLD response see Logothetis et al., 2001). This is called the haemodynamic response function (HRF, Figure 3.3A). Three points are worth noting: firstly, it takes approximately 5 s for the BOLD signal to achieve its maximum intensity, and 12 to 20 s to return to baseline; secondly, if there are multiple instances of neuronal activity before BOLD has returned to baseline, they can be assumed to add linearly (Figure 3.3B; Boynton et al., 1996); thirdly, one can go back and forth between neural activity and BOLD signal by (de)convolving with the HRF, and this principle underpins statistical analysis of the BOLD signal (Boynton et al., 1996).

To summarise, fMRI exploits the fact that neural activity is tightly coupled to an increase in blood flow. This reduces the concentration of Hb-dO_2 which in turn increases T2* signal. This fortunate situation allows the indirect measurement of neural activity simultaneously across the whole brain with high (0.5 mm or higher) spatial resolution, albeit at the cost of poor temporal resolution. For the fMRI work presented in this thesis it was particularly important that fMRI provides access to subcortical structures such as the basal ganglia, an ability not afforded by techniques such as electroencephalography (EEG) or stimulation techniques (see section 3.6).

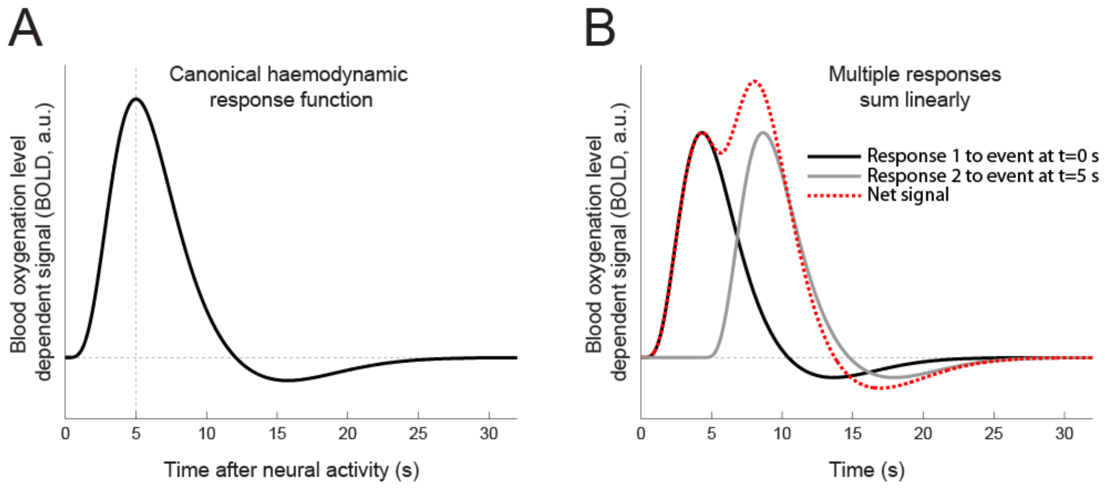


Figure 3.3: Neurovascular coupling leads to canonical blood flow response. (A) Canonical response in blood oxygenation level-dependent (BOLD) signal in response to a burst of neural input into a region at t = 0 s, as used in this thesis to predict fMRI signal. The signal peaks after approximately 5 seconds and returns to baseline over the course of 12 to 20 s. (B) Multiple events that happen before the signal has returned to baseline are assumed to simply add together.

3.2.3 fMRI preprocessing

The primary goal of preprocessing in fMRI is to transform individual volumes of BOLD signal such that a voxel at coordinates (x, y, z) in a single participant refers to the same piece of tissue across all volumes, and to the homologous part of the brain across participants. The only step that does not directly relate to the issue of localization is bias correction, which corrects for receiver coil properties affecting the intensity of the signal. For all fMRI preprocessing in this thesis I used the Statistical Parametric Mapping (SPM) software developed at the Wellcome Trust Centre for Neuroimaging, UCL, implemented in MatLab R2012a (The MathWorks, Inc.). All analyses were performed in SPM version 8 unless noted otherwise.

There are a number of potential issues with BOLD images as they come from the scanner which are alleviated in preprocessing: voxels closer to the receiver coil have a stronger signal than those further away from the receiver coils (bias correction); participants inevitably move their heads during e.g. 40 min of scanning (motion correction); the B_0 field is not perfectly homogenous due to the mechanics of the scanner and susceptibility gradients inside the head, leading to spatial misattribution of the signal during Fourier decomposition (unwarping); there are individual differences in the shape of the brain and location of the head in the field of view (coregistration, normalization and smoothing).

I note here that I did not perform slice time correction, a common step whereby the sequential acquisition of slices across the duration of the TR is corrected for. The fMRI data presented in chapters 8 and 9 were acquired through 3D imaging, which involves acquiring slices in K-space (spatial frequency space) rather than real space. As such, high spatial frequencies were acquired at the start and end of the TR, whereas low spatial frequencies were acquired around the centre of the TR. In other words, at any point in the TR information pertaining to the entire volume was being collected, and slice time correction cannot meaningfully be applied under these conditions.

3.2.3.1 Bias correction

BOLD-weighted images acquired on a 32-channel head coil from a 3 T scanner might look like Figure 3.4A, which shows an axial slice at 1.5 mm isotropic resolution with a restricted field of view. An artefact is readily observed whereby the edge of the brain has a stronger signal than the internal parts of the brain due to signal drop-off proportional to the distance from the sensors in the head coil (Figure 3.4A). This bias need not be corrected for statistical purposes as the mean of the signal is modelled independently for each voxel. Instead it aids subsequent preprocessing steps that rely on alignment and coregistration of volumes (e.g. normalization). The field is assumed to be smooth (60 mm full-width at half-maximum, FWHM), estimated for the first image in the series (Figure 3.4B), and subtracted from all images (Ashburner and Friston, 2005).

3.2.3.2 Motion correction

To correct for participant movement during the course of scanning I used the first image as reference scan, and for each subsequent image I estimated a 6 degrees-of-freedom (df) rigid-body transformation that minimises the discrepancy between images (Figure 3.4C; Andersson et al., 2001). The images were then resliced to obtain the aligned images and the transformation parameters recorded to enter into the statistical analysis as nuisance variables (Figure 3.4C, bottom; see section 3.2.4).

3.2.3.3 B_0 inhomogeneity correction

There exists a spatial inhomogeneity in the B_0 field due to imperfections in the magnets generating the field and due to air-tissue boundaries that generate local magnetic susceptibility artefacts. The latter primarily affects the inferior frontal lobe (due to air in the paranasal sinuses) and in the inferior temporal areas (due to air in the ear canal). This inhomogeneity can be measured using a ‘fieldmap’ estimated from the phase difference between the signal at a short and long TE (Andersson et al., 2001). All motion-corrected functional images are spatially ‘unwarped’ using the fieldmap (Figure 3.4D).

3.2.3.4 Coregistration of functional volumes to structural volume

The images corrected for B_0 inhomogeneities can be entered into a statistical model. However, it is often useful to first coregister the functional images to the structural image of the participant allowing for two analysis pathways both used in this thesis: firstly, the functional images can be brought into some standard space shared across participants to examine group-level effects (e.g. as in chapters 8 and 9), secondly the signal can be analysed in the space of the participant based on participant-specific regions of interest (ROIs) or diffusion data without the additional error introduced by normalization and smoothing. I performed coregistration by estimating a 6 df (rigid body) affine transformation matrix from the first functional scan to the magnetization transfer-weighted (MT) image of the same participant (Figure 3.4E). The affine transformation only involves translations and rotations as there is no need for scaling or shearing within-participant. I used the MT image here as well as during normalization because it has a higher contrast between white and grey matter, especially in subcortical structures. The transformation involved the maximization of normalised mutual information, measured as the sharpness of the 2D joint histogram of the transformed functional and structural image (Ashburner and Friston, 2005). A sharp histogram indicates that components of the signal in one image (e.g. the white matter component) can be predicted from the other, be it through a positive or negative relationship of any magnitude (see Figure 3.4E, bottom, for example of joint histogram before and after coregistration of two volumes; note that sharper edges after coregistration). The obtained transformation matrix is then applied to all B_0-corrected images.

3.2.3.5 Normalization of functional images

Every brain has its own characteristic folding of sulci and gyri, thickness of grey matter, size of subcortical nuclei, and any other number of morphological idiosyncrasies. Furthermore, every participant is placed in a slightly different location in the scanner and in the field of view of the image. In order to make inferences about the population from which the participants were drawn it is necessary to average functional signals from anatomically homologous brain regions across participants. One way to do this is by moving, rotating, scaling, shearing and warping the brain of a single participant to optimally fit a standard template. Using a standard rather than a group template has the additional advantage that a signal at coordinate (x, y, z) can be compared to coordinates from any other study that used the same standard space. This is particularly relevant for meta-analyses and the use of anatomical atlases such as the Anatomical Automatic Labeling (AAL) atlas used in chapter 9. The work presented in chapters 8 and 9 used Montreal Neurological Institute (MNI) space based on their ICBM152 maps (Mazziotta et al., 2001). The details of the normalization procedure are described in section 3.3.2 on preprocessing of structural images. The transformation parameters obtained through normalization of the MT image were then applied to all the functional images (Figure 3.4F).

3.2.3.6 Smoothing

At first thought smoothing functional images seems a waste of spatial resolution that physicists worked so hard to achieve. However, it is a necessary evil mainly for three reasons: firstly, part of the noise is independent across voxels, whereas the signal is mostly spread out over contiguous voxels, such that smoothing improves the signal-to-noise ratio by averaging out noise; secondly, the same information processing function might be located in a slightly different anatomical location in one participant compared to the other, such that even if normalization were perfect the signal would not overlap across participants; and thirdly, correcting for the multiple comparisons problem by Random Field Theory (RFT; see 3.2.4.2) requires a smoothness greater than that of unsmoothed fMRI data, such that not smoothing would lead to an excessively strict threshold for significance and a corresponding high false negative rate. Smoothing is applied using a 3D smoothing kernel with a full width at half maximum (FWHM) of 10 mm (chapter 9) or 6 mm (chapter 8; see Figure 3.4G). Ideally the size of the smoothing kernel is matched to the spatial extent of the hypothesised activation; in practice it is a function of the trade-off between spatial specificity and statistical power needed to overcome the stringent multiple comparisons correction associated with small smoothing kernels. That is, a well-powered study will be able to detect more focal signals albeit within the constraints of functio-anatomical heterogeneity across participants.

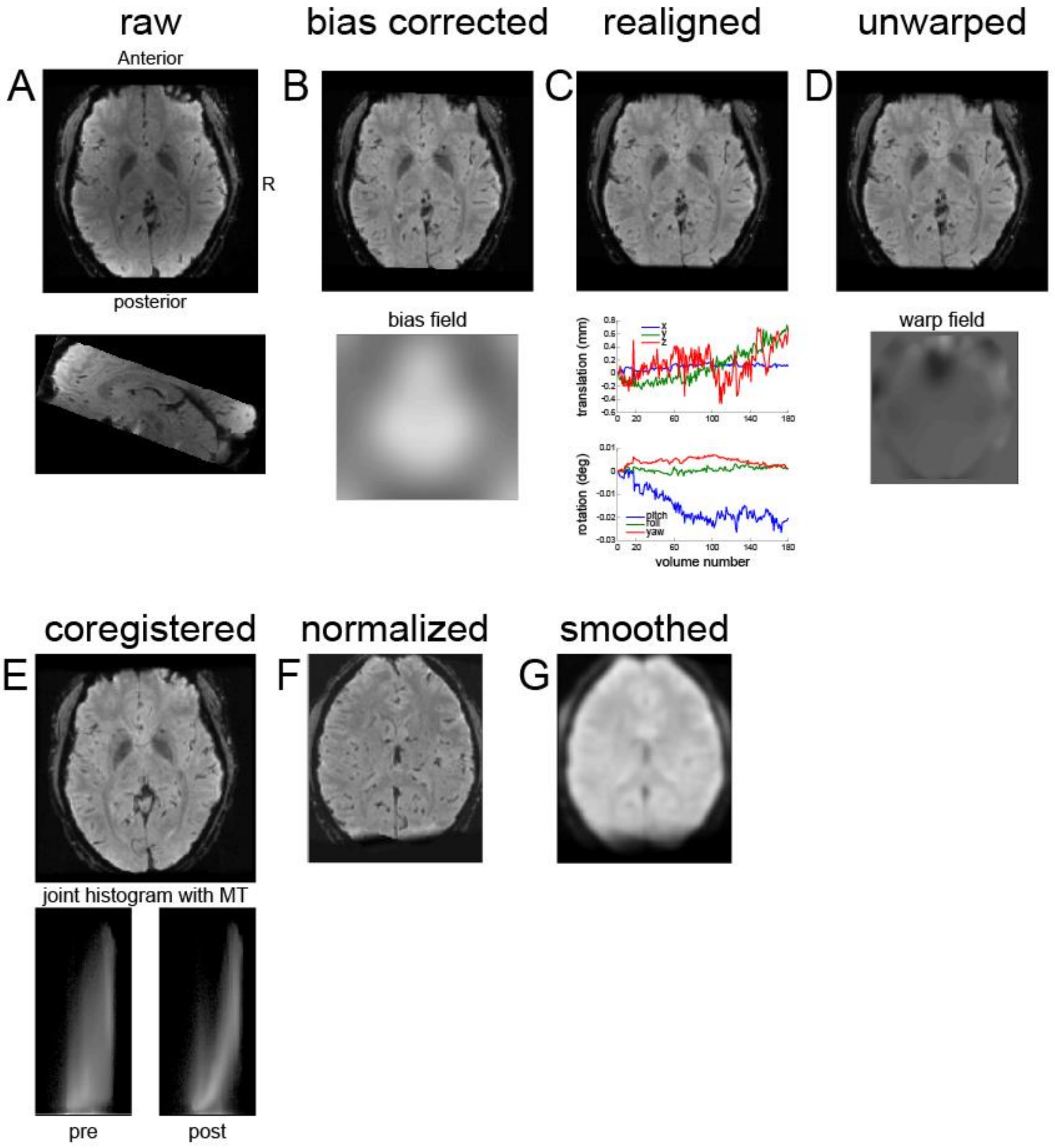


Figure 3.4: Preprocessing steps for fMRI data to prepare for statistical analysis. (A) Raw data shows a marked bias whereby the outer parts of the brain have greater signal amplitude compared to inner parts of the brain. This particular slice was from a single participant in chapter 8, which acquired data over a restricted volume. (B) Bias correction estimated a smooth bias field and generates a more homogenous signal intensity. (C) Realignment adjusts for participant movement, and the estimated movement parameters (bottom) are used in the statistics as described in section 3.2.4. (D) Fieldmaps, which estimate the imperfections in the static B_0 field, can correct distortions. (E) The functional data is coregistered to the participant's structural data by maximizing the normalised mutual information (bottom). (F) The structural volume is

normalised to the ICBM152 template using both linear and non-linear transformations. (G) The normalised functional images are smoothed to account for small errors during preprocessing as well as inter-individual differences in the precise location of functional anatomy. This image was smoothed at 6 mm full width at half maximum (FWHM).

3.2.4 Statistical analysis of fMRI images

3.2.4.1 General linear model

The goal of fMRI as used in chapters 8 and 9 was to associate aspects of cognition or action—such as action values or proactive inhibition—to changes in the BOLD signal. This is most commonly achieved through a mass univariate approach where each of the 100,000+ voxels is treated (at first) as an independent measurement. I use the general linear model (GLM) described as

$$y = x\beta + \varepsilon$$

where y is a vector containing the BOLD signal in a single voxel across all acquired volumes and x is the design matrix describing hypothesised causes of changes in y (Friston et al., 1994). A set of regression coefficients β is estimated using restricted maximum likelihood (ReML) to account for structure in the residual error ε (Glaser and Friston, 2004). This violation of independent and identically distributed errors arises from the sluggishness of the BOLD signal (Figure 3.3) and of cognition, such that the signal in volume n and $n+1$ are not independent. After estimating a single whitening matrix across all voxels, the set of coefficients β is estimated for each voxel separately, generating a statistical parametric map (SPM) for each regressor in x. Although in many studies there will only be a small number of regressors of interest, the design matrix contains many more regressors to account for as much variance in the signal as possible. Firstly, the design matrix contains regressors describing on-off events such as button presses, the appearance of a Go cue on the screen, or the onset of feedback. Secondly, optional parametric modulators on these events describe graded effects, such as reaction times on the button press, visual intensity of the Go cue, or the expected value at time of feedback. Thirdly, nuisance variables capture unwanted variance in the BOLD signal as well as possible confounds of the regressors of interest, such as head movements, breathing and heartrate. Lastly, dummy regressors account for baseline signal differences across volumes acquired over multiple runs. The scope here is limited to fast event-related designs as opposed to block designs (Friston et al., 1998), as I only used the former in the studies presented here.

The main regressors and parametric modulators are convolved by the haemodynamic response function (HRF), which links the hypothesised neural events in x to the BOLD signal in y (see Figure 3.3). Note that x remains identical across all voxels, which leads to the implicit assumption that the HRF is identical across all regions of the brain (but see e.g. Handwerker et al., 2004). The β at each voxel can be transformed into a t (or p) value to examine whether there is a significant relationship between the regressor and the BOLD signal at a particular voxel:

$$t = \frac{c\beta}{cSE_{\beta}}$$

where c is a contrast vector selecting specific regressors of interest from β. In addition to estimating SPMs for individual regressors across the brain (e.g. $c = [1\ 0]$ or $c = [0\ 1]$), contrast vectors can also be used to look for differences between regressors by adding and subtracting them (e.g. $c = [1 - 1]$).

3.2.4.2 Multiple comparisons problem

Most often we are not interested in SPMs of individuals, but rather in combining SPMs across individuals to describe the population at large where the participants were drawn from. In chapters 8 and 9 I created SPMs for contrasts of interest for each participant and performed one-sample t-tests over the β-maps from all participants. But how do we then decide what voxels show a significant response to the manipulation? This question continues to provide fertile grounds for debate across the neuroimaging community. The multiple comparisons problem arises from performing one t-test at every voxel, leading to many false positives if the false positive rate α is set to 0.05. We must therefore use a stricter α. However, as the SPMs arise from smoothed data, the t-tests are not independent and we should guard ourselves against an excessive false negative rate due to an insurmountable (e.g. Bonferroni-corrected) α. The solution used here and as implemented in SPM8 is Random Field Theory (RFT; Worsley and Friston, 1995). It uses an estimate of the smoothness of the data to calculate the expected Euler Characteristic (EC) for different thresholds. The EC is directly related to the number of clusters that would exceed threshold under the null hypothesis. If we want at most 5% of the SPMs to contain one or more false positive clusters we can set the threshold such that the EC is equal to 0.05. This properly controls for false positives while minimising false negatives within a single SPM. It should be clear that this approach does not correct for examining multiple SPMs originating from different contrasts, subsets of participants, pre-processing pipelines, design matrices, and other explorations of the data. Such practices would, and probably have (Button et al., 2013), lead to an abundance of false positives in the literature. It should thus be noted that multiple comparisons correction in fMRI is only useful insofar it is combined with sensible hypotheses and responsible analysis pipelines.

3.2.4.3 Regions of interest

A different way of alleviating the multiple comparisons problem is to use *a priori* regions of interest (ROIs). Rather than examining the entire brain, analysis is restricted to specific anatomically or functionally defined regions. In chapter 9 I averaged β-values from anatomical regions defined in Montreal Neurological Institute (MNI) space by the Automatic Anatomical Labeling (AAL) atlas (Tzourio-Mazoyer et al., 2002), and from probabilistic atlases available in the community (e.g. Keuken et al., 2014). I also used functional ROIs defined by clusters from contrasts. In chapter 8 I used anatomical ROIs that were defined in the native space of the participant rather than in standard space by means of automatic segmentation with FIRST as implemented in FSL 5.05 (Smith et al., 2004; Patenaude et al., 2011) and manual segmentation using ITK-SNAP 3.0 (Yushkevich et al., 2006). Participant-specific ROIs improve spatial specificity by accounting for inter-participant anatomical variability and obviating the need for normalization, which is inherently imperfect.

3.3 Multi-parameter mapping

3.3.1 Quantitative structural images

For each participant in chapters 8 and 9 I acquired structural multi-parameter maps (MPMs). In addition to the commonly used T1-weighted scans, the acquired volumes also include T2*-, proton density- and magnetization transfer-weighted volumes (Weiskopf and Helms, 2008). The MPM sequence and processing pipeline aspires to provide quantitative maps, meaning the signal describes some property of the tissue in the voxel independent of scanner type, head coil, and other extraneous factors (Weiskopf et al., 2013). This would allow meaningful aggregation of data across studies and research centres, be it through meta-analyses of the literature or more direct data-sharing. The work reported in this thesis does not leverage the quantitative property of the maps, so I refer the interested reader to other studies that examined this topic more closely (Draganski et al., 2011; Weiskopf et al., 2013; Callaghan et al., 2014). The goal in collecting the MPMs was to have magnetization transfer (MT) images with exceptional grey/white-matter contrast for normalization and sub-cortical segmentation, as well as T2* images for visualization of the internal and external parts of the globus pallidus, substantia nigra, and subthalamic nucleus during manual segmentation.

Acquisition of the MPMs involves B_1 mapping (i.e. mapping the RF field, Lutti et al., 2012), B_0 mapping identical to field maps used with functional scans (see section 3.2.3.3), and the acquisition of multiple echoes for each of a T1-weighted volume, a MT volume and a proton density (PD) weighted volume (Helms et al., 2008a; Helms and Dechent, 2009). All volumes were simultaneously processed in the voxel-based quantification (VBQ) toolbox implemented in SPM8, which models the tissue properties and outputs the quantitative maps (Callaghan et al., 2014).

3.3.2 Normalization

In order to make inferences at the population level it is often necessary to normalise functional images to some standard space. Although the functional images could be directly normalised to a standard template, it is more accurate to do so via a high-resolution structural image of the participant (Ashburner and

Friston, 2005). In chapters 8 and 9 I use Montreal Neurological Institute (MNI) space to normalise the MT images. The normalization procedure, as implemented in SPM8 (Ashburner and Friston, 2005), segments the MT image into grey matter, white matter and cerebrospinal fluid based on a set of tissue probability maps (TPMs) in MNI space. During this process a set of non-linear distortions as well as a full affine (12 df) transformation matrix is estimated that optimally fits the TPMs to the MT image. This mapping is then applied to the functional images to bring them into a common space.

3.4 Semi-automated segmentation of subcortical structures

An alternative to performing normalization to allow population-level inference is to produce summary statistics for each participant in their own space. For example, in chapter 9 I use regions of interest (ROIs) defined in the structural space of the participant to summarise functional activation in a particular region, and t-tests over these average activations are used to test for significant effects. The critical difference between working in standard space versus structural space is the automated versus manual identification of anatomical structures, respectively. That is, during normalization we relinquish responsibility for matching the idiosyncrasies of an individual brain to a template to an algorithm that warps each brain into standard space (see section 3.3.2); when using ROIs in the participant's structural space we define the ROIs by hand, or semi-automatically by having an algorithm make a first best guess which can then be checked and manually refined. Therefore, if algorithms existed that almost perfectly normalised images into structural space there would be no need for manual segmentation of regions, but such reliable automated identification is as of yet impossible for structures like the substantia nigra, subthalamic nucleus and internal and external parts of the globus pallidus. As these regions were of interest in chapter 8, I used a combination of automated and manual segmentation to obtain ROIs for each participant in their structural space.

The definition of the ROIs followed three steps. First, I used the toolbox FIRST as implemented in FMRIB Software Library (FSL) 5.0.5 to generate ROIs of the putamen, caudate, pallidum and nucleus accumbens for each participant in their structural space (Patenaude et al., 2011). Second, I loaded these segmentations into ITK-SNAP 3.0 (www.itksnap.org, Yushkevich et al., 2006) visualised onto the participant's MT and R2* image (Figure 3.5). I adjusted inaccuracies in the FIRST-based segmentations by hand. The combination of MT and R2* images allowed us to manually delineate the internal and external parts of the globus pallidus based on the medial medullary lamina visible on both the MT and R2* image (Figure 3.5A); the substantia nigra based on a strong contrast in the MT image with the surrounding tissue in the brain stem (Figure 3.5B); and the subthalamic nucleus based on a strong R2* contrast and a MT signal that differs from the adjacent substantia nigra (Figure 3.5B; Forstmann et al., 2012; Lambert et al., 2012). I further delineated the red nucleus to aid identification of the substantia nigra and subthalamic nucleus. In a last step I calculated the volume of each ROI and plotted a range of histograms (including volumes of single regions, right/left ratios, and between-ROI volume ratios, all across participants) to detect outliers. These were re-examined in ITK-SNAP and adjusted if necessary. The observed values were further compared to the literature, which showed that the volumes were within the expected range. Lastly, to visualise the segmented structures at the group level, I normalised the structures to MNI space and generated group probability maps by taking the mean at each voxel across participants for each mask. These were thresholded at 0.26 such that only voxels positive for 7 or more participants were retained in the mask. These normalised masks were not used for analysis, but rather for visualization.

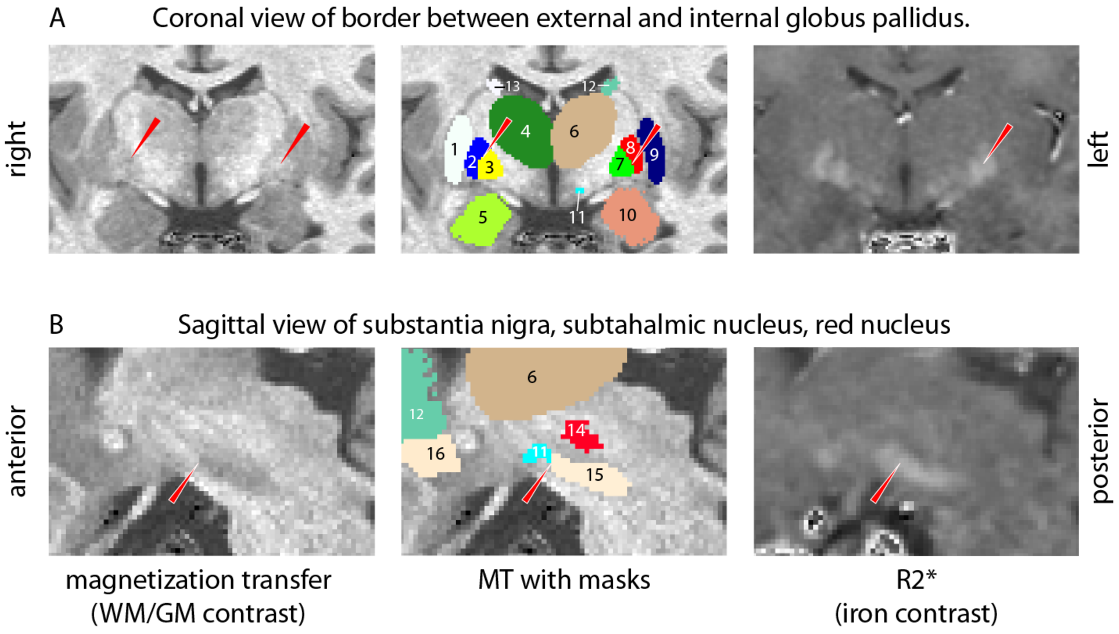

A
Coronal view of border between external and internal globus pallidus.
right
left
1
2
3
4
5
6
7
8
9
10
11
12
13
B
Sagittal view of substantia nigra, subtahalmic nucleus, red nucleus
anterior
posterior
6
12
14
16
11
15
magnetization transfer
(WM/GM contrast)
MT with masks
R2*
(iron contrast)

Figure 3.5: A single participant example of the use of magnetization transfer (MT) and R2 maps for automated and manual segmentation. (A) FSL FIRST segmented putamen, thalamus, hippocampus, caudate and pallidum from the T1w image. I then overlaid these segmentations onto the MT and R2* image in ITK-SNAP to, as a first step, split the pallidum into the internal and external part. The demarcation is the medial medullary lamina, indicated at the tip of the red arrows. On this slice the lamina is particularly well visible on the R2*, whereas on other slices the MT image showed a clear contrast. (B) On the sagittal view the substantia nigra can easily be visualised based on its dark colour in the MT image. The subthalamic nucleus is more challenging, and can be identified by its identical R2* intensity as the substantia nigra, but increased signal in the MT image. This border is indicated by the red arrow. 1: right putamen. 2: right GPe. 3: right GPi. 4: right thalamus. 5: right amygdala. 6: left thalamus. 7: left GPi. 8: left GPe. 9: left putamen. 10: left amygdala. 11: left subthalamic nucleus. 12: left*

caudate nucleus. 13: right caudate nucleus. 14: left red nucleus. 15: left substantia nigra. 16: left nucleus accumbens.

3.5 Diffusion MRI

3.5.1 Basis of the diffusion signal

Axons, the part of a neuron that carries spike trains towards synapses, bundle together to transfer information between parts of the central and peripheral nervous system. Where enough axons bundle together, the otherwise Brownian motion of water molecules is restricted orthogonal to the direction of the bundle, such that displacement along the principal direction of the bundle is, on average, larger than displacement orthogonal to it. In 1985 it was discovered such restrictions of random motion of water molecules are sufficiently large to pick up the direction of fibre bundles *in vivo* (Le Bihan and Breton, 1985; Le Bihan et al., 1986).

Diffusion-weighted imaging (DWI) fundamentally measures the loss of signal due to the movement of water molecules along a magnetic field gradient. This is perhaps best illustrated by an example: imagine a narrow tube filled with water, much like an axon bundle in the brain, in an MRI scanner. We can use an RF pulse combined with a slice-selecting gradient to excite a particular slice of the tube. Between the time of the RF pulse and readout of the transverse magnetization we could apply one of two magnetic gradients: one along the length of the tube, or one at a 90-degree angle to the tube. In the former case, water molecules will freely drift along the length of the tube, experiencing a varying magnetic field and thus de-phasing rapidly. This would result in a weak transverse magnetization signal at the time of readout. In the alternative case, with the gradient applied orthogonal to the direction of the tube, the water molecules are unable to diffuse along the direction of the gradient and will thus all experience the same magnetic field, leading to little de-phasing and a strong signal at readout time. If all we measure, then, is a weak signal for one gradient direction and a strong signal for another, we know that there is non-isotropy (i.e. non-uniformity across diffusion directions) in the sample. If we now tested another, say, 100 directions, we could more accurately estimate the orientation of the tube based on when the signal is strong and weak (Jones et al., 2013). In addition to the number of directions we can also vary the duration and strength of the gradient, together summarised by a 'b'-value, which further helps model the diffusion characteristics of the sample. We can then estimate the diffusion along a particular direction as described in the following equation:

$$\frac{S_d}{S_0} = e^{-bD}$$

where S_d is the signal measured with the diffusion gradient turned on; S_0 the signal without the diffusion gradient (i.e. $S_0 > S_d$); b is the combined strength and duration of the diffusion gradient in s mm^{-2}; and D is the variable of interest defined as the diffusion coefficient, representing the strength of diffusion. As S_d decreases with no changes in S_0 and b, D must therefore increase. That is, the loss of signal in the diffusion scan relative to the non-diffusion ('b$_0$') scan is inversely proportional to the log of the diffusion strength.

In chapter 8 I used a sequence modelled after the Human Connectome Project (HCP; Van Essen et al., 2013). I sampled across 100 gradient directions distributed over a sphere across three b-values (also called 'shells'; 900, 1800 and 2700 s mm^{-2}), along both right-left and left-right phase-encoding (PE) directions. The acquisition of both PE directions, also called 'blips', is necessary to correct for the strong distortions commonly observed in DWI. In chapter 8 I further acquired b$_0$ images along the posterior-anterior and anterior-posterior

PE directions to inform the reconstruction of the original signal from the PE-distorted images.

3.5.2 Preprocessing of diffusion images

I analysed the raw diffusion data (Figure 3.6A) using FSL’s diffusion toolbox to 1) estimate the distortion along the phase-encoding dimension from b_0 images (Figure 3.6B), 2) apply the corrections for these distortions and simultaneously correct for eddy currents and movement (Figure 3.6C), 3) estimate the diffusion tensors for each voxel to acquire fractional anisotropy maps (Figure 3.6D), and 4) estimate the distribution of diffusion parameters (Figure 3.6E) at each voxel to allow for probabilistic tractography (Figure 3.6F).

3.5.2.1 Correcting for phase-encoding distortions

The spin-echo sequence used in DWI is highly sensitive to off-resonance effects, such as magnetic susceptibility gradients caused by air in the ear canal or sinuses. Especially at higher echo times (TEs), as used in chapter 8, the signal around the inferior parts of the brain gets stretched and compressed along the PE direction due to these off-resonance effects (Figure 3.6A). Fortunately, flipping the PE direction also inverts the distortions, such that signal that was stretched along one PE direction is compressed along the opposing PE direction and vice versa (see Figure 3.6). FSL’s TOPUP function maximises the similarity between unwarped images by estimating the distortion field (Figure 3.6B), using the sum-of-squared differences between the unwarped images as goodness of fit.

3.5.2.2 Correcting for eddy currents and motion

The rapid switching of gradients in diffusion imaging leads to the induction of electric currents in components of the MR scanner, such as the headcoil. These currents themselves create small magnetic fields that affect the Larmor frequency of protons in the brain, in turn distorting the spatial reconstruction of the MR signal. The eddy currents become larger as we move from weaker to stronger gradients, e.g. from b=1000 s mm^{-2} to b=3000 s mm^{-2}. The artefacts are visible as contractions, shifts and shears, and additionally depend on the direction of the gradient. We thus need to model the distortions based on knowledge of the strength (b-value) and direction of the field. Additionally, head motion interacts with eddy currents, such that simple realignment of the head as done for functional MR images would not take into account the variable effect of eddy currents dependent on head location in the coil. I used FSL's function EDDY which uses a single model that incorporates TOPUP's field coefficients, motion, eddy currents, gradient strengths and gradient directions to correct each individual volume (Figure 3.6C).

3.5.3 Estimating the diffusion tensors

We can estimate a diffusion tensor D for each voxel that is described by 6 unique elements

$$D = \begin{bmatrix} D_{xx} & D_{xy} & D_{xz} \\ D_{yx} & D_{yy} & D_{yz} \\ D_{xz} & D_{yz} & D_{zz} \end{bmatrix}$$

where D_{xx}, D_{yy} and D_{zz} are the diffusion coefficients in the scanner's frame of reference, and D_{xy}, D_{xz} and D_{yz} reflect the correlations in displacement along the dimensions. These 6 elements can be estimated independently for each voxel by regressing the signal attenuation A in volume i on the strength and direction of the diffusion gradient associated with the same volume (Basser et al., 1994; Le Bihan et al., 2001):

$$\ln(A) = -\begin{bmatrix} D_{xx} \; D_{yy} \; D_{zz} \end{bmatrix} \begin{bmatrix} b_{xx} \\ b_{yy} \\ b_{zz} \end{bmatrix} - 2\begin{bmatrix} D_{xy} \; D_{xz} \; D_{yz} \end{bmatrix} \begin{bmatrix} b_{xy} \\ b_{xz} \\ b_{yz} \end{bmatrix}$$

$$\ln(A) = -b_{xx}D_{xx} - b_{yy}D_{yy} - b_{zz}D_{zz} - 2b_{xy}D_{xy} - 2b_{xz}D_{xz} - 2b_{yz}D_{yz}$$

Matrix D can be diagonalised, which yields eigenvalues λ and eigenvectors. The latter describes the three principal directions of diffusion, and λ describes the diffusivity for each direction independent of the other directions. The fractional anisotropy then quantifies the degree to which diffusion is different along x, y and z:

$$FA = \sqrt{\frac{1}{2}} * \frac{\sqrt{(\lambda_1 - \lambda_2)^2 + (\lambda_2 - \lambda_3)^2 + (\lambda_3 - \lambda_1)^2}}{\sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}}$$

Note that if $\lambda_1 = \lambda_2 = \lambda_3$ then the numerator and therefore FA equals zero, indicating there is no anisotropy. In the converse, e.g. $\lambda_1 \gg \lambda_2 = \lambda_3$, the FA will approach one, indicating all diffusion occurs along the direction described by the first eigenvector. Typical values for FA in grey matter are between 0 and 0.1, whereas FA in white matter can reach as high as 0.9 (Figure 3.6D). I used FSL's *dtifit* function to estimate the diffusion tensor.

3.5.4 Generating distributions for the diffusion parameters

Although FA maps are widely used to compare diffusion properties of tissue across individuals, the diffusion tensor cannot be used for probabilistic tractography. The reason is that the tensor characterises the principal diffusion direction, but does not contain a measure of the confidence that can be placed in this direction, i.e. it returns a point estimate rather than a distribution. However, probabilistic tractography as implemented in FSL works by sampling this distribution of principal diffusion directions. I used FSL's *bedpostx* to generate, using Markov Chain Monte Carlo sampling, the distributions of up to three principal directions per voxel (Behrens et al., 2003b; Behrens et al., 2007b; Sotiropoulos et al., 2011; Jbabdi et al., 2012). The peak of this distribution is shown for each of the three fibres in Figure 3.6E. These estimates are then used for probabilistic tractography (e.g. from the putamen, Figure 3.6F).

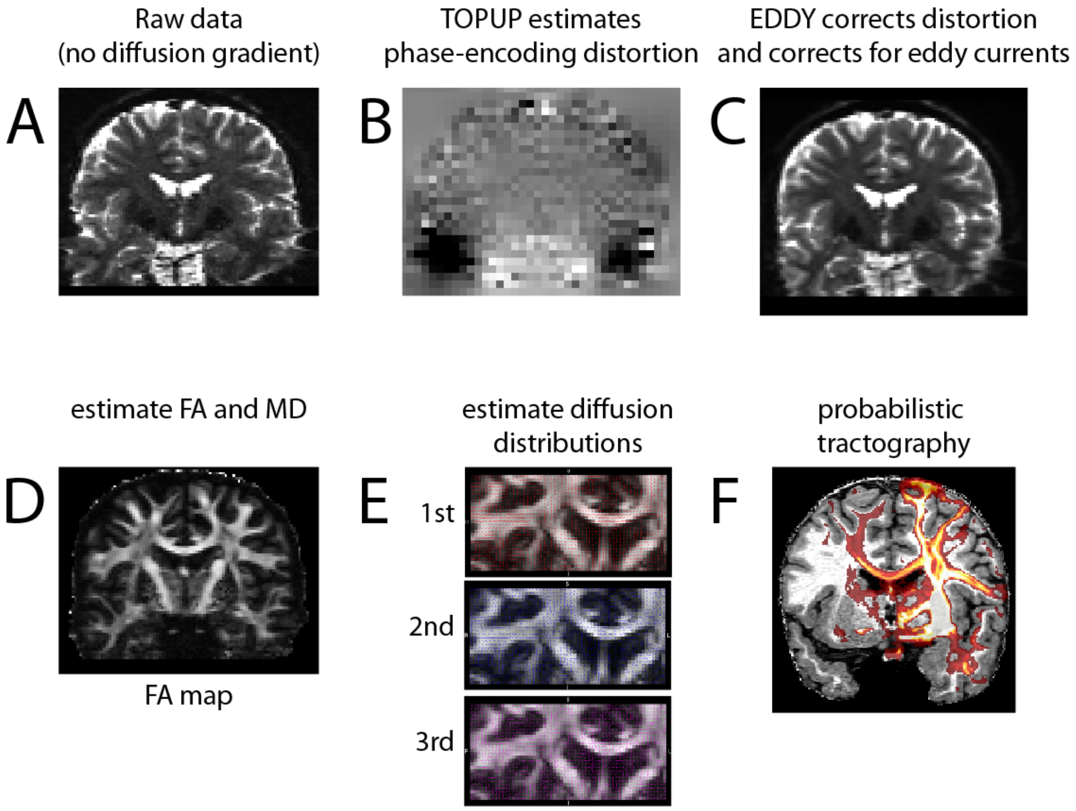


Figure 3.6: diffusion data preprocessing pipeline. (A) The raw data, here shown in the absence of a diffusion signal for best signal-to-noise ratio, shows distortions along the phase encoding direction (x) especially in ventral parts of the brain. (B) TOPUP estimates these distortions, generating a map of field coefficients. Note the black areas indicating strong distortion, as can be seen in A. (C) EDDY combines information from TOPUP with estimate of eddy currents and movement into a single model. The same slice as in A is shown here, but after correction by EDDY. Note the alleviated distortion in the ventral parts of the brain. (D) Using function dtifit I estimated the diffusion tensors that best

explained the data, without an estimate of the uncertainty. This yields fractional anisotropy (FA) and mean diffusivity (MD) maps, amongst others. (E) The eddy-corrected data also allows estimation of the distributions of fibre directions as implemented in bedpostx. Shown here are the peaks of these distributions, but critically, the entire distribution is estimated to allow for probabilistic tractography. (F) An example of probabilistic tractography, seeded from every voxel in the right putamen. Hotter colours indicate more streamlines passing through.

3.5.5 Probabilistic tractography

3.5.5.1 Rationale

The anatomical connectivity of the brain provides a blueprint for its function. Prior to the 21st century the only way of studying these connections was through post-mortem dissection and vivisection (as in ancient Greece by Herophilus, the first known anatomist; Bay and Bay, 2010) or white matter lesions and tracing studies in model organisms (Mesulam, 1978). None of these techniques can be used to study the relationship between anatomy and function in healthy humans. Furthermore, single-neuron tracing studies in non-human primates are painstaking (Markov et al., 2012), although recent developments promise more efficient tracing (Brainbow; Livet et al., 2007)

The development of diffusion MRI paved the way for in vivo tractography. By knowing the direction of fibres in each voxel of the brain it becomes possible to estimate what the likely white matter pathways are across multiple voxels. The characteristics of these pathways can then be linked to function to provide support for the widely assumed notion that connectivity underpins function in the central nervous system (Johansen-Berg et al., 2005; Neubert et al., 2010; Coxon et al., 2012; Saygin et al., 2012; e.g. Chowdhury et al., 2013).

3.5.5.2 Method

The first methods for constructing white matter pathways from diffusion data took a deterministic approach, reconstructing a single best guess based on the data (Mori et al., 1999; Basser et al., 2000). However, this approach did not reflect the often considerable uncertainty in the direction of fibres, leading to the introduction of probabilistic tractography (Behrens et al., 2003b). Here, thousands of 'streamlines' are drawn for each seed voxel, sampling from the fibre distribution from *bedpostx* at each subsequent step. Over many streamlines this builds up a map of the brain quantifying the number of streamlines that pass through each voxel. Further development of this method achieved more accurate tractography for multi-shell data (Behrens et al., 2007b). This approach has been validated against non-human primate data (Croxson et al., 2005; Jbabdi et al., 2013), suggesting that a reconstruction of white matter pathways in this stochastic framework can accurately capture even detailed characteristics of the anatomy.

3.5.5.3 Limitations

"It seems strange, therefore, that the connectivity of the brain depends on the parameters of the MR experiment." Derek K. Jones, 2013

An initial excitement with in vivo tractography has led, in some cases, to an over-interpretation of the data (Jones et al., 2013). There are many ways in which (probabilistic) tractography can run into problems and yield misleading results. For example, crossing white matter fibres often look isotropic, i.e. lacking directional diffusion, despite the underlying pathways. Solving this has been a major focus of developments in tractography from the start (Basser et al., 2000; Behrens et al., 2003b; Behrens et al., 2007b; Wedeen et al., 2008) and is still ongoing (Sotiropoulos et al., 2013). Another issue, also present in data in chapter 8, is strong distortion with longer acquisition times at high resolution along the phase encoding direction (Van Essen et al., 2012; Glasser et al., 2013). This can be particularly severe in ventral areas of the brain. Lastly, the probabilistic nature of the reconstruction makes it particularly hard to reconstruct long fibres. Taken together, the measurement of connectivity between two voxels directly depends on the quality and quantity of the MR data. For this reason, these values cannot directly represent connection probability or fibre density (Jones et al., 2013). However, by using the same MR sequence for each participant as well as relative measures of connectivity (e.g. compared to some participant-specific baseline, which accounts for MR quality) it is possible to get informative estimates of how reliably two regions are connected. Such estimates can then be related to individual differences in behaviour and functional data (chapter 8).

3.6 Neurostimulation techniques

In most of cognitive neuroscience our only way of manipulating neural activity is through the sensory channels. For example, we can cause a change in BOLD signal in occipital cortex by flashing a checkerboard pattern in the participant’s eyes, or in S1 by tactile stimulation of the skin. This contrasts with neurostimulation, as noted by Bestmann et al. (2008): “[Neurostimulation] can bypass the sensory pathways that provide the conventional alternative source of causal inputs.” By directly stimulating regions of interest we can study the *necessary* role for those regions in information processing, which is impossible in fMRI. Neurostimulation has been used in humans for over a century (Thompson, 1910), but especially the last three decades (Merton and Morton, 1980; Barker et al., 1985) have seen major advances in the protocols (Hoogendam et al., 2010) and widespread adoption in research and clinic (Kolbinger et al., 1995; Pascual-Leone et al., 2005; Rossi et al., 2009; Freitas et al., 2011). Here I used transcranial magnetic stimulation (TMS) to disrupt processing in a cortical area for tens of minutes (chapter 6; Huang et al., 2005) as well as transcranial direct current stimulation (tDCS) to increase excitability of the same region of cortex for a similar amount of time (chapter 7; Nitsche and Paulus, 2001; Nitsche et al., 2008). It remains largely unknown how these neurostimulation techniques exert their effects (though see Stagg and Nitsche, 2011; Stagg et al., 2013) and their efficacy is primarily based on changes in excitability of motor cortex as measured by motor-evoked potentials (MEPs, e.g. Penfield and Boldrey, 1937; Nitsche and Paulus, 2000). Indeed, a meta-analysis published recently suggests tDCS does not have any effect on cognition (Horvath et al., 2015). Additional methodological details regarding positioning, stimulation parameters and blinding can be found in chapters 6 and 7.

3.7 Oral levodopa to alter dopamine levels

Levodopa (L-DOPA) is transformed into the neuromodulator dopamine by decarboxylation inside dopaminergic cells. Critically, the carboxylic acid on L-DOPA allows it to cross the blood-brain barrier, unlike its derivative dopamine. Much of the development of this molecule as a drug was due to its remarkable benefit to Parkinson’s disease patients (Cotzias et al., 1969). In chapter 5 I used a modern version of the drug, which is co-administered with the peripheral decarboxylation inhibitor benserazide, to increase dopamine levels across all dopaminergic neurons in the brain (Everett and Borcherding, 1970). This has particularly pronounced effects in the striatum (Lloyd et al., 1975), but also directly or indirectly affects prefrontal function (Cools et al., 2002). The half-life of L-DOPA/benserazide is approximately 1.5 hours (Fabbrini et al., 1987), its concentration in the blood plasma peaking approximately 1 hour after oral administration, though it depends on factors such as last meal ingestion time (Baruzzi et al., 1987). The experiment in chapter 5 was performed double-blind with counterbalanced administration of placebo.

4 Empirical studies on model-based and model-free control

4.1 Introduction

An overarching view of adaptive behaviour is that humans and animals act to maximise reward and minimise punishment as a consequence of their choices. There are multiple ways this can be realised (see section 2.2) and mounting evidence indicates model-based and model-free forms of reinforcement learning (RL) contribute to behavioural control (Doya, 1999; Daw et al., 2005; Balleine and O'Doherty, 2010; Redgrave et al., 2010; Boureau and Dayan, 2011; Wunderlich et al., 2012b).

To study these two forms of control I used a task first developed by Daw et al. (2011). The defining feature of the task is that it has an associative structure that can be exploited by a model-based, but not a model-free, controller. The extent to which participants use this structure during choice is used as a measure of model-based control; conversely, the extent to which they ignore the structure during value updating is used as a measure of model-free control. Here I describe the task and analysis methods used in 3 experiments involving an L-DOPA manipulation (chapter 5), transcranial magnetic stimulation (chapter 6) and transcranial direct current stimulation (chapter 7).

4.2 Task

Each trial consisted of two stages, both requiring a choice between two stimuli. Each choice option was represented by a fractal in a coloured box on a black background (Figure 4.1). At every choice, participants had to respond within two seconds using the left/right cursor keys or the trial was aborted. Participants rarely missed a trial (e.g. in chapter 5, mean proportion of missed trials: 0.4%, SD: 1.5%), and those missed trials were omitted from analysis. Choice at the first stage always involved the same two stimuli left/right randomised. After participants made their response the rejected stimulus disappeared from the screen and the chosen stimulus moved to the top of the screen. After a delay (see Table 4.1 for experiment-specific settings) one of two second stage stimulus pairs appeared, with the transition from first to second stage following fixed transition probabilities. Each first stage option is more strongly (with a 70% transition probability) associated with one of the two second stage pairs, a crucial factor in allowing the dissociation of model-free from model-based behaviour. After the second choice, the chosen option remains on the screen, together with a reward symbol (a pound coin) or a ‘no reward’ symbol (a red cross). Each of the four stimuli in stage 2 had a reward probability between 0.2 and 0.8. Those reward probabilities drifted slowly and independently for each of the four second stage options in every trial through a diffusion process with Gaussian noise (mean 0, SD 0.025). The walks were not truly random as I selected a number of walks that would ensure that participants needed to keep learning and switching their preference for second-stage cues throughout the experiment (see section 4.5 for validation of these walks). For experiment-specific settings see Table 4.1.

Prior to the experiment participants were given explicit information about the task structure; namely that for each stimulus on the first stage one of the two transition probabilities was higher than the other, and that these transition probabilities remained constant throughout the experiment. Participants were also told that reward probabilities on the second stage were independent of each other and would change slowly over time. Before starting the testing session, participants practiced 50 trials with different stimuli and outcome probabilities.

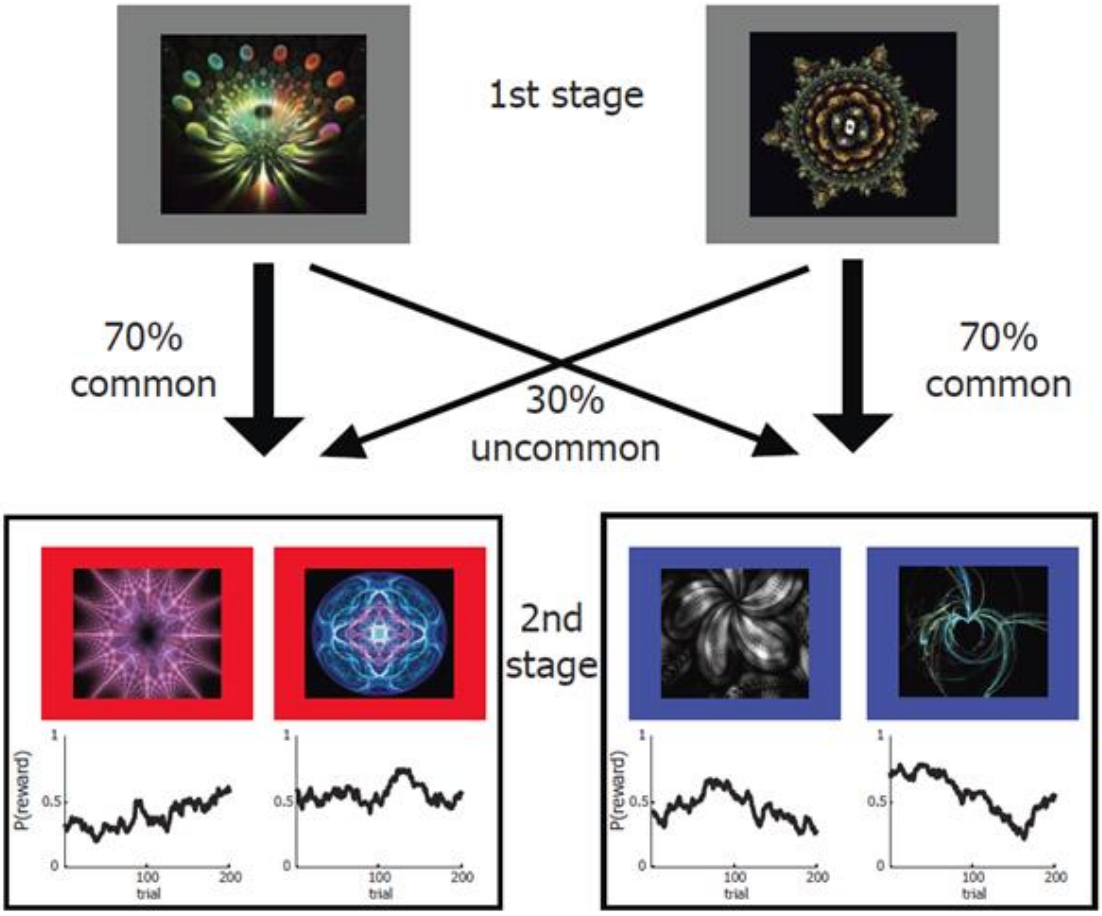


Figure 4.1: Two-step task design (A) On each trial a choice between two stimuli led probabilistically to one of two further pairs of stimuli, which then demanded another choice followed by reward or no-reward according to the p(reward) of the chosen second-stage stimulus that fluctuated over time. Importantly, participants could learn that each first-stage stimulus led more often (70/30%) to one of the pairs; this task structure could then be exploited by a model-based, but not by a model-free controller. (B) Model-based and model-free strategies for reinforcement learning predict differences in feedback processing particularly after uncommon transitions. If choices were exclusively model-free, then a reward would increase the likelihood of staying with the same stimulus on the next trial, regardless of the type of transition (left). Alternatively, if choices were driven by a model-based system, the impact of reward would interact with the transition type (middle). As shown previously behaviour in healthy participants resembles a hybrid of model-based and model-free control (right; Daw et al., 2011). I can thus quantify model-free control by estimating the

main effect of reward, and model-based control by estimating the reward-by-transition interaction.

Table 4.1: Experiment-specific settings. The timings in the TMS and tDCS experiment were faster to allow for more choices per unit time in the experiment. ITI = inter-trial interval; L-DOPA = levodopa; TMS = transcranial magnetic stimulation; tDCS = transcranial direct current stimulation; $U(1{,}3)$ *= uniform distribution between 1 and 3; nRewards = total number of rewarded trials in the session.*

Study	**Sessions**	**different sets of walks**	**double blind**	**settings per session**						
				performance-based pay-out (£)	**trials**	**# breaks**	**ITI (s)**	**time for choices (s)**	**transition time (s)**	**reward duration (s)**
L-DOPA (Ch. 5)	2 (on/off)	2	yes	0.25*(nRewards - 85)	201	2	$U(1{,}3)$	2.0	1.5	2.0
TMS (Ch. 6)	3 (vertex, left/right dlPFC)	3	no	0.25*(nRewards - 85)	201	2	$U(1{,}2)$	2.0	0.5	1.5
tDCS (Ch. 7)	2 (sham/active)	3	yes	0.2*(nRewards - 170)	350	4	$U(1{,}2)$	2.0	0.5	1.5

4.3 Basic analysis of behaviour

The logic of the task is based on the fact that a dependence on model-based or model-free strategies predicts different patterns through which feedback obtained after the second stage should impact future first stage choices (Daw et al., 2011). A model-free reinforcement learning strategy predicts a main effect of reward on stay probability. This is because model-free choice works without considering structure in the environment; hence rewarded choices are more likely to be repeated, regardless of whether that reward followed a common or rare transition. A reward after an uncommon transition would therefore adversely increase the value of the chosen first stage cue without updating the value of the unchosen cue. In contrast, under a model-based strategy I expect a crossover interaction between the two factors, because a rare transition inverts the effect of a subsequent reward (Figure 4.1). Under model-based control, receiving a reward after an uncommon transition increases the propensity to switch. This is because the rewarded second stage stimulus can be more reliably accessed by choosing the rejected first stage cue than by choosing the same cue again.

Thus, the influence of the controllers can be inferred in terms of the main effect of reward (model-free) and the interaction between reward and transition likelihood (model-based) on the probability of staying with the same first-stage stimulus on the next trial (as in Daw et al., 2011; Figure 4.1). This “1 trial back” analysis can be performed using a simple ANOVA over p(stay|reward,transition) with factors reward, transition and manipulation (e.g. L-DOPA or placebo in chapter 5), or in a logistic regression as in chapters 6 and 7. In the latter, the dependent variable is stay (1) or switch (0), and I used the Linear Mixed Effects

4 toolbox for R (R Development Core Team, 2008; R Core Team, 2011; Bates et al., 2012) to estimate population coefficients for each regressor. I could then use the *esticon* function in package “doBy” (Højsgaard, 2012) to test contrasts of interest. In chapter 7 I further extended this framework to examine influences of both controllers extending more than 1 trial back, by coding the dependent variable as “Chose A” (1) or “Chose B” (0), where A and B are the two first-stage stimuli. The regressors code for whether a model-free or model-based agent, based on the first-stage choice, reward and transition in trials prior to the choice, would promote choosing A (1) or B (-1; see Table 6.1 for a list of all regressors).

4.4 Reinforcement learning models

In the following I denote the model-free value Q_{s1}^{MF} and the model-based value Q_{s1}^{MB} for first stage stimuli $s1 \in \{1,2\}$. The hybrid model computes the actual value Q_{s1}^{Hybrid} used in determining choice as weighted linear combination

$$Q_{s1}^{Hybrid} = \omega * Q_{s1}^{MB} + (1 - \omega) * Q_{s1}^{MF}$$

where $\omega \in [0,1]$ quantifies the extent of model-based and model-free control. Values for the four stimuli at the second stage (stimuli $s2 \in \{3,4,5,6\}$) are updated identically for both controllers according to reward prediction errors (Watkins, 1989):

$$Q_{s2}(t+1) = Q_{s2}(t) + \alpha_2 * (r - Q_{s2}(t))$$

Where $\alpha_2 \in [0,1]$ is the learning rate, and $r \in \{0,1\}$ the absence or presence of a reward on trial t. At the first stage, model-free ‘cached’ values are updated according to on-policy temporal difference learning with reward prediction errors and eligibility traces (Rummery and Niranjan, 1994):

$$Q_{s1}^{MF}(t+1) = Q_{s1}^{MF}(t) + \alpha_1 * \left(Q_{s2\,chosen}(t) - Q_{s1}^{MF}(t)\right) + \lambda * \alpha_1 * \left(r - Q_{s2}^{MF}(t)\right)$$

Here $\alpha_1 \in [0,1]$ is the learning rate for the first stage, and $\lambda \in [0, +\infty)$ is a gain parameter representing the eligibility trace.

Model-based values are calculated anew for each and every trial in a forward looking manner by multiplying the state values of the better option at the second stage with the state transition probabilities:

$$Q_1^{MB} = 0.7 * \max(Q_3, Q_4) + 0.3 * \max(Q_5, Q_6)$$

$$Q_2^{MB} = 0.3 * \max(Q_3, Q_4) + 0.7 * \max(Q_5, Q_6)$$

Based on simulations by the authors of the original task I likewise simplified model-based learning under the premise that learning of state transitions quickly converges to stable values and hence transition probabilities were not updated by explicitly modelling state prediction errors (see supplemental material in Daw et al., 2011).

I computed the probability P of choosing stimulus 1 (in a choice between stimulus 1 with value Q_1 and stimulus 2 with value Q_2) at stage 1 according to a softmax choice rule, which depends on the relative stimulus values and choice in the previous trial (*C* = 1 if s1 was chosen on the previous trial, -1 if s2 was chosen)

$$P(s_1) = \frac{1}{1 + \mathrm{e}^{-\beta_1 * (Q_1 - Q_2) - \pi * C}}$$

and similarly in stage 2, e.g. when observing the s3-s4 pair:

$$P(s_3) = \frac{1}{1 + \mathrm{e}^{-\beta_2 * (Q_3 - Q_4)}}$$

The model thus includes 7 possible parameters to optimise over: two inverse temperatures β_1 and β_2, two learning rates α_1 and α_2, an eligibility trace λ, perseverance π, and a parameter ω for the relative degree of model-based versus model-free control. The learning rate α captures the extent to which new information at outcome is used for learning, i.e. the learning speed; β measures the discriminability between two options, with a larger value pertaining to more predictable choices (i.e. a more predictable link between values and choices); the persistency π is an index of the tendency to choose the same option as in the previous trial regardless of value (Lau and Glimcher, 2005; Kable and Glimcher, 2007), and parameter ω represents the extent to which one or other system drives a participant's behaviour. Reduced versions of the model were compared using model comparison techniques in chapter 5, and a 5-parameter model was used in section 4.5 to simulate data from the task

I applied logistic or exponential transformations before fitting parameters to transform bounded parameters into transformed space which spanned $[-\infty, +\infty]$ to accommodate maximum likelihood (ML) and expectation maximization (EM) estimation. I transformed α and ω using the logistic function

$$\alpha_{bounded} = \frac{1}{1 + e^{-\alpha_{unbounded}}}$$

$$\omega_{bounded} = \frac{1}{1 + e^{-\omega_{unbounded}}}$$

and β and λ using the exponential function

$$\beta_{bounded} = e^{\beta_{unbounded}}$$

$$\lambda_{bounded} = e^{\lambda_{unbounded}}$$

4.5 Validation of random walks

A purely model-free agent can, under restricted circumstances, generate data that contains a reward-by-transition interaction in the 1-back analysis described above (p(stay) analysis). This confound arises when the second-stage reward probabilities are relatively static, because the participant might settle on the best first-stage stimulus rather than switch between the two first-stage stimuli on a regular basis. This would then lead to a situation where most common transitions are rewarded, and most uncommon transitions are unrewarded, when choosing the best stimulus. In both cases, the participant likely stays with the same first-stage stimulus on the first trial, even if the participant is completely model-free, thus leading to a reward-by-transition interaction that is inferred as model-based control. This confound is quickly alleviated when the second-stage reward probabilities become less static, or indeed truly random. In chapter 6 I used 3 random walks randomly assigned to sessions. To confirm that these walks (and by extension the walks from my other studies, which were generated with similar levels of drift) are not confounded I generated data from a model-free agent playing on these random walks and examined the inferred levels of model-based and model-free control. The prediction was that data from a model-free agent should not show model-based characteristics.

First, I generated choices on the three walks using a reinforcement learning model identical to that used by Otto et al. (2013), which has 5 parameters: a model-free learning rate shared between the first and second stage (α), inverse temperature for the softmax choice rule (β), eligibility trace which carries the second-stage prediction error over to the first-stage stimuli (λ), a weighting parameter between model-based and model-free control (ω) and a perseverance parameter which accounts for a propensity to stay regardless of previous events (π). For details of the model see Otto et al. (2013). I selected representative values for all but α: $\beta = 4$, $\lambda = 0.6$, $\omega = 0$ (i.e. purely model-free), $\pi = 0.1$ (Daw et al., 2011). As the potential confound can depend on α, I generated data for α between 0.001 and 0.600 in steps of 0.001, simulating 3000 datasets of 201 trials for each configuration of parameters. I then calculated p(stay) for each of the four reward/transition conditions and calculated the magnitude of the main effect of reward, and reward-by-transition interaction from these p(stay) values. Note this is a different approach from the hierarchical analysis, but along identical lines of reasoning. Crucially, I predicted that a model-free agent should not show any reward-by-transition interaction in any of the walks, as that would indicate a potential confound in the walks. As expected, these walks did not show such a confound (Figure 4.2), such that the level of model-based control was close to zero for all learning rates. Interestingly, the level of inferred model-free control scaled close to linear with learning rate. In conclusion, in the random walks used in this thesis the estimation of model-based control is not confounded by model-free influences on behaviour.

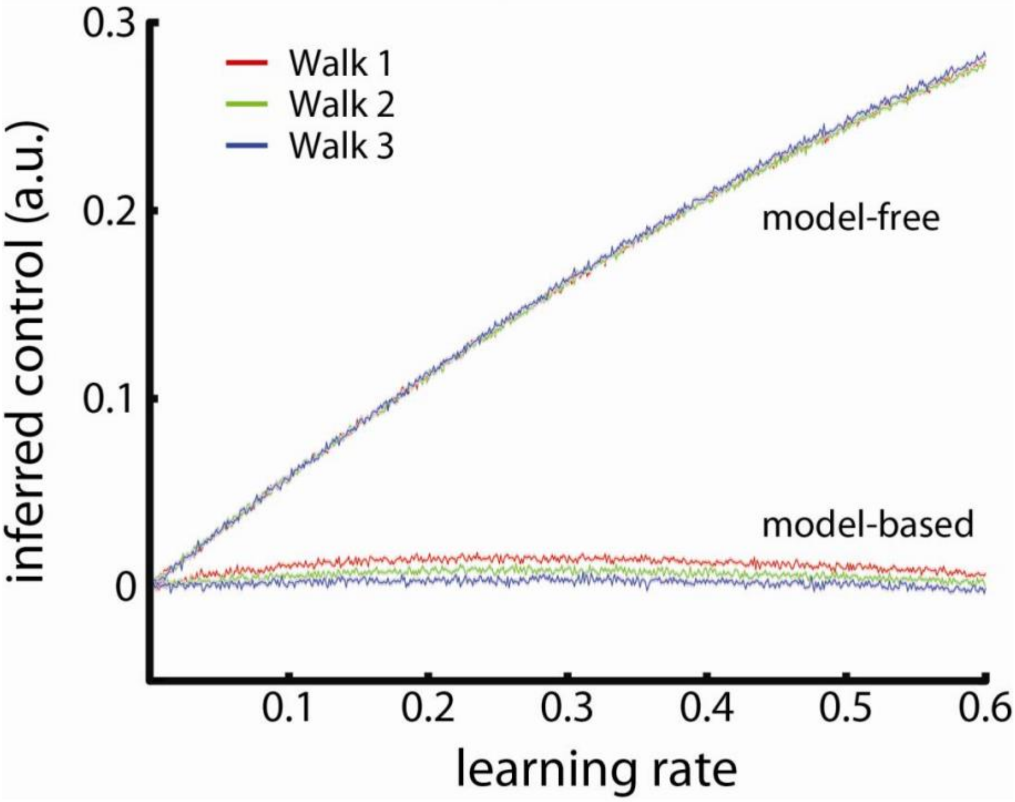


Figure 4.2: inferred level of model-based and model-free control in a purely model-free agent as a function of learning rate of this agent. I simulated 3000 agents playing the 2-step task for every learning rate to verify that the analysis method would not infer model-based control even though the underlying generative model was purely model-free. For all three walks used in chapter 6, the analysis correctly estimated model-based control to be around zero irrespective of the learning rate. Random walks in the other experiments had similar generative settings.

5 Dopamine enhances model-based over model-free choice behaviour

5.1 Abstract

Dopamine has been implicated in virtually all aspects of reward-guided learning and choice, including reward prediction errors, motivation and planning. Here, rather than studying the role of dopamine in one specific process, I asked how a dopaminergic challenge could alter the extent to which model-based and model-free decision-making strategies are used. I used the two-stage Markov decision task, as outlined in the introduction to these chapters, to quantify model-based and model-free control. I found that administration of L-DOPA promoted model-based over model-free choice, specifically by strengthening model-based control in response to the absence of reward. In contrast, I observed no effect of L-DOPA on model-free control.

5.2 Introduction

Previous research had focused on the role of dopamine in model-free learning, and value updating via reward prediction errors. For example, phasic firing of dopaminergic VTA neurons encodes reward prediction errors in reinforcement learning (Schultz et al., 1997; Hollerman and Schultz, 1998), and in humans, drugs enhancing dopaminergic function (e.g. L-DOPA) augment a striatal signal that expresses reward prediction errors during instrumental learning (Pessiglione et al., 2006). In so doing, L-DOPA increases the likelihood of choosing stimuli associated with greater monetary gains (Frank et al., 2004; Pessiglione et al., 2006; Bodi et al., 2009). However, dopamine’s role in model-based choice remains poorly understood, most likely due to its widespread rather than isolated effects. For example, it is unknown if and how it impacts on performance in model-based decisions, and on the arbitration between model-based and model-free controllers. This is the question I addressed in the present study where I formally tested whether dopamine influences the degree to which behaviour is governed by either control system.

5.3 Methods

5.3.1 Participants

18 healthy males (mean age: 23.3 (SD: 3.4)) participated in two separate sessions. Data from two additional participants were not included in the analysis as those participants misunderstood instructions and performed at chance level. The UCL Ethics committee approved the study and participants gave written informed consent before both sessions.

5.3.2 Dopamine drug manipulation

Participants were tested in a double-blind, fully counterbalanced, repeated measures setting on L-DOPA (150 mg $_{L}$-3,4-dihydroxyphenylalanine / 37,5 mg benserazide; Madopar®, Roche UK), and on placebo (500mg calcium carbonate; Calcit®, Procter and Gamble) dispersed in orange squash (see section 3.7 for a description of levodopa). The task was administered 55.0 (SD: 4.7) minutes after drug administration. Session 1 and 2 were approximately one week apart (at least 4, but no more than 14 days) with both sessions at the same time of day. All participants except one participated in the morning to minimise time-of-day effects. I assessed drug effects on self-reported mental state using a computerised visual analogue scale immediately before starting the task (Bond and Lader, 1974)

5.3.3 Task & Analysis

I used the task and analysis strategies as described in chapter 4.

5.4 Results

Using repeated measures ANOVA I examined the probability of staying at the first stage dependent on drug state (L-DOPA or placebo), reward on previous trial (reward or no-reward), and transition type on previous trial (common or uncommon) (Figure 4.1). A significant main effect of reward, $F(1,17) = 23.3$, $p < 0.001$, demonstrated a model-free component in behaviour (i.e. reward increases stay probability regardless of the transition type). A significant interaction between reward and transition, $F(1,17) = 9.75$, $p = 0.006$, revealed a model-based component (i.e. participants also take the task structure into account). These results show both a direct reinforcement effect (model-free) and an effect of task structure (model-based) and replicate previous findings (Daw et al., 2011).

The key analyses here concerned whether L-DOPA modulated choice propensities. Critically, I observed a significant drug*reward*transition interaction, $F(1,17) = 9.86$, $p=0.006$, reflecting increased model-based behaviour under L-DOPA treatment. I also observed a main effect of drug, $F(1,17) = 7.04$, $p = 0.017$, showing that participants were less perseverative under L-DOPA treatment. Interactions between drug and transition, $F(1,17) = 4.09$, $p = 0.06$, or drug and reward (which would indicate a drug-induced change in model-free control), $F(1,17) = 1.10$, $p = 0.31$, were not significant.

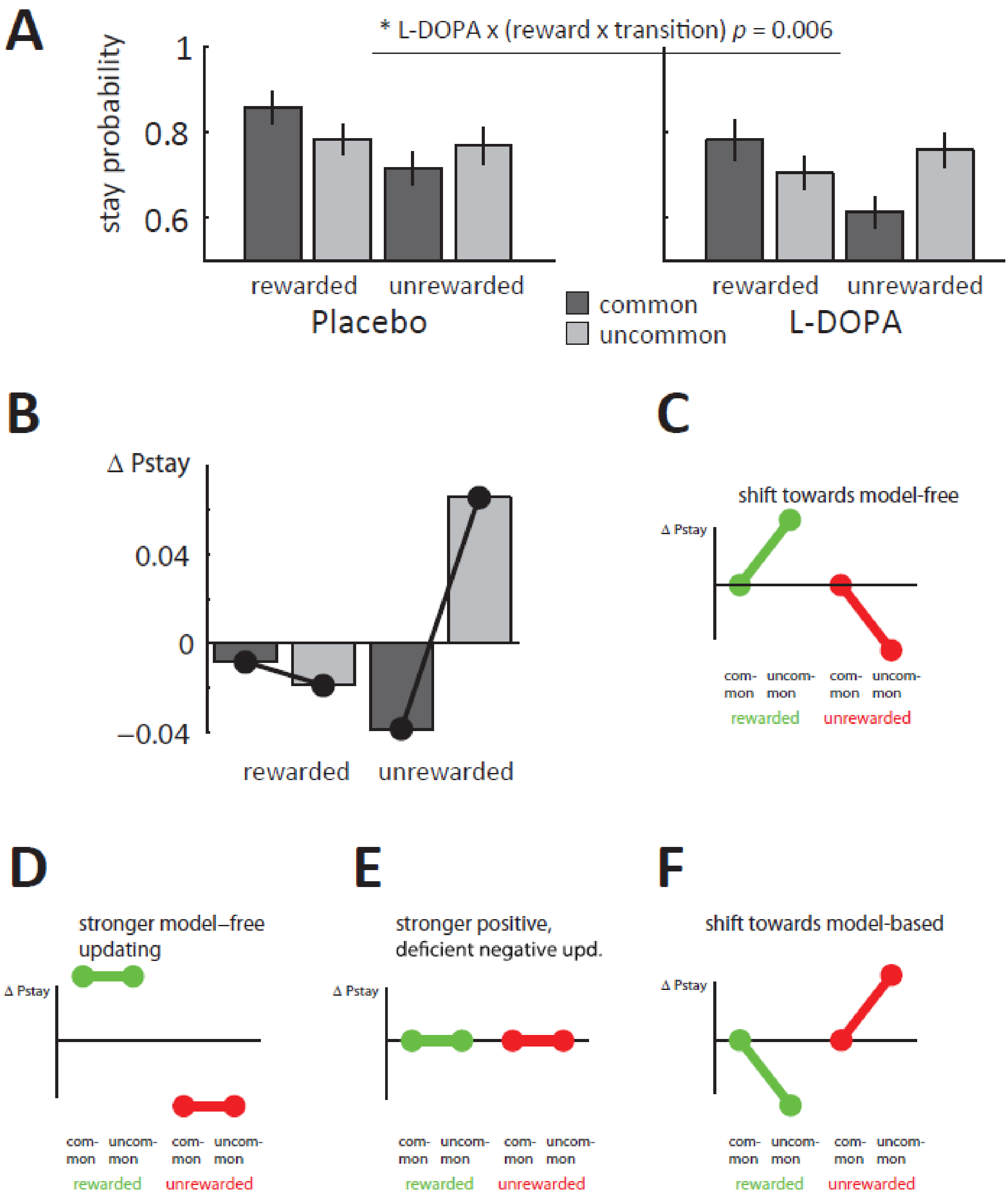


*Figure 5.1: (A) Participants' task behaviour showed characteristics of both model-free and model-based influences, demonstrating that participants combined both strategies in the task. The reward*transition interaction (a measure of the extent to which participants consider the task structure) was significantly larger in L-DOPA compared to placebo, indicating stronger model-based behaviour. (B) Difference in stay probability between L-DOPA and placebo condition, corrected for the main effect of drug. The observed interaction indicates a shift towards model-based choice (see F) while there is*

no resemblance to any of the three effects implicating the model-free system (see C-E). (C-F) Illustration of expected differences in stay probability for hypothetical drug effects. See Figure 5.2 and Table 5.1 for validation of these hypotheses. (C) Trials after uncommon transitions (2nd and 4th bar) are discriminatory between model-free and model-based choice, whereas both models make equal predictions for trials after common transitions (cf. Figure 4.1). A shift towards model-free control would be indicated by an increased propensity to stay with the chosen pattern after uncommon rewarded trials and an increase in switching after uncommon unrewarded trials. (D) Stronger or faster model-free learning would increase the reward-dependent effect and be expressed as general increase to stay after rewarded trials, and general decrease to stay after unrewarded trials. (E) A selective enhancement of positive updating paired with impairment in negative updating might not change mean-corrected stay probabilities. This is because enhanced positive updating leads to a stronger propensity to stay after rewarded trials, while impaired updating of unrewarded trials decreases the propensity to switch after such trials. (F) Opposite to C, a shift towards model-based control is expressed by enhanced sensitivity to the task structure.

Figure 5.1B shows the difference in stay probability between drug states corrected for a main effect of drug. Note that dopamine treatment particularly affected choices after unrewarded trials and a post-hoc contrast, testing for a differential drug effect after unrewarded compared to rewarded trials, confirmed this was significant, $F(1,17) = 12.68$, $p = 0.002$. Figure 5.1C-F illustrate how a number of hypothesised effects of L-DOPA might manifest itself in a stay-switch analysis (see Figure 5.2 for a validation of these hypotheses using simulations). Qualitatively, the data in Figure 5.1B resemble a shift towards model-based control, most notable after unrewarded trials. In contrast, my results do not resemble any of the putative model hypotheses that invoke modulation of a model-free system.

To confirm that the winning model can capture key behavioural findings (i.e. the drug*reward*transition interaction on stay-switch behaviour) I generated data for 500 virtual participants on this task using representative parameters from the hybrid reinforcement learning model from section 4.4. These data were then subjected to a stay-switch analysis. I found an identical pattern of effects in these generated data as observed empirically in the participants (Figure 5.2A). Most importantly, the data generated by the model showed a significant 3-way interaction, indicating that the model indeed captures key components of the data (see Table 5.1). Note that, as expected, the model did not replicate the asymmetry in rewarded versus unrewarded trials shown in Figure 5.1B.

The idealised hypotheses put forward for the stay-switch analysis in Figure 5.1C-F were based on ideas derived from previous literature. To validate these hypotheses I generated choices for virtual participants, but now with adjustments to parameters based on specific hypotheses (Figure 5.2B-D). The key hypotheses are fully supported by these simulations, showing that the computational models capture the key behavioural signatures of model-free and model-based behaviour. The data generated by this model was subjected to the same ANOVA as the participant data, showing the same effects as found in participants, most notably the three-way interaction that supports the claim that L-DOPA enhances model-based behaviour. The model thus provides a reasonable account of the data. Identical patterns exist between the two datasets, given the statistical model used.

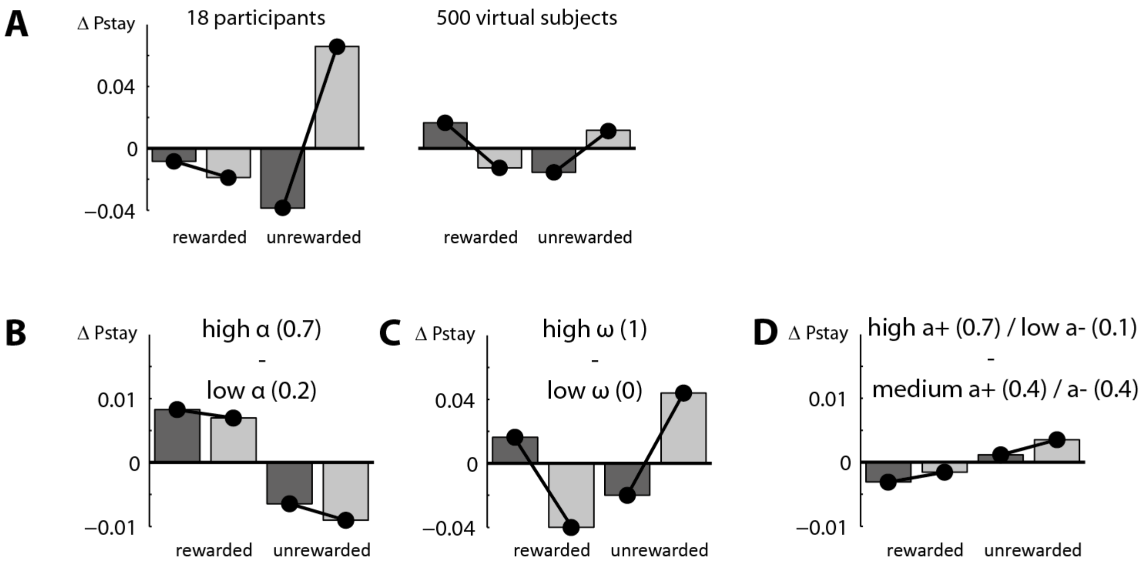


*Figure 5.2: (A) Mean-corrected P(stay)$_{ON}$ - P(stay)$_{OFF}$ for 18 participants reported in the study (left) and 500 virtual participants using representative parameters for the reinforcement learning model (right). (B) Modulation of the model-free learning rate α. A change in learning rate alters stay probability after rewarded versus unrewarded trials, but does not interact with transition. This is equivalent to Figure 5.1D. (C) Model-based (ω = 1) versus model-free agent (ω = 0) shows a stronger reward*transition interaction. This is equivalent to Figure 2F. (D) Increase in positive learning rate and decrease in negative learning rate does not change relative stay probabilities, similar to the prediction in Figure 5.1E.*

Table 5.1: statistical comparison of model-generated versus participant data (related to Figure 5.2)

Effect	**18 participants**		**500 virtual participants**	
	F(1,17)	p	*F*(1,499)	p
drug	7.04	= .02	83.00	< .001
reward	23.30	< .001	6.01	= .02
transition	< 1	~	< 1	~
drug x reward	1.10	= .31	< 1	~
drug x transition	4.09	= .06	< 1	~
reward x transition	9.75	= .006	561.79	< .001
drug x reward x transition	9.86	= .006	16.62	< .001

There was no evidence for differences in drowsiness or general alertness (Bond and Lader, 1974) between sessions (paired t-tests over each score; smallest $p > 0.1$) or in average response times between drug states (first stage $RT_{L\text{-}DOPA}$ = 593ms, $RT_{Placebo}$ = 586ms; paired t-test, $p = 0.70$).

Finally, I tested for order effects by repeating the analyses with session instead of drug as factor. There were no significant differences in stay-switch behaviour (repeated measures ANOVA: main effect of session $F(1,17) < 1$; session*reward, $F(1,17) < 1$; session*(reward*transition), $F(1,17) = 1.37$, $p = 0.26$). Thus these results provide compelling evidence for an increase in the relative degree of model-based behavioural control under conditions of elevated dopamine.

5.5 Discussion

It is widely believed that both model-free and model-based mechanisms contribute to human choice behaviour. In this study I investigated a modulatory role of dopamine in the arbitration between these two systems and provide the first evidence that L-DOPA increases the relative degree of model-based over model-free behavioural control.

The use of systemic L-DOPA combined with a purely behavioural approach precludes strong conclusions about the precise anatomical location of physiological changes that led to the observed shift to model-based control. Nevertheless, I provide a number of possible explanations for how this effect might be mediated in the brain that could guide further studies. First, increased dopamine levels may improve performance of component processes of a model-based system. Dopamine has previously been associated with an enhancement of prefrontal cognitive functions such as reasoning, rule learning, set shifting, planning and working memory (Cools et al., 2002; Lewis et al., 2005; Mehta et al., 2005; Clatworthy et al., 2009; Cools, 2011), and these processes are most likely co-opted during model-based decisions. Previous theoretical considerations link a system's performance to its relative impact on behavioural control, such that the degree of model-based versus model-free control depends directly on the relative certainties of both systems (Daw et al., 2005). Increased processing capacity might enhance certainty in the model-based system and would thus predict the observed shift in behavioural control that I detail here.

Second, a more conventional account is that increased dopamine exerts its effect through an impact on a model-free system. According to this view, excessive dopamine disrupts model-free reinforcement learning, which is then compensated for by increased model-based control. Specifically, elevated tonic dopamine levels may reduce the effectiveness of negative prediction errors (Frank et al., 2004; Voon et al., 2010). However, this explanation fails to account for the results presented here. Firstly, a disruption of negative prediction errors under L-DOPA would change stay probabilities independent of transition type (Figure 5.1E), which is incompatible with the reward*transition interaction observed here (Figure 5.1B). This argues against the idea that L-DOPA in this study enhanced the relative degree of model-based behaviour through a disruption of the model-free system.

Finally, dopamine could facilitate switching from one type of control to the other akin to the way it decreases behavioural persistence (Cools et al., 2003). It is known that over the course of instrumental learning the habitual system assumes control from the goal-directed system (Adams and Dickinson, 1981; Yin et al., 2004), but the goal-directed system can quickly regain control in unforeseen situations (Norman and Shallice, 1986; Isoda and Hikosaka, 2011). This could explain why I observed a stronger switch to model-based behaviour following unrewarded trials: the lack of rewarding feedback may prompt the need to re-evaluate available options and invest more energy to prevent another non-rewarding event by switching to model-based control. Note that it is possible and indeed likely that a facilitation of control switching under L-DOPA works in concert with an enhancement of the model-based system itself.

The predominant view in computational and systems neuroscience holds that phasic dopamine underlies model-free behaviour by encoding reward prediction errors. On the other hand, animal and cognitive approaches emphasise a role for dopamine in model-based behaviour such as planning and reasoning (Berridge, 2007; Robbins and Everitt, 2007; Clatworthy et al., 2009; Cools, 2011). Contrasting with interest in the model-free and model-based system separately is the lack of data on the arbitration between these two behavioural controllers. Our experiment fills this gap by pitting model-free and model-based control against each other in the same task and in so doing provides strong evidence for an involvement of dopamine in the arbitration between model-free and model-based control over behaviour.

Our findings advocate an effect of L-DOPA on the arbitration between model-based and model-free control, without a modulation of the model-free system itself. Note that the majority of studies reporting enhanced or impaired learning under dopaminergic drugs used either Parkinson’s disease (PD) patients (Frank et al., 2004; Voon et al., 2010) or involved agents that primarily act at D2 receptors (Cools, 2006; Frank and O'Reilly, 2006). In contrast with these studies, I did not find evidence for any modulation by L-DOPA of model-free learning or indeed evidence of impaired model-free choices. These deviations might partly be explained by PD patients' more severely reduced dopamine availability off their dopamine replacement therapy (in contrast to the placebo condition), and the much higher doses of medication involved in PD treatment. Consistent with this explanation is that the effect of L-DOPA on instrumental learning in healthy volunteers was found to be significant only when compared to an inhibition of the dopamine system (via haloperidol) but not when compared to placebo (Pessiglione et al., 2006).

Dopamine itself is a precursor to norepinephrine and epinephrine, potentially contributing to the observed effects. However, L-DOPA administration causes a linear increase in dopamine levels in the brain without affecting norepinephrine levels (Everett and Borcherding, 1970). Another possibility would be that L-DOPA exerts effects through interactions with the serotonin system. Such an interaction, between dopamine and serotonin, is known to play a role in a range of higher-level cognitive functions (Boureau and Dayan, 2011).

These data open the door to further experiments aimed at elucidating the precise neural mechanisms underlying the arbitration between both controllers. In the following chapters 6 and 7 I focused in on the dorsolateral prefrontal cortex as one such neural mechanism of model-based control specifically.

6 Disruption of dorsolateral prefrontal cortex impairs model-based control

6.1 Abstract

In chapter 6 I hypothesised that L-DOPA might increase model-based control through a modulation of prefrontal cortex. Here I set out to test whether I could achieve the opposite effect, whereby a disruption of prefrontal cortex might impair model-based but not model-free control. I specifically focused on dorsolateral prefrontal cortex (dlPFC), and show it is possible to reduce model-based control by disruption of right dlPFC via transcranial magnetic stimulus (TMS). In contrast, disruption of left dlPFC impaired model-based performance only in those participants with low working memory capacity. Neither left nor right dlPFC disruption had an effect on the level of model-free control, in line with the notion of dissociable neural circuits supporting model-based and model-free control.

6.2 Introduction

In this study the goal was to manipulate the relative balance between model-based and model-free control in human participants. I focused on the dorsolateral prefrontal cortex (dlPFC) as a substrate for model-based processes based on previous evidence for its role in the construction and use of associative models (Gläscher et al., 2010; Wunderlich et al., 2012b; Xue et al., 2012) and the coding of hypothetical outcomes (Abe and Lee, 2011).

In addition I took into account work from studies in non-human primates which also implicated the dlPFC as a site for convergence of reward and contextual information (Lee and Seo, 2007), while lesions of rat prelimbic region—argued by some to be equivalent to primate dlPFC (Uylings et al., 2003; Fuster, 2008)—abolish flexible decision-making (Killcross and Coutureau, 2003).

Therefore, while the literature suggests a crucial role for this region in model-based control at the time of the study there was no evidence for a necessary role that might support this hypothesis. Here I used a transient lesion model, as engendered by theta burst transcranial magnetic stimulation (TBS), to provide evidence for a necessary role of dlPFC in model-based behaviour.

6.3 Methods

6.3.1 Participants

I recruited 25 human participants (mean age (SD): 24.2 (4.0) years; 15 females) to perform the two-step task (see chapter 4 and Daw et al., 2011). All participants had normal or corrected-to-normal vision, and without a history of psychiatric or neurological disorder. All participants provided written informed consent prior to start of the experiment, which was approved by the Research Ethics Committee at University College London (UK). No participants were excluded over the course of the experiment.

6.3.2 Theta burst stimulation

Participants received TBS (see methods section 3.6) over the right dlPFC, left dlPFC, and vertex on three separate occasions, with site order counterbalanced across 24 participants, and the 25th participant received a randomly selected session order. I identified stimulation sites as follows: the MNI coordinates for the right dlPFC (x = 37, y = 36, z = 34) were taken from a previous study that used a combination of individual anatomy and fMRI results to pinpoint the dlPFC (Feredoes et al., 2011). For the left dlPFC (x = -37, y = 36, z = 34) I took the negative of the right dlPFC x-coordinate. These MNI coordinates were transformed to coordinates in native space by taking the inverse normalization parameters from unified segmentation of a previously acquired T1w structural image as implemented in SPM8 (Wellcome Trust Centre for Neuroimaging, UCL, UK). I visually confirmed that the coordinates in native space corresponded to middle frontal gyrus (as in Feredoes et al., 2011). These coordinates were then entered as targets into Visor2 (ANT B.V.), which uses a 3D camera to guide the stimulation coil (Magstim) to the target coordinate. The vertex was set to the Cz of the 10-20 system. To mimic the stimulation experience for the participant, I entered the vertex coordinates into Visor2 and used 3D navigation to target the stimulation coil.

I administered stimulation in 5 Hz bursts of 3 pulses set 20 ms apart, for 40 s, amounting to a total of 600 pulses. Stimulation intensity was set for each individual participant as 90% of active motor threshold (AMT). AMT was defined as the lowest stimulation intensity, expressed as % of max output of the Magstim equipment that reliably (3/5 times) yielded a visible muscle twitch in the hand when stimulating the hand area of the contralateral motor cortex with a single pulse. During this procedure participants held (lightly) an item in the hand contralateral to the stimulation site. For technical and safety reasons, the maximum stimulation intensity was set to 51% of maximum output; as such, any participant with an AMT > 56% received TBS at 51% of maximum output. Note that such reduced stimulation will make it less likely to find significant effects of TMS. The average stimulation intensity was 49% (range: 40 - 51%) of maximum output.

6.3.3 Task & analysis

I used the task and analysis strategies that are described in the introduction to this thesis in chapter 4.

6.3.4 Baseline working memory capacity

On the first session, before any TBS or practice on the main task, participants performed a 7-minute task to establish visuospatial working memory capacity. In short, participants had to remember the location of 5 simultaneously presented dots in a circular array of 16 positions. After a delay the participant was asked whether, for one of the 16 locations, a red dot was presented. From these data I calculated a K-value, reflecting the amount of information that the participant can store in working memory. For details of the task and analysis, see McNab and Klingberg (2008).

6.4 Results

Participants' first-stage choices for all three TBS conditions qualitatively reflected a hybrid of model-based and model-free control (Figure 6.1, cf. Figure 4.1). I estimated the main effect of reward and the reward-by-transition interaction for each TBS site using hierarchical logistic regression, with all coefficients taken as random effects across participants (see Table 6.1 for list of regressors, and section 4.3 for a description of the regression analysis).

Table 6.1: Regressors for hierarchical logistic regression on stay (coded as 1) or switch (coded as 0) for each first-stage choice. Reward is coded as 1 and -1 for presence and absence, respectively; transition is coded as 1 and -1 for common and uncommon, respectively. The main effect of vertex is subsumed in the intercept.

Intercept
left dlPFC
right dlPFC
left dlPFC * reward
right dlPFC * reward
vertex * reward
left dlPFC * transition
right dlPFC * transition
vertex * transition
left dlPFC * reward * transition
right dlPFC * reward * transition
vertex * reward * transition

I observed positive coefficients for the reward and reward-by-transition regressors for all three TBS sites (all p < .006), confirming that behaviour comprised a hybrid of model-free and model-based control. Levels of model-based and model-free control after left and right dlPFC TBS were then contrasted with vertex (Figure 6.1B). I observed that TBS to either left (p = .52) or right (p = .20) dlPFC did not significantly change model-free control compared to vertex. By contrast, model-based control was disrupted following TBS to right (p = .01) but not left (p = .89) dlPFC compared to vertex. I observed no difference in model-based control between left and right dlPFC (p = .13).

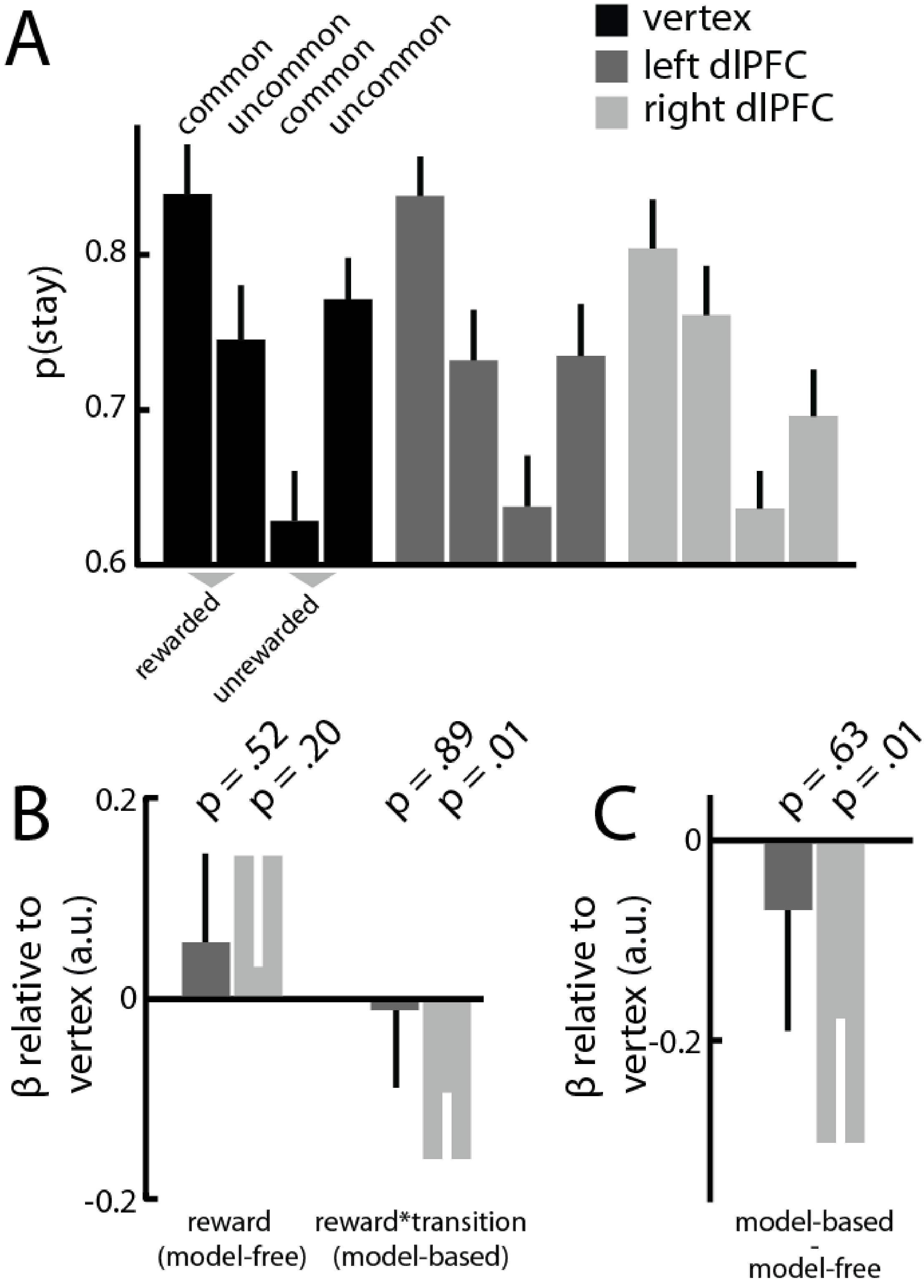


Figure 6.1: (A) The probability of repeating the same first-stage choice is shown as a function of reward and transition experienced on the previous trial. The pattern of choices qualitatively resembles influences of both model-based and model-free control for all three stimulation sites (cf. Figure 4.1). (B) I quantified

model-free and model-based control as the main effect of reward and the reward-by-transition interaction, respectively, in a hierarchical logistic regression on stay/switch behaviour on each trial. Disruption of right dlPFC reduced model-based control compared to vertex. TBS did not significantly affect model-free control. (C) The relative balance between the controllers was calculated as $\beta_{model\text{-}based} - \beta_{model\text{-}free}$*. The balance significantly shifted towards model-free control after disruption of right, but not left, dlPFC compared to vertex. Error bars indicate SEM.*

I also computed a measure of the relative balance between these two systems as $\beta_{\text{model-based}} - \beta_{\text{model-free}}$ (Figure 6.1C). This showed a significant shift towards model-free control caused by TBS to right ($p = .01$) but not left ($p = .63$) dlPFC compared to vertex. I observed no difference between left and right dlPFC ($p = .11$). Together these results provide evidence that right dlPFC exerts a causal role in model-based control, and show that the balance between model-based and model-free control can be manipulated through prefrontal disruption via TBS.

I then repeated these analyses to examine order effects. In pairwise session comparisons I found no effect of session on model-free or model-based control, or on the balance between model-based and model-free control (all $p > .14$), except for an increase in model-free control in session 3 compared to session 1 ($p = .04$).

Model-based control is thought to depend on a number of processes including prefrontal working memory (WM) capacity. Given that studies of WM report lateralised functionality (e.g. Mull and Seyal, 2001) I asked whether the magnitude of a TBS effect might be related to WM capacity. To examine such inter-individual differences I could not use the population parameter estimates obtained through the regression. Instead, I extracted the numerical magnitude of the main effect of reward, the reward-by-transition interaction and the difference between the two from each participant's average stay probability in each of the four reward/transition conditions in each stimulation condition.

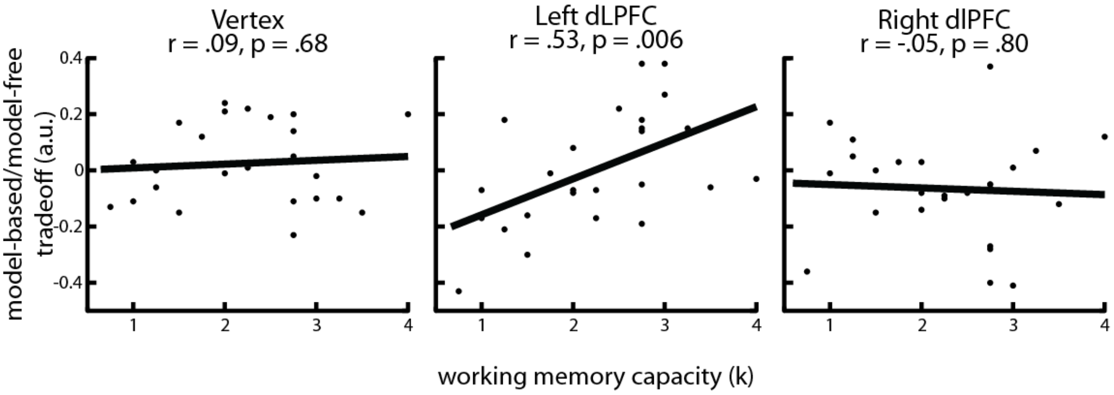


Figure 6.2: Working memory capacity interacts with stimulation in left dlPFC. Working memory (WM) capacity did not predict the balance between model-based and model-free control after disruption of vertex (left) or right dlPFC (right). In contrast, higher WM was associated with relatively stronger model-based control after disruption of left dlPFC (middle) with the correlation being significantly more positive than for right dlPFC (permutation test, p = .009) or vertex (p = .06).

I first asked whether model-free or model-based control independently correlated with WM in any of the 3 stimulation conditions. Only the magnitude of the reward-by-transition interaction, inferred as model-based control, correlated with WM following disruption to left dlPFC (r = .45, p = .02; all other p > .10). I then correlated the balance between the two systems in all stimulation conditions with WM. Strikingly, only behaviour after disruption of left dlPFC was WM-dependent (Figure 6.2; vertex, r = .09, p = .68; left dlPFC r = .53, p = .006; right dlPFC, r = -.05, p = .80). Pairwise permutation tests revealed the correlation was significantly more positive in left compared to right dlPFC (10^5

permutations, p = .009), marginally more positive in left dlPFC compared to vertex (p = .06), and not significantly different between right dlPFC and vertex (p = .52). Taken together, these data show that the effect of left dlPFC disruption on the balance between model-based and model-free control depends on WM capacity, with high WM participants retaining more model-based control compared to those with low WM.

Whereas first-stage choices allowed me to dissociate model-based from model-free control, both types of control make equivalent predictions for second-stage choices as there is no task structure to exploit. It has, however, been shown that TBS to left, but not right, dlPFC modulates probabilistic instrumental reward learning (Ott et al., 2011). I therefore sought to explore the effects of TBS on 1-step reward learning here as well (Figure 6.3). I examined second-stage choices using hierarchical logistic regression similar to the analysis of first-stage choices: stay-switch behaviour was regressed against reward received on the most recent trial involving that second-stage pair. Transition was not included as a factor because second-stage choices are assumed to be independent of the transition type that led to the state. I observed that TBS to left dlPFC affected second-stage choices by making them more perseverative (p = .02) and more sensitive to reward (p = .006) compared to vertex (see Figure 6.3). No such effect was found for right dlPFC (p = .11 and p = .10, respectively). There was no difference between left and right dlPFC (perseveration: p = .35, effect of reward: p = .20).

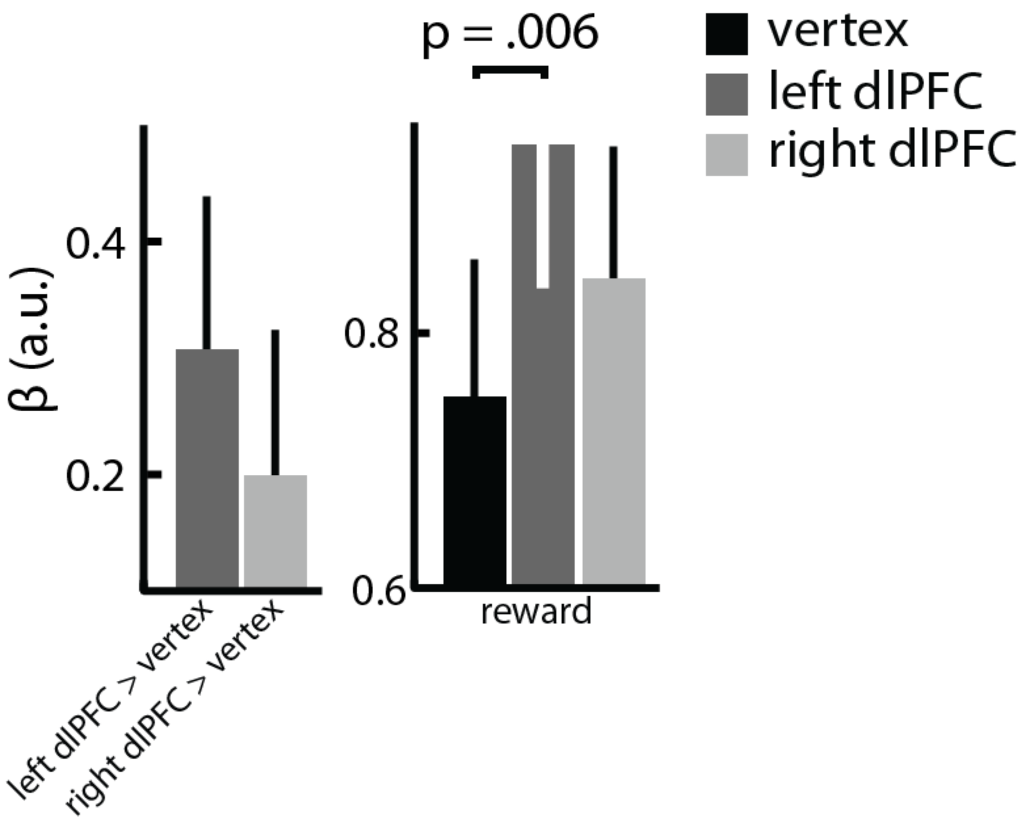


Figure 6.3: analysis of second-stage choices. The main effect of each stimulation site (left) captures the propensity to stay with the same stimulus irrespective of reward, relative to the vertex condition. Participants become more perseverative after left dlPFC TBS compared to vertex (p = .02) on second-stage choices. Note that the main effect of vertex is subsumed in the intercept of the regression, such that a coefficient significantly different from zero indicates a significant deviation from vertex. The main effect of reward in each stimulation condition (right) indicated participants tended to stay with a rewarded stimulus more than with an unrewarded stimulus (all p < .001), but this propensity was stronger after left dlPFC TBS compared to vertex (p = .006). Error bars indicate SEM.

6.5 Discussion

The balance between model-based and model-free control is often framed as a competition between a flexible, forward-looking, system and a simpler retrospective stimulus-response-based system (Daw et al., 2005). These results show that the balance between these two systems can be causally manipulated in the human brain by a disruption of prefrontal cortex. The data suggest that TBS to right dlPFC impairs a key node in a network that underpins model-based control (cf. Killcross and Coutureau, 2003; Gläscher et al., 2010). I further show an involvement of left dlPFC in model-based control that is related to individual differences in working memory, suggesting differential roles for left and right dlPFC in the functional architecture underlying deliberative choice.

Animal lesion and human imaging work suggest that sectors of prefrontal cortex are involved in high-level cognition and decision-making (Miller and Cohen, 2001). These studies have shown correlates of model-based control in ventromedial prefrontal cortex and dlPFC as well as outside the prefrontal cortex, e.g. dorsomedial striatum (Gallagher et al., 1999; Killcross and Coutureau, 2003; Hikosaka, 2007; Boorman et al., 2009; de Wit et al., 2009; Gläscher et al., 2010; Liljeholm and O'Doherty, 2012; Wunderlich et al., 2012b; Xue et al., 2012). In contrast, model-free control is most strongly associated with the dorsolateral striatum and infralimbic cortex (Yin et al., 2004; Balleine and O'Doherty, 2010; Wunderlich et al., 2012b). Furthermore, a strong dependence of model-based control on prefrontal systems is hinted by a finding that its dominance can be abolished during dual-task performance (Otto et al., 2013). However, up to now the key human evidence for dlPFC involvement in model-based control has been based on correlational evidence using functional imaging (fMRI). Here I show that model-based control is impaired by a transient disruption of the right dlPFC, providing causal evidence for its involvement in complex, flexible, decision-making. I note this effect was significant only when compared to the vertex, the control site, but not when compared to left dlPFC. I speculate this might be due to individual variation in the role of the left dlPFC in model-based control, or in the strategies employed by the participants to solve the task.

An influential hypothesis about the balance between model-based and model-free control states that their individual influence over behaviour is governed by their respective uncertainties (Daw et al., 2005). Within this framework, my results can be interpreted as emerging out of a disruption to a key component process of model-based control (e.g. the utilization of associative models, Gläscher et al., 2010). This would lessen the certainties of model based predictions leading to an attenuated dominance over behaviour—similar to that observed when participants are distracted by a dual task (Otto et al., 2013). However, whereas disruption of right dlPFC led to an unambiguous impairment of model-based control, the effect of TBS on the left dlPFC was dependent on baseline WM capacity. Specifically, higher WM capacity conferred a degree of protection against a shift towards model-free control upon disruption of left dlPFC, whereas participants with low WM capacity appear to require an uncompromised left dlPFC for the exercise of model-based control. I acknowledge uncertainty as to what precise factors might explain this finding.

An increase in perseveration might be caused by a reduction in striatal dopamine after left TBS (Ko et al., 2008), which is known to affect behavioural flexibility and perseverance (Cools et al., 2006). It is, however, unclear why such a reduction in striatal dopamine would be associated with *improved* reward learning. However, this finding replicates a previous study that found improved reward learning after left, but not right, TBS (Ott et al., 2011). Arguing against a role for dopamine in this increase in reward sensitivity is a null effect of dopamine administration on second-stage choices shown previously (Wunderlich et al., 2012a).

The effect of TBS on sub-cortical dopamine might also play a role in first-stage choices. The reduction in dopamine might interact with baseline dopamine levels that are known to co-vary with WM capacity (Cools et al., 2008), such that high WM participants are more resilient against TBS-induced decreases in dopamine than low WM participants. I, with colleagues, previously showed that dopamine levels modulate the balance between model-based and model-free control (de Wit et al., 2011; de Wit et al., 2012b; Wunderlich et al., 2012a), and a TBS-induced depletion in low WM (i.e. low dopamine) individuals might have a more pronounced effect than a similar depletion in high WM (i.e. high dopamine) individuals. However, given that I did not directly measure dopamine levels, future work could usefully explore potential interactions between WM and model-based control to fully understand the effect reported here.

The findings speak to the literature on goal-directed and habitual behaviours (Balleine and O'Doherty, 2010). Although model-based/model-free and goal-directed/habitual control are not synonymous, the former provides a computational framework that can encompass key features of goal-directed and habitual control (for a review, see Dayan and Niv, 2008). I would predict a disruption of right dlPFC would also impair goal-directed behaviour in devaluation and contingency degradation tests in humans, as has been shown in rats (Balleine and O'Doherty, 2010).

In summary, I provide evidence for a necessary role of the right dlPFC in flexible, model-based decision-making. Our findings invite the question as to whether naturally occurring variation in dlPFC function and connectivity is a marker for predisposition towards model-free as opposed to model-based control, and whether an enhancement of dlPFC function (e.g. through other stimulation protocols) might improve rather than impair model-based control. I set out to test whether a putative improvement of right dlPFC would indeed lead to stronger model-based control in the next chapter.

7 Transcranial direct current stimulation does not affect model-based control

7.1 Abstract

There is broad consensus that the prefrontal cortex supports goal-directed, model-based decision-making. Consistent with this I showed in chapter 6 that model-based control can be impaired through transcranial magnetic stimulation of right dorsolateral prefrontal cortex in humans. Here I tested the hypothesis that an enhancement of model-based control could be achieved by anodal transcranial direct current stimulation of the same region. I tested 22 healthy adult human participants in a within-participant, double-blind design in which participants were given Active or Sham stimulation over two sessions. I show active stimulation had no effect on model-based or on model-free control compared to Sham stimulation. I also introduced a novel regression analysis that examines model-based and model-free influences multiple trials into the past, which also showed no effect of stimulation. These null effects are substantiated by a power analysis, which suggests that the study had at least 60% power to detect a true effect, as well as a Bayesian model comparison, which favours a model of the data that assumes stimulation had no effect over models that assume stimulation had an effect on behavioural control. Although I cannot entirely exclude more trivial explanations for the null effect, for example related to (faults in) the experimental setup, these data suggest that anodal transcranial direct current stimulation over right dorsolateral prefrontal cortex does not improve model-based control, despite existing evidence that transcranial magnetic stimulation can disrupt such control in the same brain region.

7.2 Introduction

Electrical stimulation of the human brain has received widespread attention over recent years. It has been used to study the function of healthy cortex (Marshall et al., 2004), connectivity between regions (Mars et al., 2009), as an avenue for treatment in disorders such as depression, Parkinson’s disease and stroke (Fregni et al., 2005b; Boggio et al., 2006; Boggio et al., 2008; Baker et al., 2010a), and to improve normal function such as in skill learning (Nitsche et al., 2003; Reis et al., 2009).

Here I used transcranial direct current stimulation (tDCS), a technique whereby two electrodes are placed on the skull and a fixed current level is applied (also see Methods chapter, and Nitsche and Paulus, 2001). This technique is reported to increase and decrease the excitability of the neural tissue underlying the anodal and cathodal electrode respectively (Nitsche and Paulus, 2001; Nitsche et al., 2003). A number of studies have suggested that high-level cognition can be improved by anodal stimulation of the prefrontal cortex. Specifically, stimulation of the dorsolateral prefrontal cortex (dlPFC) has been shown to decrease risk-taking (Fecteau et al., 2007a), improve working memory (Fregni et al., 2005a; Mulquiney et al., 2011) and improve classification learning (Kincses et al., 2004).

I focused on the right dlPFC based on evidence for its role in model-based processes such as the construction and use of associative models (Gläscher et al., 2010; Wunderlich et al., 2012b; Xue et al., 2012) and the coding of hypothetical outcomes (Abe and Lee, 2011). Work on non-human primates also implicates the dlPFC as a site for convergence of reward and contextual information (Lee and Seo, 2007). Furthermore, in chapter 6 I showed that right, but not left, dlPFC is necessary for model-based control, evidenced by a reduction in model-based control after disruptive theta-burst transcranial magnetic stimulation to the right dlPFC (Smittenaar et al., 2013b). Here, to complement these previous findings, I sought to enhance, rather than disrupt, model-based control through anodal stimulation. I used a task which has been shown to quantify model-based and model-free control (Daw et al., 2011; Wunderlich et al., 2012a; Otto et al., 2013) and tested participants undergoing anodal or Sham tDCS stimulation to the right dlPFC in a double-blind, counterbalanced design. I hypothesised that anodal stimulation would improve model-based control without affecting model-free control, an effect driven by an enhancement of a component process of model-based control subserved by the right dlPFC.

7.3 Methods

I recruited 23 healthy participants to participate in an experiment over 2 sessions. All participants had normal or corrected-to-normal vision and no history of psychiatric or neurological disorders. One participant was excluded from analysis due to failed stimulation after an increase in resistance from drying electrodes, leaving 22 participants (11 female, mean age ± SD: 22.5 ± 5.3 years, all participants were at least 18 years of age at the time of consent) for analysis. Written informed consent was obtained from all participants prior to the experiment and the UCL Research Ethics Committee approved the study (project number 3450/003).

7.3.1 Setup of experiment and double-blinding procedure

Participants were tested on 2 occasions between 3 and 8 days apart, going through the same procedure on each day: after obtaining informed consent I determined the electrode locations, explained the task, guided participants through a short practice session, placed the electrodes on the scalp, turned on stimulation, and started the task. The experiment was double-blind, with both experimenter and participant unaware of the stimulation condition (Active or Sham). This was achieved through a system of blinding codes embedded in the stimulation machine (NeuroConn, Germany). First, researcher GP selected 24 pairs of 5-digit codes, each pair containing one code associated with Active and one code associated with Sham stimulation as programmed into the stimulation machine. These were then permuted such that half the pairs had Active stimulation on session 1 and Sham stimulation on session 2, whereas the other half of pairs had the reversed order. GP kept the unblinded version of the codes and handed the permuted set to PS, who acquired the data. Each participant was assigned a pair in order of testing date. When the participant was prepped for stimulation, their session-specific code was entered into the stimulation machine, which then administered the corresponding Active or Sham protocol without any indication as to the stimulation condition. I tested the participant's awareness of the stimulation condition at the end of the experiment (see below). PS was deblinded after acquisition of all 23 datasets.

7.3.2 Task & analysis

Whereas in previous chapters I examined only influences from 1 trial back using regressions, here I expanded on this approach to examine model-based and model-free influences that go up to 3 trials in the past. This provides a more fine-grained dissection of the influences of each system on behaviour. The dependent variable for trial *t* was 1 when stimulus A was chosen and 0 when stimulus B was chosen in the first stage. Each regressor then described whether events on trial t-1, t-2, and t-3 would increase (coded as +1) or decrease (coded as -1) the likelihood of choosing A according to a model-based or model-free system. If a trial contained a common transition the model-based and model-free system would make identical predictions, whereas on trials with uncommon transitions these predictions would be inverted. I additionally modelled the main effect of transition type (common as +1, uncommon as -1) on trial t-1, t-2 and t-3, which I predicted would have no effect on the propensity to choose stimulus A. I also tested 3 alternative models that used 1) one set of model-based regressors for both conditions, 2) one set of model-free regressors for both conditions and 3) one set of model-based and one set of model-free regressors for both conditions (‘null model’). These models allowed me to test whether the additional complexity of having separate regressors for the stimulation conditions was appropriate. These models were compared using the BIC and AIC values provided by the lme4 package.

I performed contrasts over the population coefficients to test for differences between conditions in model-free and model-based control. All p-values reported in the manuscript that pertain to the logistic regression were estimated using the “esticon” procedure in the “doBy” package which relies on the chi-square distribution (Højsgaard, 2012). Power analyses were performed using the Matlab 7.12.0 ‘sampsizepwr’ function and G*Power 3.1.7 (Faul et al., 2007; Faul et al., 2009). Other tests were performed in SPSS 17.0.

7.3.3 Stimulation

On both sessions the anodal electrode was placed over right dlPFC and the cathodal electrode over the inion. The inion was chosen for cathodal electrode placement in order to maximise current flow through the dlPFC. The right dlPFC was located using the 10/20 system, which is appropriate given the limited level of spatial resolution of tDCS (Herwig et al., 2003). In brief, I first located Fpz, Fz and Oz as 10%, 30% and 90% of the nasion-inion distance, measured from the nasion. I then located F8 as 30% of the distance between Fpz and Oz, measured from Fpz passing over the ears. Electrode F4, commonly used for the right dlPFC (Herwig et al., 2003), was then determined as 50% of the distance between F8 and Fz. I used conductive rubber electrodes inserted in a sponge cover measuring 7.5 by 6 cm, secured to the head using a bandage. I placed the electrode along the gyrus, i.e. the electrode was placed in superior-medial to inferior-lateral direction.

I used a DC-stimulator system (NeuroConn, Germany). In the Active condition a 2 mA current was delivered for 25 minutes with 15 s ramping-up and ramping-down. In the Sham condition the current ramped up then down over 15 s, and then performed continuous impedance testing. This manipulation made it very hard for the participant to tell which type of stimulation was given at what time. I confirmed this by giving a 2-alternative forced-choice at the very end of the experiment asking which session contained the Active stimulation. This test showed that participants as a group were not significantly different from chance at determining the session that contained Active stimulation (10 out of 22 participants guessed correctly, binomial test, p = .83). I employed a number of post-hoc checks to safeguard against experimental error. Firstly, I monitored the resistance reported by the DC-stimulator throughout the experiment, rejecting one participant for whom stimulation was stopped after a strong increase in resistance (>55 kΩ). Secondly, after the experiment I confirmed for a random set of 4 sham and 4 active codes that they were correctly linked to the sham or active stimulation procedure by examining the current with an amperometer.

This was the case for all 8 codes. Thirdly, I note that of the 100,000 possible codes that can be entered into the DC-stimulator only 200 are allowed, minimizing the possibility of erroneously entered codes.

After turning on stimulation the participant waited for 10 minutes before starting the task in order to ensure the effects of stimulation were fully established (Nitsche and Paulus, 2001). Altogether participants received 25 minutes of stimulation at 2 mA. It is known that cortical excitability changes outlast such stimulation durations by over an hour (Nitsche and Paulus (2001), though see Stagg et al. (2013)). The window of stimulation therefore need not fully overlap with the task, and in the design stimulation ended approximately halfway through the task. It should be noted that choices for stimulation parameters are based on studies of motor cortex stimulation. It is possible that these parameters, when used on frontal areas, have different effects. To my knowledge there is no published data on this, though I note this protocol is similar to that of other studies using tDCS on dlPFC (Kincses et al., 2004; Fecteau et al., 2007a).

7.4 Results

Participants earned £8.25 ± 2.56 during Active stimulation and £8.30 ± 2.39 during Sham stimulation (no difference in paired samples t-test, t(21) < 1). Participants missed 0.10 ± 0.37% of trials during Active stimulation and 0.09 ± 0.18% of trials during Sham stimulation (no difference in paired sampled t-test, t(21) < 1).

For comparison to previous studies using this task I plotted the stay probabilities based on reward/no-reward and common/uncommon transition on the previous trial (Figure 7.1). Qualitatively the pattern in both the Active and Sham condition resembles that of a hybrid controller (Figure 4.1, right) in which choices are influenced both by model-based and model-free control.

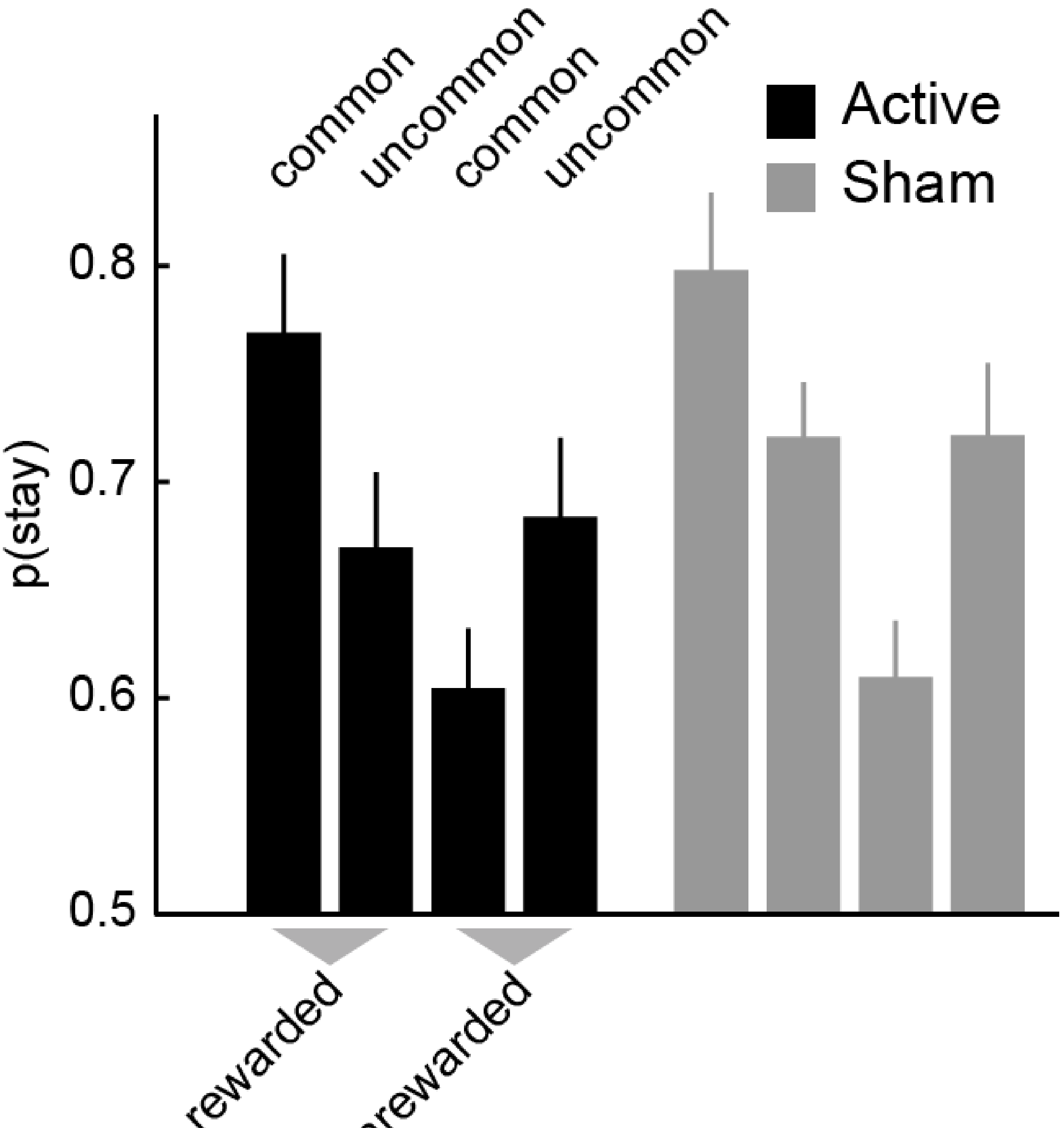


Figure 7.1: Stay probabilities as a function of reward and transition on previous trial. Participants showed a pattern of stay probabilities characteristic of hybrid model-based/model-free control during both Sham and Active stimulation of dlPFC. Error bars indicate SEM.

To quantify these influences and examine effects of trials that extend beyond the previous (lag-1) trial, I performed a hierarchical regression analysis (see Table 7.1 for regressors).

Table 7.1: Regressors in the full model for first-stage choices. MF = model-free; MB = model-based; SE = standard error. Lag denotes the effect of time. Bold-face indicates p < .05 uncorrected for multiple comparisons.

regressor	**estimate**	**SE**	**z-value**	**p**
intercept	0.25	0.03	7.81	**<0.0001**
Active	-264.18	194.46	-1.36	0.1743
Active MF Lag-1	287.02	62.06	4.63	**<0.0001**
Active MF Lag-2	293.64	50.73	5.79	**<0.0001**
Active MF Lag-3	172.87	51.73	3.34	**0.0008**
Active MB Lag-1	244.48	72.35	3.38	**0.0007**
Active MB Lag-2	180.58	66.90	2.70	**0.0069**
Active MB Lag-3	200.76	44.92	4.47	**<0.0001**
Sham MF Lag-1	374.51	51.11	7.33	**<0.0001**
Sham MF Lag-2	287.55	54.85	5.24	**<0.0001**
Sham MF Lag-3	246.79	59.53	4.15	**<0.0001**
Sham MB Lag-1	226.13	64.93	3.48	**0.0005**
Sham MB Lag-2	207.15	77.43	2.68	**0.0075**
Sham MB Lag-3	170.37	60.91	2.80	**0.0052**
Active transition Lag -1	-4.62	36.24	-0.13	0.8985
Active transition Lag -2	9.20	32.34	0.28	0.7760
Active transition Lag -3	-19.03	34.09	-0.56	0.5767
Sham transition Lag -1	-6.61	42.27	-0.16	0.8758
Sham transition Lag -2	15.68	33.42	0.47	0.6389
Sham transition Lag -3	-2.77	36.88	-0.08	0.9400

This revealed that all model-based and model-free regressors were significantly larger than zero, meaning both systems rely on events at least 3 trials into the past (Figure 7.2; see Table 7.1 for statistics).

Table 7.2: Contrasts performed on the full model. MF = model-free; MB = model-based; SE = standard error; χ^2 = chi-square distribution; df = degrees of freedom; Lag denotes the effect of time. Bold-face indicates p < .05 uncorrected for multiple comparisons.

contrast	**estimate**	**SE**	**χ^2 (1 df)**	**p**
MF Active > Sham	-155.32	119.50	1.69	0.1937
MB Active > Sham	22.17	131.42	0.03	0.8661
MF/MB x Active/Sham	-177.49	192.33	0.85	0.3561
MF Lag-1 Active > Sham	-87.49	55.46	2.49	0.1146
MF Lag-2 Active > Sham	6.09	54.82	0.01	0.9115
MF Lag-3 Active > Sham	-73.93	50.87	2.11	0.1461
MB Lag-1 Active > Sham	18.35	59.86	0.09	0.7592
MB Lag-2 Active > Sham	-26.57	60.15	0.20	0.6587
MB Lag-3 Active > Sham	30.39	54.31	0.31	0.5758
Lag MF Active	114.16	55.61	4.21	**0.0401**
Lag MF Sham	127.72	45.43	7.90	**0.0049**
Lag MB Active	43.72	60.32	0.53	0.4686
Lag MB Sham	55.76	45.04	1.53	0.2157
Lag MF > MB	142.40	124.64	1.31	0.2532
Lag MF Active > Sham	-13.57	65.62	0.04	0.8362
Lag MB Active > Sham	-12.04	70.89	0.03	0.8651
Lag MF/MB x Active/Sham	-1.53	102.76	0.00	0.9882

Contrary to my hypothesis I did not find a difference between the Active and Sham stimulation conditions in any of the contrasts (Table 7.2). I therefore report the absence of evidence for an effect of anodal tDCS to right dlPFC on model-free or model-based control. In subsequent analyses I explored whether this null effect was due to a lack of power in the experiment or due to an inability of tDCS to right dlPFC to modulate model-based or model-free control.

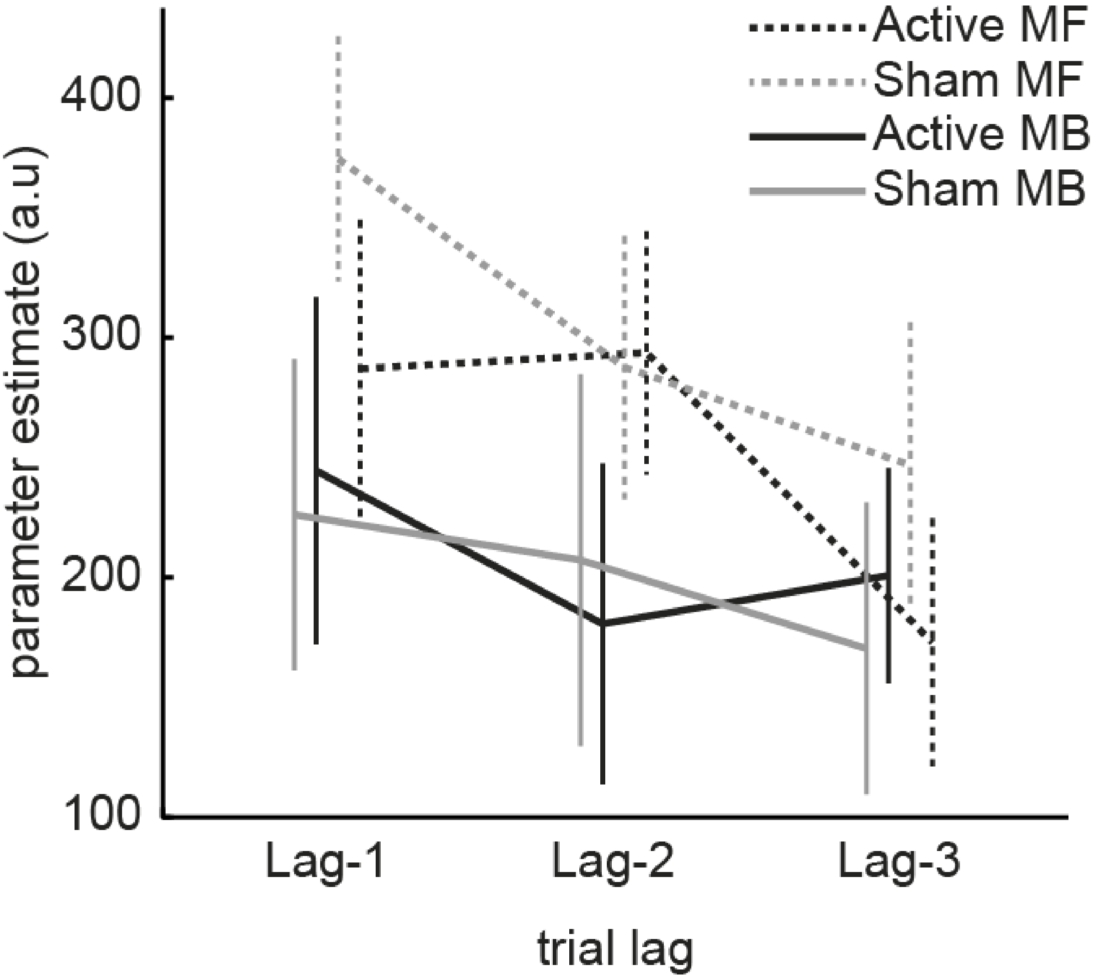


Figure 7.2: Model-based and model-free influences on choice. I estimated the dependence of a choice at trial t *on reward and transition events in trials t-1 up to t-3. These regression coefficients can be interpreted as model-based and model-free influences on choice, and larger coefficients indicate a stronger influence over choice. Firstly, all regression coefficients in the plot are significantly larger than zero, suggesting that model-based and model-free systems did not just rely on events on the previous trial but rather on events as far as 3 trials in the past. I did not observe any difference between Active and Sham conditions. Error bars indicate SEM.*

To estimate the power in the experiment I gathered effect size estimates in the published literature for manipulations involving the 2-step task (Wunderlich et al., 2012a) and for two tDCS experiments on dlPFC: an enhancement of working memory (Fregni et al., 2005a) and a reduction in risk-taking (Fecteau et al., 2007a). I was unable to extract effect size estimates from three other tDCS studies on the dlPFC (Kincses et al., 2004; Boggio et al., 2007; Fecteau et al., 2007b). For purposes of the power analyses I assumed that a tDCS effect on model-based control has an effect size, expressed in Cohen's d, similar to these studies. Our power to detect this effect, given a two-tailed alpha of 0.05 and sample size of 22, was then at least 0.60 (Figure 7.3). Although this is not as high as the normative power of 0.80, it is considerably higher than many studies in cognitive neuroscience (Button et al., 2013). However, to support my claim that tDCS to right dlPFC does not affect model-based and model-free control I formally tested this hypothesis in a model comparison.

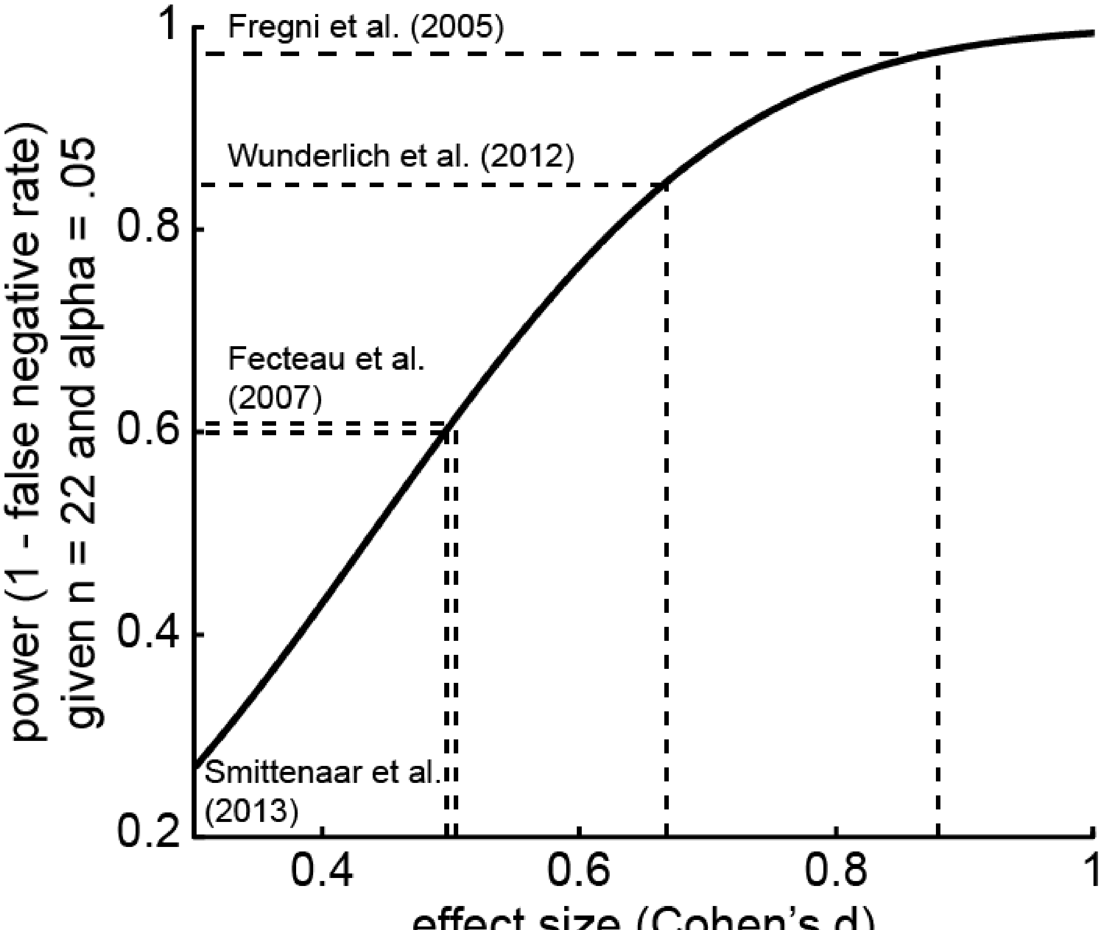


Figure 7.3: Statistical power to detect true effects. I estimated statistical power in the study based on effect size estimates taken from the published literature. I could then compute the power in the study based on 22 participants and a false positive rate of 0.05 (two-sided alpha). Assuming any true effect of tDCS would have a similar magnitude as the studies shown in the figure, the current study had a power of 50-80%.

The analyses presented above rely on a frequentist approach and hence are framed in terms of null hypothesis testing, which precludes strong conclusions being drawn about the absence of an experimental effect. Hence, based on the preceding analyses I cannot decisively conclude that the null model is more likely compared to the full model that allows for differences in model-free or model-based control in Active versus Sham conditions. Bayesian statistics, by contrast, allow inferences to be made about the absence of experimental effects, and I thus exploited this approach to further probe the results. Thus, I fit three models to the data that were identical to the full model, except that the model-free and/or model-based regressors were assumed identical between stimulation conditions. The first model contained a single set of model-free regressors for both stimulation conditions; the second contained a single set of model-based regressors for both stimulation conditions; and the third ('null') contained a single set of model-based and a single set of model-free regressors for both stimulation conditions (see Table 7.3 for the regressors in the null model).

Table 7.3: Regressors in the null model which contains the same MB and MF regressors for the Active and Sham stimulation conditions. MF = model-free; MB = model-based; SE = standard error. Lag denotes the effect of time. Bold-face indicates p < .05 uncorrected for multiple comparisons.

regressor	**estimate**	**SE**	**z-value**	**p**
intercept	0.24	0.03	7.78	**<0.0001**
Active	-269.68	179.42	-1.50	0.1328
MF Lag-1	332.27	48.71	6.82	**<0.0001**
MF Lag-2	285.59	43.58	6.55	**<0.0001**
MF Lag-3	208.50	48.78	4.27	**<0.0001**
MB Lag-1	234.64	61.35	3.82	**0.0001**
MB Lag-2	194.46	64.68	3.01	**0.0026**
MB Lag-3	180.81	45.37	3.99	**0.0001**
Active transition Lag -1	-11.12	35.88	-0.31	0.7566
Active transition Lag -2	7.89	31.01	0.25	0.7993
Active transition Lag -3	-20.15	33.11	-0.61	0.5428
Sham transition Lag -1	0.98	40.73	0.02	0.9809
Sham transition Lag -2	15.32	32.40	0.47	0.6365
Sham transition Lag -3	2.99	35.05	0.09	0.9320

I then performed Bayesian model selection using the Bayesian Information Criterion (BIC) and Aikaike Information Criterion (AIC) that are returned by the lme4 package for each model (Table 7.4). Although derived within different frameworks, both the BIC and AIC can be thought of as approximations to the true model evidence (Penny, 2012), both containing a term reflecting the likelihood of the model given the data (the ‘accuracy’ term) and a penalization term reflecting the number of parameters in the model (the ‘complexity’ term). As such, the difference in the values of the Information Criteria between models approximates the log Bayes factor, which is the ratio of probabilities of the model given the data. The BIC difference was 900 in favour of the null model when compared to the full model that contains a separate set of model-based and model-free regressors for the Active and Sham condition. This indicates the null model was e^{900} times more likely than the full model. The AIC, which penalises model complexity less harshly than the BIC, was 100 in favour of the null model compared to the full model, i.e. the null model was e^{100} times more likely. I found a similar pattern of results for the model-free clamped and model-based clamped models which were $>e^{29}$ and $>e^{44}$ less likely than the null model, respectively. Therefore I can conclude that it is significantly more likely that tDCS had no effect on model-based or model-free control than that it did.

Table 7.4: Model comparison between a null model (one set of model-based and model-free regressors for both stimulation conditions) and more complex models that allow for an effect of tDCS on model-based control, model-free control, or both, which shows the null model is significantly more plausible than any of the models that allow for an effect of tDCS on behavioural control. The second column refers to the number of regressors in the hierarchical regression at the individual participant level (cf. Table 7.1 and Table 7.3). BIC: Bayesian Information Criterion; AIC: Aikaike's Information Criterion.

model	**No. of regressors per participant**	**BIC**	**ΔBIC**	**AIC**	**ΔAIC**	**Bayes factor in favour of null model based on AIC**
null model	13	18553	0	17752	0	-
separate model-free regressors for Active and Sham	16	18962	409	17796	44	1.3×10^{19}
separate model-based regressors for Active and Sham	16	18947	394	17781	29	3.9×10^{12}
full model	19	19453	900	17852	100	2.7×10^{43}

To test for session effects I performed a hierarchical logistic regression with identical regressors as those described in Table 7.1, but instead of Active and Sham I coded the regressors as session 1 and 2, respectively. The equivalent contrasts to Table 2 were all $p > .15$ except effect for Lag on MF in session 1, $p = .003$, and session 2, $p = .06$. This suggests that model-based and model-free control do not change with additional exposure to the task, which replicates previous chapters (Wunderlich et al., 2012a; Smittenaar et al., 2013b).

Both model-based and model-free control make equivalent predictions for second-stage choices as there is no task structure to exploit. I nevertheless explored the effects of stimulation on 1-step reward learning. I examined second-stage choices using hierarchical logistic regression similar to the analysis of first-stage choices: stay-switch behaviour was regressed against reward received on the most recent trial involving that second-stage pair (i.e. lag-1 only). Transition was not included as a factor because second-stage choices are assumed to be independent of the transition type that led to the state. I observed that in both stimulation conditions there was a main effect of reward, such that if a particular stimulus was rewarded in the most recent encounter with that second-stage pair it was more likely to be chosen again (Active, mean ± SE = 0.96 ± 0.13, $p = 9.4 \times 10^{-13}$; Sham, mean ± SE = 0.82 ± 0.11, $p = 5.46 \times 10^{-13}$). There was a trend-level effect of stimulation-by-reward suggesting a stronger influence of reward under Active stimulation (mean ± SE difference = 0.14 ± 0.08; $p = .07$), but given the large amount of statistical tests performed I do not further consider this marginal effect. Together, these results suggest stimulation had no effect on second-stage choices.

7.5 Discussion

Here I provide evidence that tDCS to right dlPFC does not affect model-based or model-free control in an established behavioural paradigm. In a double-blind design I confirmed that participants used both model-free and model-based strategies to solve the task, and I could quantify the extent to which either strategy was used. A putative enhancement of right dlPFC activity through Active compared to Sham anodal tDCS stimulation did not significantly change the level of model-based or model-free control. Formally testing this null effect, I provide evidence that a null model predicting no effect of stimulation performed significantly better than more complex models predicting an effect of stimulation on model-based control, model-free control, or both.

I hypothesised that an enhancement of right dlPFC would improve model-based control, similar to beneficial tDCS effects observed on risk taking (Fecteau et al., 2007a), probabilistic learning (Kincses et al., 2004) and working memory (Fregni et al., 2005a). Based on published tDCS studies and studies of model-based control, I estimated this study had more than 60% statistical power to detect such an effect were it to exist. Although the power was potentially lower than the often cited 80% power standard (e.g. Cohen, 1992), it was considerably higher than >75% of neuroscience studies as determined recently in a meta-analysis (Button et al., 2013). Despite this, I observed a null effect of tDCS on model-based control. However, frequentist statistics do not allow me to conclude the null hypothesis was a significantly better explanation than the alternatives in which stimulation does have an effect. I therefore performed a complementary model comparison using information-theoretic measures to formally show this (Stephens et al., 2005). Together, these analyses support the conclusion that tDCS to right dlPFC has no effect on model-based or model-free control.

There is a modest literature on improvement in cognition through tDCS of the right dlPFC, and this begs the question why no effect was found in this experiment. This is even more surprising because the dlPFC is implicated in model-based processes (Lee and Seo, 2007; Gläscher et al., 2010; Abe and Lee, 2011; Wunderlich et al., 2012b; Xue et al., 2012) and when the region is transiently disrupted using transcranial magnetic stimulation, model-based control is selectively impaired (Smittenaar et al., 2013b). Here I speculate that the null result is most likely due to an inability of tDCS to improve the specific component processes of model-based control subserved by the dlPFC.

Firstly, little is known about the physiological effects of tDCS in prefrontal cortex (Stagg and Nitsche, 2011), though this is a rapidly developing field (Stagg et al., 2013). While there is evidence that anodal stimulation over M1 increases the motor evoked potential (MEP) size elicited by TMS (Nitsche and Paulus, 2000), it is not clear how the cellular physiology of the dlPFC is changed following anodal stimulation, nor what the physiological underpinnings of model-based control in the dlPFC are. Despite these unknowns, I suggest here that the neural mechanisms for model-based control in right dlPFC are not amenable to improvement through anodal tDCS.

Secondly, I used a task to assess model-based control that has previously been shown to be susceptible to manipulation (Wunderlich et al., 2012a; Otto et al., 2013; Smittenaar et al., 2013b). I used a set of stimulation parameters that are widely used in the tDCS community (Nitsche et al., 2008), and replicated previous observations of dual control by model-based and model-free systems. Together, this suggests the null result is not due to the introduction of uncertain elements (e.g. novel task or novel stimulation parameters) into the study design.

Despite the use of established methods, I cannot exclude methodological issues as the cause of the null effect altogether. Although I am confident the null effect is not due to faulty equipment or errors in the double-blinding procedure (see Methods), potential other issues might include inaccurate electrode placement, a problem that can be alleviated by stereotactic navigation using anatomical scans as commonly used in transcranial magnetic stimulation, and unpredictable current flow based on electrode placement, which might be alleviated by computational models of current flow (Wagner et al., 2007).

I was particularly careful to employ a double-blinded design to eliminate any stimulation-dependent influence from the experimenter on task performance. The task used here requires relatively extensive involvement of the experimenter in the task instructions. In a double-blinded design, these effects can be most reliably attributed to the experimental manipulation of interest rather than to unintended information biases (Schulz and Grimes, 2002). I note that no published work has manipulated the instruction of the 2-step task to examine its influence on model-based and model-free performance.

In conclusion, I provide evidence that anodal stimulation of the right dlPFC by tDCS does not alter model-based or model-free control in the paradigm. This observation was made in the context of extensive and causal evidence for a role of right dlPFC in model-based control in humans. As such, my results should not be interpreted as providing evidence that the right dlPFC is *not* involved in model-based control; rather, my main finding is that anodal stimulation does not necessarily enhance this function. An open question is whether tDCS might improve performance on tasks that are more taxing on the model-based system (e.g. Huys et al., 2012).

8 Predicting striatal reward signals from corticostriatal connectivity

8.1 Abstract

A defining feature of the basal ganglia structures are their anatomical organization into multiple corticostriatal loops, which themselves subdivide into direct and indirect pathways in the basal ganglia. A central tenet of this framework is that local striatal function is determined by its connectivity with cortex, which creates the functional topography that is mirrored within cortex and striatum. In this chapter I formally test this notion by asking whether it is possible to leverage the information contained in corticostriatal anatomical connectivity to predict local function of the striatum in a reinforcement learning task. Using high-resolution functional and diffusion MRI, combined with leave-one-out cross-validation methods, I show that connectivity profiles can indeed predict reward and action value signals in the caudate nucleus. I then describe the cortical regions that contribute most strongly to this prediction. Future work can explore in more detail the precise mechanisms by which structural connectivity between the striatum and specific cortical regions predict functional activity, including studying functional representations across the corticostriatal network.

8.2 Introduction

In chapter 2 I went into some detail explaining the anatomical layout of the basal ganglia, including the remarkable parallel corticostriatal loops that comprise its defining structural and functional feature at the macro-scale (Alexander et al., 1986; Haber, 2003). We can take a more abstract view of the brain and consider that the function of any neural region, and indeed any neuron, is to a large extent governed by its inputs. This has led to the prediction that knowledge of the ‘connectivity fingerprint’ of the brain is sufficient to predict its ‘functional fingerprint’ (Passingham et al., 2002). This notion was most directly tested in a combined fMRI/DTI study by Saygin et al. (2012). The authors predicted functional responses to face stimuli for individual voxels in the fusiform gyrus from connectivity fingerprints of these same voxels. Critically, functional responses for an individual participant were better predicted by a connectivity fingerprint than by the average functional response of the group. Although this approach has been extended to other functions of the visual system (Osher et al., 2015), it has not been applied to higher cognitive functions or subcortical structures.

In this chapter I ask whether functional responses across the striatum during value-based learning show a reliable relationship with anatomical inputs from cortex (Haber and Behrens, 2014). The striatum has been suggested to serve as a focal point for associative, reward and motor information, though with each input defining only partially overlapping functional regions (Haber et al., 2006). Instrumental learning is widely accepted to engage the striatum both in animal models (Samejima et al., 2005; e.g. Lau and Glimcher, 2007; Samejima and Doya, 2007) and human studies (e.g. O'Doherty et al., 2004; Wunderlich et al., 2012b). However, the relationship between function and structural connectivity has not been explored in detail in humans, instead there has been a relatively exclusive focus on purely structural connectivity of the corticostriatal network (e.g. Leh et al., 2007), or between-participant correlations of structure and function (e.g. corticostriatal connectivity predicts habitual versus goal-directed control; de Wit et al., 2012a). In this chapter I use a method applied between-voxels that attempts to understand what makes some parts of the striatum respond differently from other parts based on structural connectivity fingerprints.

I predicted that reward signals associated with action values during choice as well as rewards and expected value at the time of outcome can be predicted from cortical connectivity fingerprints for voxels in the caudate nucleus.

8.3 Methods

8.3.1 Participants

Twenty-four adults participated in the experiment (14 female; age range 18-36 years; mean ± SD = 22.5 ± 4.5 years). All participants were right hand dominant, had no history of psychiatric or neurological disorder, were not taking any medication known to affect neural or cognitive function, had normal or corrected-to-normal vision and passed the safety requirements to enter a MRI scanner. All participants provided written informed consent prior to the start of the experiment, which was approved by the Research Ethics Committee at University College London (UK). One further participant was excluded due to excessive movement (images could not be realigned successfully).

8.3.2 Overview of the approach

The goal of the experiment was to test the notion that corticostriatal input governs representations of action values, reward and expected value in the striatum. To understand the link between the anatomical pathways and their contribution to reinforcement learning I estimated for each voxel in the striatum its functional response to reward and expected values. These same voxels were characterised in terms of their structural connectivity to 148 cortical regions through diffusion imaging techniques. This allowed a prediction of functional activation from structural connectivity with the cortex. All these analyses were performed in participant space, with only summary statistics for each participant taken to the group level (see Figure 8.1 for overview of study design).

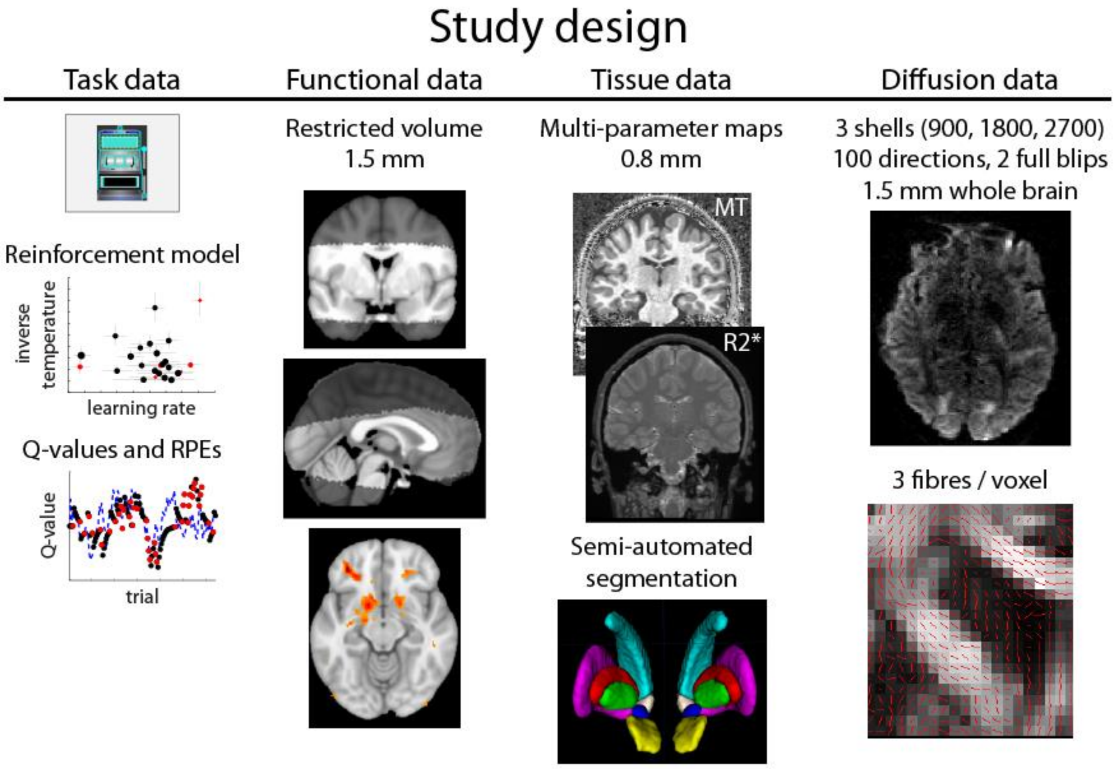


Figure 8.1: overview of the acquired data and processing steps central to this chapter. Each of these steps is further expounded in the methods section.

8.3.3 Task

The task required a participant to track stimulus-specific action values in order to probe how these action values are represented and updated in neural structures during choice and feedback (Figure 8.2A). Participants had to learn two separate two-armed bandits which were distinguished by their colour (red or blue; see Figure 8.2A). On each trial, one of these two slot machines was presented to the participant, requiring a response using either right index finger or right ball of the foot. Binomial feedback was then presented which indicated a reward or no-reward. The probability of reward given a bandit s and action a,

$p(r\,|s_i, a_j)$ where $i \in \{1,2\}$ and $j \in \{1,2\}$, changed slowly over trials, forcing participants to keep exploring throughout the experiment in order to maximise the number of rewards obtained.

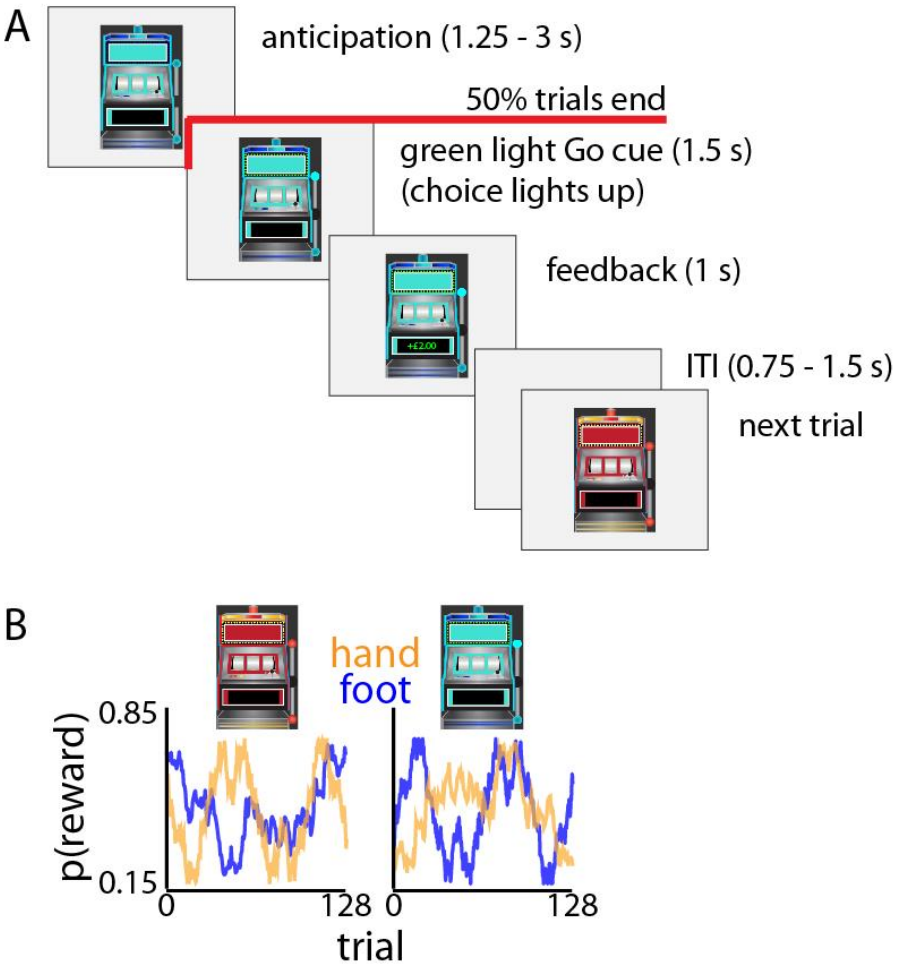


Figure 8.2: reinforcement learning task involving right hand and right foot responses. (A) A single trial consisted of the following sequence: a fixation cross (inter-trial interval) was presented for 750-1500 ms, drawn from a uniform distribution; either the red or blue slot machine was presented for 1250-3000 ms, drawn from a uniform distribution; on half the trials ('abort' trials) the slot machine disappeared and the next trial started; on the other half ('response' trials) lights on the slot machine would turn green, serving as a Go signal; participants responded within 1500 ms by depressing force-sensitive buttons with either their right hand or foot, and upon reaching the force threshold the corresponding lever immediately became brighter until the 1500 ms were up;

feedback was then presented on the slot machine for 1000 ms, consisting of either "+ £2.00" in green, or "+ £0.00" in red. (B) The probability of obtaining the reward varies over time per response per slot machine. This meant participants were required to track 4 random walks that could go between 0.15 and 0.85.

Participants performed 512 trials (approximately 42 minutes) consisting 128 red-abort, red-response, blue-abort, and blue-response trials each (Figure 8.2A). The order of these four trial types was randomly determined and only constrained such that no trial type occurred for more than 3 trials in a row.

Participants came in 1 to 20 days before the scanning session to perform a full set of 512 trials (mean ± SD = 7 ± 4.4 days). A different set of reward probabilities was used each day but otherwise the parameters of the experiment were identical. Participants could also use the training session to get used to the foot and hand force buttons.

8.3.3.1 Fixed reward walks

The $p(r\ |s_i, a_j, t)$, where *t* indicates trial number, was generated by a Gaussian random walk for each action *a* and stimulus *s* as follows:

$$p(r\ |s_i, a_j, t+1) = p(r\ |s_i, a_j, t) + N(0{,}0.01)$$

where for the first trial the probability was randomly drawn from *U(0.15,0.85)*. The walks were not generated anew for each participant—rather, one set of two pairs was used for each participant's practice, and one set was used for each participant's scanning session. However, the assignment of these two pairs to the red and blue slot machine was randomised, and the subsequent assignment of random walk to the two available actions was also randomised. This meant that volatility and availability of reward were matched between participants. The walks were constrained in their upper (0.85) and lower (0.15) values and in their mean value (between 0.4 and 0.6; see Figure 8.2B). The highest correlation between any two of the four walks was 0.38, forcing participants to learn about the value of each option through trial-and-error rather than inferring the value of choice options based on a level of correlation between the walks.

8.3.3.2 Cancelling half the trials

Examining value representations in the BOLD signal at both choice and outcome phase can be challenging due to the sluggishness of the BOLD response (see section 3.2.2), and the resulting correlated regressors in the design matrix if the choice and feedback are presented close together in time. I considered two options to minimise this potential confound: a slow design where choice and feedback events are separated by at least 8 s (e.g. Behrens et al., 2008), and a fast design in which half the trials are cancelled at any point between choice and feedback phase (e.g. Guitart-Masip et al., 2011). Pilot data with both designs (data not shown) suggested participants were more accurate at learning reward probabilities in the fast design, possibly due to disengagement from the task when participants are faced with long pauses. Also, a slow design might lead to non-striatal learning mechanisms dominating behaviour, whereas I was specifically interested in such striatal mechanisms (Foerde et al., 2012). I thus opted for the fast design.

8.3.4 Reinforcement learning models

I used temporal difference (TD) reinforcement learning models as described in chapter 2 to model participants' behaviour and estimate quantities that might be represented in the BOLD signal in the striatum, most notably rewards and action values. Each slot machine *i* defines a state s_i where two actions a_j are available. The reward *r* on trial *t* can be either 0 or 1. The value of action *j* in state *i* is updated after feedback by:

$$Q_{s_i,a_j}(t+1) = Q_{s_i,a_j}(t) + \alpha * \partial(t)$$

where α = 0 for all states and actions that did not occur on trial *t*-1. As the reward probabilities change independently for each state and action, the participant only learns about the chosen action in the current state, rather than inferring changes in value for non-chosen state-action pairs in a ‘model-based’ way (except for value decay—see below). $\partial(t)$ represents the RPE at trial *t*, defined as

$$\partial(t) = r(t) - Q_{s_i,a_j}(t)$$

The probability of each action given these cached values Q are then given by the softmax equation with inverse temperature β:

$$p(a_j \mid s_i) = e^{\beta * Q_{s_i,a_j}} \Big/ \sum_{k=1}^{2} e^{\beta * Q_{s_i,a_k}}$$

I used an expectation maximization (EM) approach as implemented in Guitart-Masip et al. (2012) to simultaneously fit parameters at the level of participants and population.

In addition to this basic model with a learning rate and inverse temperature I examined a number of more complex models that might provide a better explanation for the data. For each of these models I estimated the negative log-likelihood and Bayesian Information Criterion (BIC) to select the model that optimally described the participant’s behaviour on this task. The additional parameters are described in Table 8.1. All parameter combinations were tested.

Table 8.1: additional parameters for the reinforcement learning model.

Parameter name	**Description**
Negative learning rate	Separate learning rate for negative and positive feedback
Effector bias	A fixed bias towards hand or foot responses
Lapse rate	A value that constrains the softmax between ε and 1-ε rather than 0 and 1 to account for occasional lapses
Decay	Implements the notion that unsampled actions do not maintain their value but decay back to 0.5. The parameter describes the time constant of exponential decay.
Perseverance	A tendency to stick with the same action for a given stimulus, irrespective of value.

8.3.5 Magnetic resonance imaging

For each participant I acquired 1.5 mm isotropic restricted volume echo-planar imaging (EPI) data during task performance, 0.8 mm isotropic whole-brain multi-parameter maps (MPMs) consisting of a T1-, proton density- and magnetisation transfer-weighted volume, 1.5 mm isotropic whole-brain diffusion weighted images, 1.1 mm restricted volume diffusion weighted images, and B0 field maps to correct for field inhomogeneity for the EPI data. The parameters of these scans are detailed in Table 8.2. I also acquired a single whole-brain volume using otherwise identical settings for the EPI sequence. Cardiac rate was recorded using an MRI-compatible pulse oximeter (Model 8600 F0, Nonin Medical), and respiration was monitored using a pneumatic belt positioned around the abdomen. I processed these data as described in the literature (Hutton et al., 2011) and included them as regressors of no interest in the first-level general linear models (see below).

Table 8.2: MRI acquisition parameters.

Sequence	**Parameters**
B0 field map	Double echo FLASH sequence (matrix size = 64 x 64; 64 slices; spatial resolution = 3 x 3 x 3 mm^3; gap = 1 mm; short TE = 10 ms; long TE = 12.46 ms; TR = 1020 ms) to correct EPI images for distortion in the B0 field (Weiskopf et al., 2006).
Functional, EPI	Restricted volume, 44 slices (40 in slab with 10% oversampling), FoV read 192 mm, transverse slices tilted 20 degrees, anterior-posterior phase encoding, 12% phase oversampling, 10% slice oversampling, 40 slices per slab, voxel size 1.5 mm isotropic, TR = 78 ms, TE = 37.3, GRAPPA2 along phase encoding (144 PE ref. lines, 44 3D ref. lines), 180-185 volumes per block depending on duration of block over 4 ~10 min blocks in total.
Multi-parameter maps	Proton density (PD)-weighted, T1-weighted, and magnetization transfer (MT)-weighted images at 0.8 mm isotropic resolution for each participant using multi-echo 3D FLASH (Helms et al., 2008a). A B1-map was acquired using a 3D SE/STE EPI method (Lutti et al., 2012) to correct for the effects of inhomogeneous radio-frequency excitation on the quantitative maps. Total time of acquisition was ~40 min.
Diffusion-weighted, whole-brain	Whole-brain 1.5x1.5x1.5 mm^3 resolution diffusion-weighted images with settings similar to the Human Connectome Project (Van Essen et al., 2012; Sotiropoulos et al., 2013). Three shells (b=900/1800/2700) for both right-left and left-right phase-encoding directions. Each of these 6 scans contained 10 images with no diffusion weighting (b=0) and 100 directions spread out over a full sphere. I used multiband 3 but no further acceleration. Acquisition time was 10 min 20 s for each of the 6 scans. No phase oversampling, 75 transverse slices, FoV read 192 mm, FoV phase 100%, slice thickness 1.5 mm with 0 distance between slices, TR 5440, TE 130 ms. I additionally acquired a single b0 image with identical settings, but phase encoding along anterior-posterior and along posterior-anterior. These additional phase encoding directions should aid in estimating distortions due to distortions along the phase encoding direction.
Diffusion-weighted, restricted volume	These images were acquired but not further analysed in this chapter. 47 slices, distance factor 10%, transverse orientation, anterior-posterior phase-encoding, 35% phase oversampling, FoV read 156 mm, FoV phase 40.8 %, 1.1 mm slice thickness, TR = 7200 ms, TE = 87.6 mm, b = 900, 100 directions over full sphere, 10 b0 images interspersed, acquisition time 13 min 19s per scan, two averages acquired. Additional single b0 image acquired with posterior-anterior phase encoding to correct for distortions along the phase-encoding direction, otherwise identical parameters.

8.3.5.1 Multi-parameter maps processing

Fully quantitative maps of the MR parameters MT, R1, PD and R2* were extracted from the acquired data as described previously (Helms et al., 2008a). I extracted a brain mask in structural space from the T1w image using BET implemented in FSL (Smith, 2002).

8.3.5.2 Semi-automatic segmentation of basal ganglia substructures

Whereas the striatum can be reasonably defined using automated algorithms, other parts of the basal ganglia require manual segmentation. These were the globus pallidus pars interna (GPi) and externa (GPe), subthalamic nucleus (STN) and substantia nigra and ventral tegmental area (SN/VTA). I used FSL FIRST to automatically segment the bilateral caudate and putamen (Patenaude et al., 2011), and ITK-SNAP to segment the remaining regions (Yushkevich et al.). Note that segmentation was performed bilaterally for each participant as it is unclear to what extent basal ganglia function is lateralised (e.g. Scholz et al., 2000).

8.3.5.3 Automatic segmentation of cortex using FreeSurfer

To obtain cortical targets for tractography I used FreeSurfer's RECON-ALL pipeline to generate 148 cortical labels in structural (participant) space following the Destrieux atlas (Destrieux et al., 2010; Fischl, 2012). These were transformed into volumetric ROIs. Two participants lacked 1 and 3 labels, respectively, so these were added as empty ROIs for tractography (see below). The FreeSurfer segmentation pipeline has been described in detail elsewhere (Fischl et al., 2004).

8.3.5.4 FMRI preprocessing

I analysed the fMRI data in SPM8 (Wellcome Trust Centre for Neuroimaging, UCL, London; www.fil.ion.ucl.ac.uk/spm). The images were corrected for signal bias at low spatial frequencies, realigned to the first functional image and distortion corrected using the B0 field maps. The first functional image was coregistered to the MT image for its superior subcortical performance in white- and grey-matter segmentation compared to T1-weighted images (Helms et al., 2009) and these transformation parameters were then applied to all restricted-volume functional images to bring them into structural space. Notably, SPM's coregistration of the restricted-volume EPI to the MT image worked well, obviating the need for an intermediate step involving the whole-brain EPI images. For additional analyses of group-level responses I applied normalization parameters to the functional images to bring them into MNI space and applied a 6 mm full-width-half-maximum (FWHM) smoothing kernel. All participant-level statistics were performed on voxels within an explicit mask (rather than the more commonly used implicit mask) to prevent brain voxels with low signal from being excluded. The explicit mask for structural (i.e. native) space was constructed by restricting the whole-brain mask (see multi-parameter maps) to the volume of the EPI sequence using SPM's IMCALC.

8.3.5.5 FMRI general linear model

The preprocessed images were analysed in an event-related design using a general linear model (GLM). The first model contained 8 explanatory variables of interest (EVs) defined at the onset of the visual stimulus (2 identical EVs), the 'go' cue when choosing hand (1 EV) or foot (1 EV), the onset of feedback after choosing hand (2 identical EVs), and the onset of feedback after choosing foot

(2 identical EVs). A number of identical EVs were entered to be able to add multiple, non-orthogonalised parametric modulators to specific events. These parametric modulators were the Q-value for the hand and foot at visual stimulus; the Q-value for the hand and foot on the respective response EVs, and whether reward was received for the respective feedback EVs.

I added the following nuisance regressors: 1 regressor for trials where no response was recorded in the 1500 ms response window, 1 regressor when the trial was aborted, 6 movement regressors produced by the realignment procedure, 14 physiological regressors for cardiac and respiratory variables (Hutton et al., 2011), and 3 block regressors covering run 1 to 3, respectively. The 4th block was subsumed in the constant of the design matrix. The GLM was estimated separately for each participant. All EVs (but not physiological regressors) were convolved with a canonical haemodynamic response function (Friston et al., 1995).

8.3.5.6 Diffusion weighted imaging preprocessing

The diffusion data was preprocessed using FSL (Smith et al., 2004). I estimated the distortions along phase-encoding directions by entering 8 b0 images into TOPUP (1 from each of 2 blips * 3 shells + 1 AP + 1 PA blip; Andersson et al., 2003). The field coefficients were then supplied to EDDY, which corrects for the phase-encoding distortion, movement, and eddy currents in all 660 volumes (3 shells * 2 phase-encoding directions * 110 images each). The corrected b0 volume from TOPUP was entered into BET to obtain a brain mask. I used DTIFIT to estimate fractional anisotropy (FA) maps and BEDPOSTX to estimate up to three fibres per voxel using custom settings for multishell data (Behrens et al.; Behrens et al., 2007b; Jbabdi et al., 2012).

8.3.5.7 Probabilistic tractography

I used PROBTRACKX2 implemented in FSL to estimate connectivity profiles for each 0.8 mm isotropic voxel in the striatum. Each voxel was seeded with 10k streamlines and standard parameter settings. I then extracted connectivity profiles for voxels at coordinates specified by the anatomical masks. The locations of these voxels were recorded and used later to extract functional signals from identical locations.

8.3.6 Relating structure to function

For the left and right caudate I extracted functional signals for the reward, Q-value at choice, and Q-value at outcome contrasts at voxel locations identical to the diffusion data. I then used a leave-one-out cross-validation (LOOCV) approach to predict functional activation in participant n based on the relationship between structure and function in participants n-1 (Figure 8.3). All functional data were smoothed at 6 mm FWHM and z-scored before entering the regression (though leaving the data unsmoothed does not drastically alter results, cf. Saygin et al., 2012). The design matrix for each participant contained 149 columns (1 intercept and 148 target regions) and the number of rows corresponded to the number of voxels in the seed region. Each value indicated the number of samples that reached the target region, z-scored across voxels for each region separately. The dependent variable was each voxel’s functional response to a contrast, also z-scored. The regression coefficients for n-1 participants were averaged and used to predict each voxel’s functional response in participant n based on its connectivity profile. Each voxel had some error in its predicted value, and the mean absolute error (MAE) was calculated for each participant (Saygin et al., 2012). This was used as a standardised measure of predictive capacity. I also performed an identical analysis to the connectivity LOOCV approach, but instead randomly permuted the regression coefficients before estimating functional signals for the n-th participant. By doing this permutation 10k times I built up a null distribution for comparison against the true connectivity model.

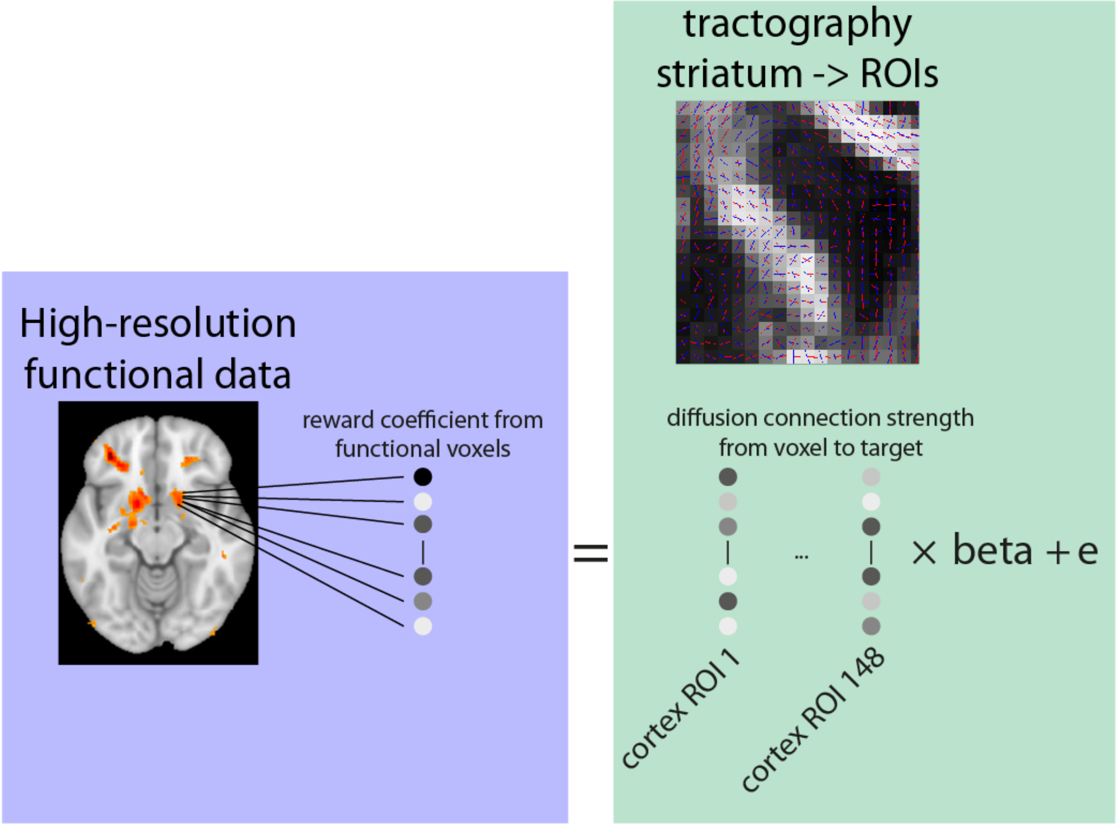

High-resolution functional data
reward coefficient from functional voxels
=
tractography striatum -> ROIs
diffusion connection strength from voxel to target
...
× beta + e
cortex ROI 1
cortex ROI 148

Figure 8.3: overview of regression approach. The betas are estimated based on n-1 participants and used to predict the reward signals in participant n. The mean absolute error (MAE) of the prediction is recorded and the approach is repeated for every participant. In this example the reward response is predicted, and I also used this method on action values at the time of choice and expected values at the time of outcome.

8.4 Results

8.4.1.1 Reinforcement learning model

The model comparison revealed that a separate learning rate for negative feedback and a decay parameter for unchosen values are consistently present in the best models (Table 8.3). Adding further parameters did not yield sufficient improvements to warrant additional complexity, such that the winning model was a four-parameter model including two learning rates, an inverse temperature and a decay rate. Table 8.4 shows descriptive statistics for the parameters fit to the behavioural data from the scanning session.

Table 8.3: model comparison results with only the five best models shown here. Each reinforcement learning model had a single learning rate and inverse temperature parameter. Added to this base model was perseverance, effector bias, separate learning rate for positive and negative feedback ('neg α'), a lapse rate, and exponential decay for unchosen options back to $Q = 0.5$*. The integrated Bayesian Information Criterion was estimated for 200k samples each from the practice and scanning session, and summed over both sessions and participants to arrive at final BICi. For details of this approach see Guitart-Masip et al. (2012).*

Additional parameters	**BICi**	**δBICi**
neg α, decay	12393	0
perseverance, neg α, decay	12400	+7
lapse rate, neg α, decay	12427	+34
perseverance, lapse rate, neg α, decay	12435	+42

Table 8.4: parameter estimates from winning model for the scanning session.

Parameter	**25th percentile**	**median**	**75th percentile**
Positive learning rate	0.54	0.61	0.72
Negative learning rate	0.20	0.32	0.38
Inverse temperature	3.12	5.01	5.87
Decay	0.36	0.55	0.73

8.4.1.2 Semi-automated segmentation

The volumes for the segmented basal ganglia structures are presented in Table 8.5. These values are compared to values from the literature, which shows no discrepancies. Figure 8.4 shows, for illustration purposes, a thresholded probabilistic map of normalised ROIs.

Table 8.5: Average region of interest volumes. Values indicate volume of each structure averaged over left and right, with 95% CI across participants. [1]

Keuken et al. (2014), [2] approx. values from Lenglet et al. (2012), [3] average value between left and right from Ahsan et al. (2007).

Structure	**volume (mm^3) ± 95% CI**	**volumes (mm^3) from literature ± SD**
Caudate	3679 ± 153	[2] 4.1e3, [3] 4102
Putamen	4796 ± 233	[2] 4.5e3, [3] 4615
Accumbens	460 ± 33	[3] 341
Globus pallidus pars externa	1152 ± 42	[1] 918 ± 123, [2] 1.2e3
Globus pallidus pars interna	532 ± 31	[1] 366 ± 60, [1] 405 ± 68, [2] 0.7e3
Subthalamic nucleus	81 ± 5	[1] 56 ±16
substantia nigra & ventral tegmental area	490 ± 24	[3] 373

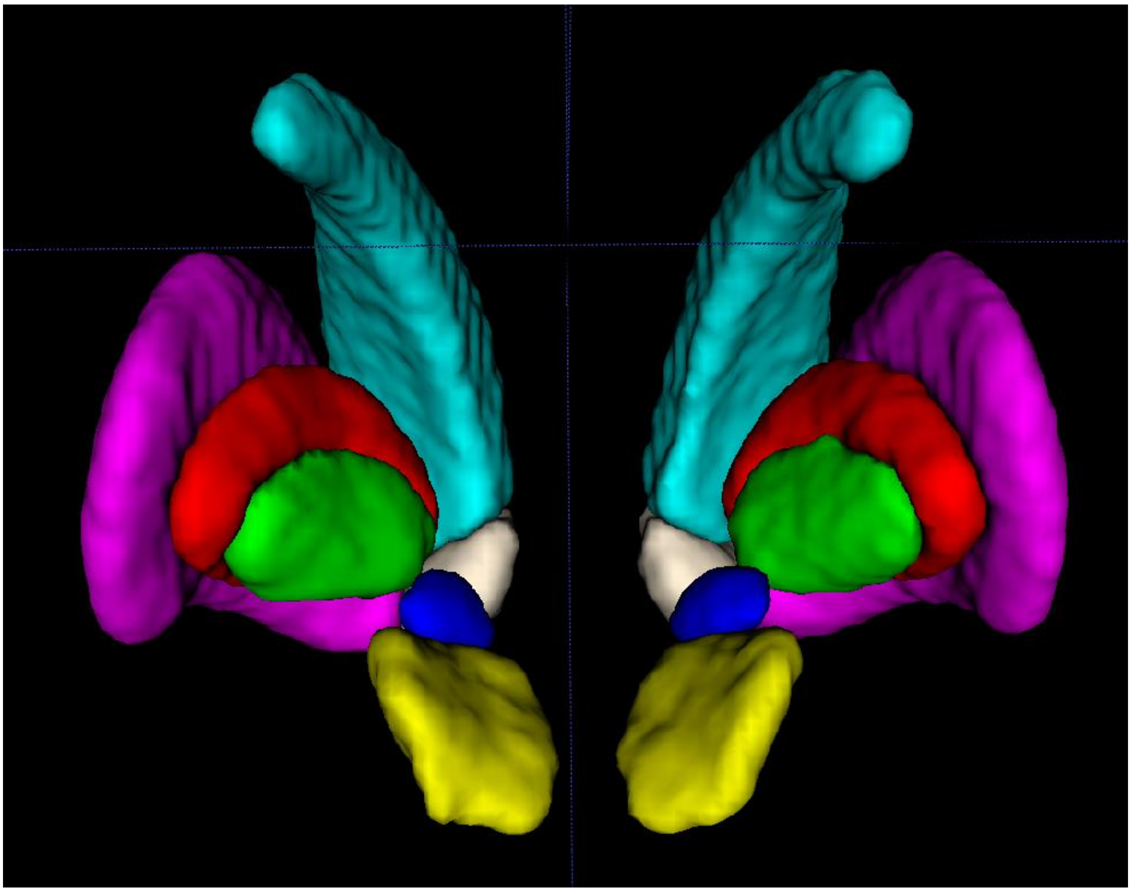


Figure 8.4: normalised ROIs thresholded at p = 0.27 viewed from caudal looking rostral. Cyan = caudate nucleus; pink = putamen; beige = accumbens; red = globus pallidus pars externa; green = globus pallidus pars interna; yellow = substantia nigra and ventral tegmental area; blue = subthalamic nucleus.

8.4.1.3 Functional results in ROIs

The extracted contrast values from the anatomical ROIs in each participant's native space showed a largely familiar pattern of reward and value signals (Figure 8.5). Action values at the time of choice (irrespective of actual choice), if represented, did not survive averaging across the ROI, with only the caudate showing a weak signal at p = .08. The reward prediction error is calculated as reward - expectation, and a region representing the RPE should thus show a positive effect of reward and negative effect of expectation. This was indeed the case in the putamen and caudate nucleus, whereas the accumbens only showed a reward signal but lacks an expected value signal. None of the other ROIs showed any significant effects, including a null effect for reward and expected value at outcome in the SNVTA (Figure 8.5). Together, this suggests the task was able to elicit value signals similar to those reported previously in the literature.

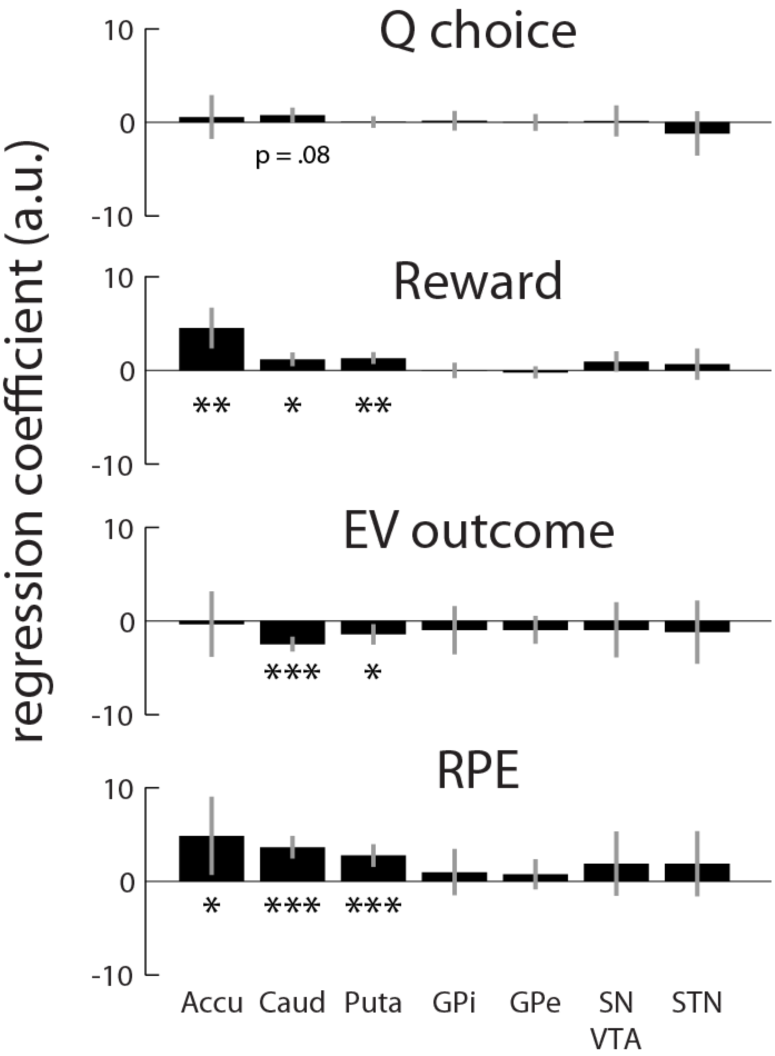


Figure 8.5: extracted betas for four reinforcement learning contrasts. Using anatomically defined regions of interest I extracted regression coefficients in participant space. The Q value at choice, which is the representation of the action value irrespective of choice, shows only a weak effect in the caudate. Significant activations are reward across the striatum and nucleus accumbens; negative expected value at outcome, but only in caudate nucleus and putamen; and reward prediction error again across the striatum and nucleus accumbens. The EV at outcome, which together with reward is what makes up the reward prediction error, is not observed in the accumbens. Reassuringly, the other regions of the basal ganglia show no value-related signals, despite e.g. the GPe bordering the putamen directly. This emphasises the specificity of these activations and potentially of the computations performed in these regions. Error

*bars indicate 95% CI. Stars indicate p-values from 1-sample t-test against zero: * < .05, ** < .001, *** < .0001.*

8.4.1.4 Functional-DTI relationship

The aim of this chapter was to predict functional signals in the striatum based on corticostriatal connectivity. As described in the methods I calculated the mean absolute error (MAE) for the connectivity model to compare against the null model from the permutation. The focus here is on the caudate as this structure is most likely to represent the variables of interest and is more amenable to tractography than the putamen due to its shape. As can be seen in Figure 8.6 the connectivity prediction was significantly more accurate than the permuted connectivity prediction (Cohen’s d = 0.46, 1.04 and 0.88 for action values, reward and EV outcome, respectively; all p < .003 in paired t-test). This provides evidence that at the level of a single participant, knowledge of structural connectivity contains information pertaining to the functional signals.

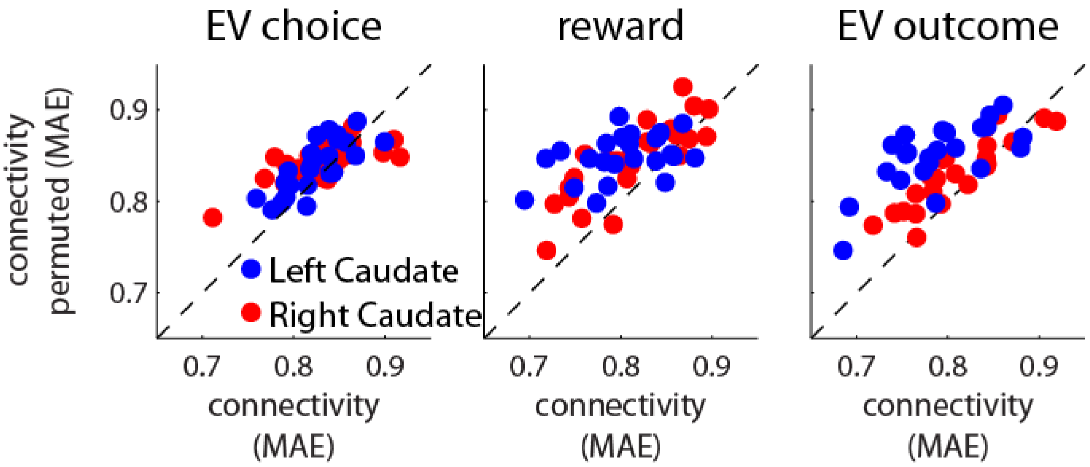


Figure 8.6: Using structural connectivity to predict functional activity. Functional signals in individual voxels related to action values at choice, reward, and expected value at outcome ('EV outcome') were predicted from corticostriatal connectivity in those same voxels. A 10k permutation test in which the regression weights were shuffled before calculating the prediction revealed significantly better predictions by connectivity compared to chance (all p < .003).

I then examined what cortical regions contributed to the functional prediction by testing all regression coefficients against zero across participants. The cortical regions that significantly contributed to functional signals in the right caudate (at p < .05 uncorrected for illustration purposes) are shown in Table 8.6. This statistical test measures the magnitude and reliability of the structure-function relationship across participants. As an example, voxels in the right caudate that are more strongly connected, as measured by probabilistic tractography, to the left middle frontal gyrus have weaker reward responses (Table 8.6). Notably, each individual region contributes only weakly to the prediction as evidenced by none of the regions surviving Bonferroni correction.

Table 8.6: overview of regions most strongly contributing (in terms of absolute standardised coefficient magnitude) to the prediction of the functional contrast. In the LOOCV approach each region has a standardised regression coefficient, i.e. weight, in its prediction of functional activity for each contrast. Data shown here are for the right caudate only. rh = right hemisphere; lh = left hemisphere; G = gyrus; S = sulcus; CI = confidence interval across participants.

Region	**Coefficient**	**lower 95% CI**	**upper 95% CI**
Action values			
rh G Ins lg and S cent ins	0.037	0.010	0.064
lh G temp sup-Lateral	0.015	0.003	0.027
rh S calcarine	0.058	0.012	0.105
rh G temp sup-Plan tempo	-0.040	-0.073	-0.006
rh S front sup	0.047	0.006	0.087
lh G front inf-Triangul	0.056	0.007	0.106
lh Pole occipital	0.030	0.003	0.057
rh S front inf	-0.084	-0.161	-0.006
rh G temporal middle	-0.074	-0.144	-0.005
lh G and S transv frontopol	-0.037	-0.071	-0.002
lh G and S cingul-Ant	-0.039	-0.078	-0.001
Reward			
rh G cingul-Post-ventral	-0.087	-0.128	-0.045
lh G front middle	-0.105	-0.188	-0.023
lh S intrapariet and P trans	-0.024	-0.043	-0.005
rh G and S frontomargin	0.157	0.028	0.286
EV outcome			
rh S collat transv ant	-0.096	-0.161	-0.030
rh S temporal inf	-0.045	-0.076	-0.014
lh S oc sup and transversal	0.013	0.002	0.023
rh G and S cingul-Mid-Ant	0.043	0.007	0.079
lh S occipital ant	-0.005	-0.009	-0.001
lh S orbital med-olfact	0.058	0.007	0.108
lh G front sup	-0.074	-0.139	-0.009
rh G oc-temp lat-fusifor	0.020	0.002	0.038
rh G cingul-Post-ventral	0.047	0.003	0.091
lh G temporal middle	-0.013	-0.025	-0.001

In Figure 8.7 to Figure 8.9 I show all 148 regression weights averaged across participants projected back into the 148 masks used for probabilistic tractography. These figures show the weights for the structure-function relationship in the right caudate. That is, positive values indicate that a voxel in the right caudate has a relatively stronger response to the contrast if it is relatively stronger connected to that part of cortex. I note a number of observations: coefficients for action values at choice are particularly strong in right dorsolateral and bilateral ventromedial PFC; coefficients for reward signals are particularly large in right cingulate cortex and right anterior temporal lobe, and show a negative-to-positive gradient from motor to occipital cortex; and coefficients for EV at outcome are particularly large in ventromedial prefrontal and orbitofrontal cortex. Across all three contrasts connectivity to the dorsal regions of right parietal, sensory and motor cortex seem to show a strong predictive power—sometimes positive, sometimes negative (see slices with high z-coordinates).

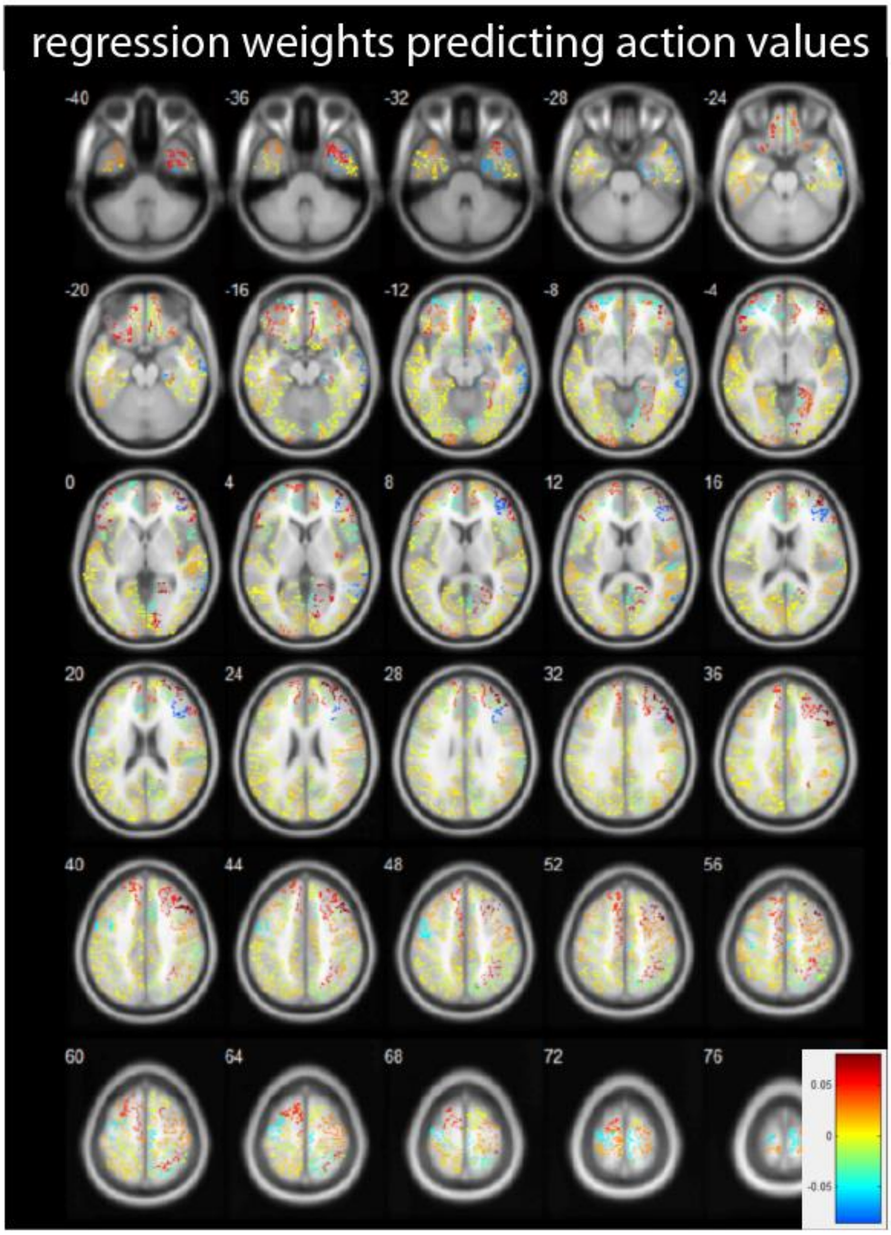


Figure 8.7: regression weights averaged across participants predicting action values signals in the right caudate nucleus from structural connectivity. The regression from Figure 8.6 yields 1 regression coefficient for each of 148 cortical regions. In this figure these weights are averaged across participants and projected onto a normalised set of 148 cortical ROIs, here overlaid onto the MNI152T1 template. Warm colours, such as those in ventral prefrontal and dorsolateral prefrontal cortex, indicate connectivity to these areas is associated with stronger action values responses in those caudate nucleus voxels. The units on the colour bar are standardised regression coefficients, the coordinates are z-coordinates in MNI space.

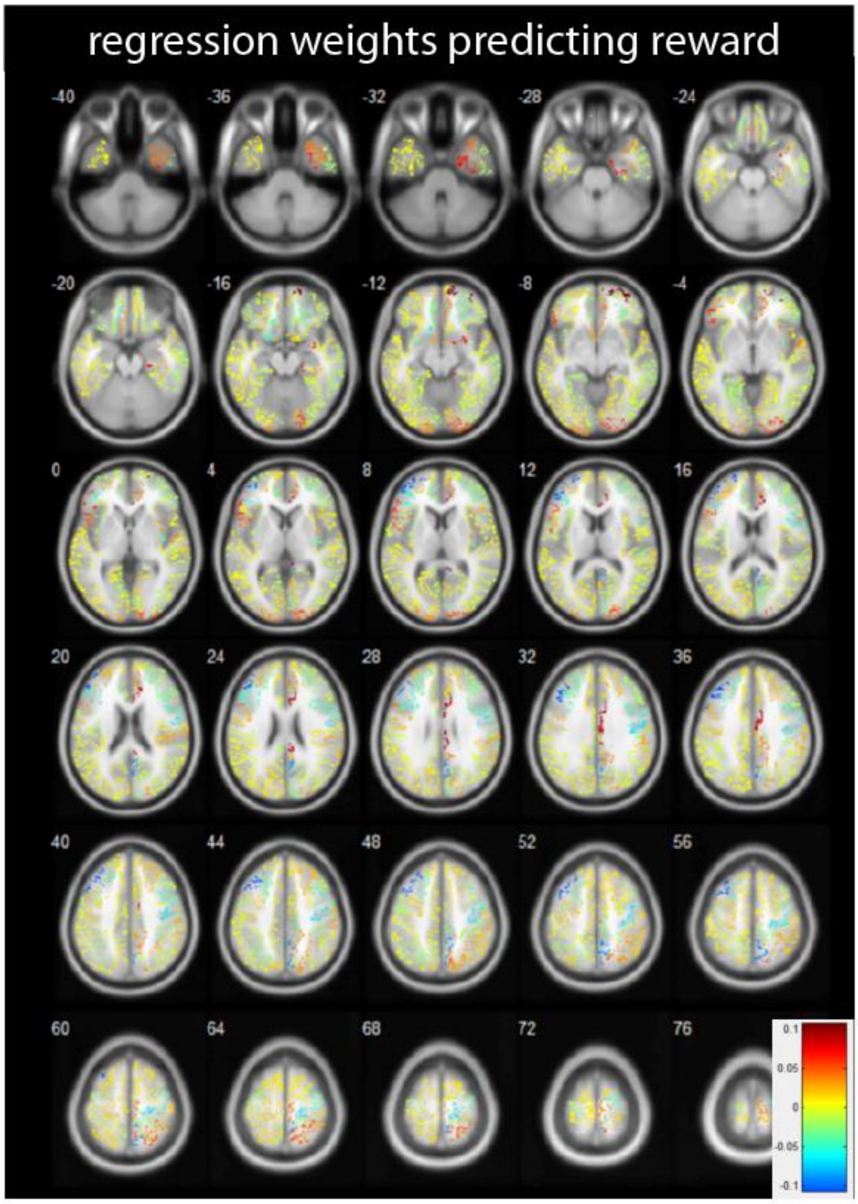


Figure 8.8: structural connectivity regression weights for reward signals in the right caudate. See Figure 8.7 for details.

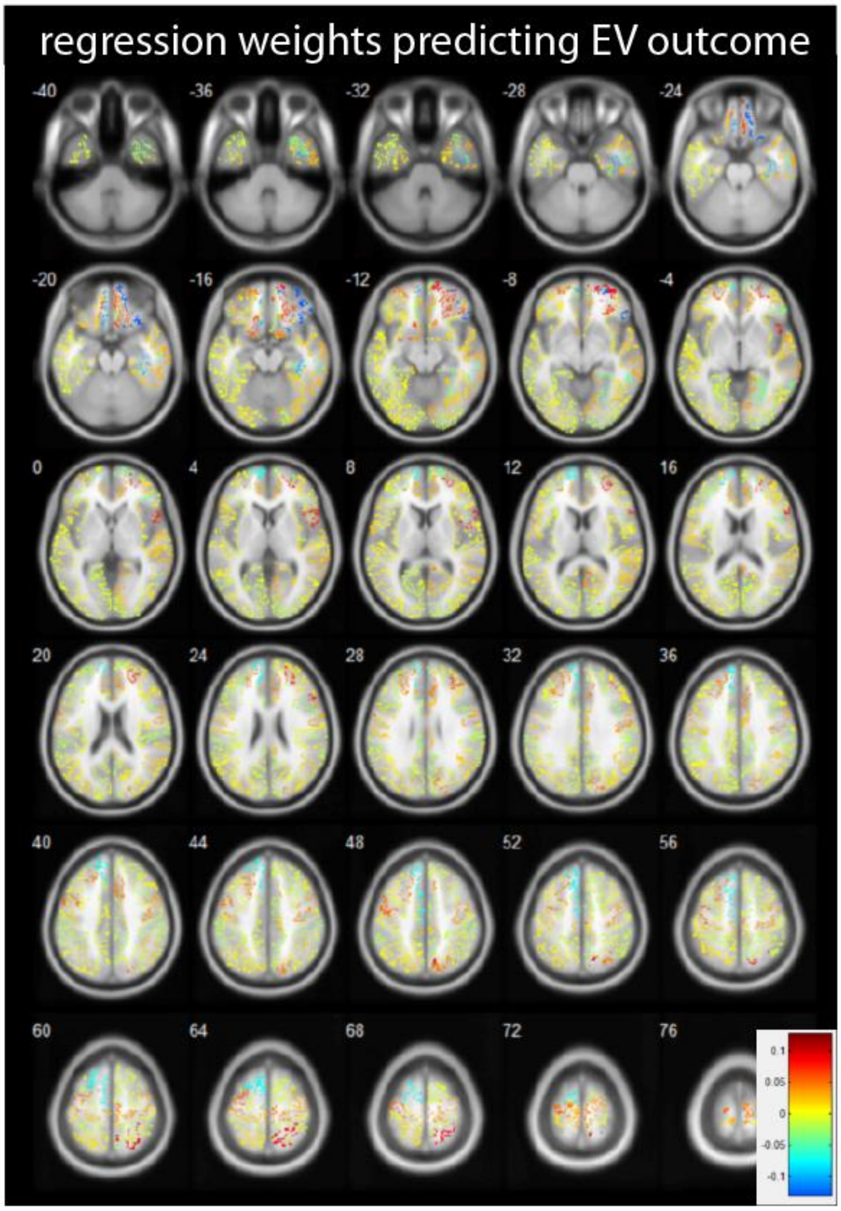


Figure 8.9: structural connectivity regression weights for EV outcome signals in the right caudate. See Figure 8.7 for details.

8.5 Discussion

In this chapter I used a reinforcement learning task to elicit BOLD responses in the striatum related to action values, rewards and expected values. The aim was to explain local variation in these responses based on local variation in the corticostriatal connectivity fingerprint. Focusing on the caudate nucleus, the results show that the connectivity fingerprint of a voxel can be used to predict its response to action values, reward and expected value better than chance. This result supports the widely held belief that partially distinct functional zones in the striatum are determined by inputs from cortex (Alexander et al., 1986; Haber, 2003; Draganski et al., 2008; Averbeck et al., 2014; Haber and Behrens, 2014).

Using an anatomical ROI approach I observed reinforcement learning-related signals in the striatum, replicating previous work (O'Doherty et al., 2004; Tricomi et al., 2004; Rutledge et al., 2009; Jessup and O'Doherty, 2011; Guitart-Masip et al., 2012). However, regions downstream of the striatum such as the internal and external globus pallidus, subthalamic nucleus and SN/VTA showed no such BOLD modulation by task variables. This can be considered surprising for various reasons: firstly, a basic view of brain function would assume that a change in neural activity in the striatum is propagated through the basal ganglia network to effect some change in cortical excitability (e.g. Mink, 1996). The lack of propagation as expressed through average BOLD signal suggests a more subtle mechanism of excitation/inhibition (Cui et al., 2013) or effects through oscillatory mechanism (Brown, 2003), both of which are virtually impossible to measure through fMRI. The second surprise is a lack of reward prediction error signals in the dopaminergic SN/VTA complex (D'Ardenne et al., 2008; Klein-Flugge et al., 2011). It is unclear why the current study does not show such signals though it might relate to the relative unreliability of BOLD in the midbrain (though see Duzel et al., 2009). In any case, the results from the current study suggest a strong dissociation between striatum and its downstream regions in their representation of reinforcement-related values.

The main goal was to understand how these striatal functional signals arise from cortical inputs. The notion that anatomical connectivity determines function is pervasive in neuroscience, and in the striatum it is known that cell populations with projections along the direct and indirect pathway have distinct functional roles in movement (Kravitz et al., 2010; Cui et al., 2013) and reinforcement learning (Kravitz et al., 2012). In humans connectivity fingerprints have been used to segment individual brain structures with remarkable similarity to functional zones (Behrens et al., 2003a; Johansen-Berg et al., 2004; Lyness et al., 2014). This same technique has revealed anatomical parcellation of the striatum (Draganski et al., 2008; Georgiou-Karistianis et al., 2011; Verstynen et al., 2012; Tziortzi et al., 2014), but this has not been directly linked to functional activations.

Saygin et al. (2012) and Osher et al. (2015) introduced a cross-validation technique to assess the predictive power that connectivity has over functional signals. In this chapter I applied this method to reward and value signals in the caudate nucleus. I show that spatial variance in reinforcement learning signals *within* this region can in part be explained by differences in corticostriatal connectivity.

The predictions of functional activity arise from a linear regression model with weights attributed to each corticostriatal connection (Figure 8.6). The distribution of these weights across the cortical surface provides a glimpse of what parts of cortex might be involved in driving activity in the striatum—or in this study, the right caudate nucleus. At a macro-anatomical level, the topography of cortex is maintained in striatal topography (Alexander and Crutcher, 1990; Haber, 2003), though at a finer scale there is also evidence for ‘hot spots’ of convergence where widely separated cortical regions converge on a single striatal patch of tissue (Averbeck et al., 2014). At a microscopic scale it has been suggested direct and indirect pathway medium spiny neurons (MSNs) differentiate in their cortical inputs (Wall et al., 2013). They show that motor cortex projects more strongly to the indirect pathway, whereas somatosensory cortex projects more strongly to the direct pathway. Curiously, in the prediction of reward activity (Figure 8.8) a similar gradient can be observed along motor-to-somatosensory cortex. The results further revealed various value-related prefrontal regions that might contribute to functional activity in the caudate nucleus (Rangel et al., 2008; Rangel and Hare, 2010; Haber and Behrens, 2014). A careful study of the relationship between frontostriatal functional and anatomical connectivity could help understand what information is transferred between cortex and striatum along specific anatomical connections.

There are a number of limitations of this study. Firstly, diffusion connectivity is not ideally placed to pick up on crossing connections, of which there may be many in the striatal system. This weakness was also discussed in chapter 2, but it is noted again here to acknowledge the difficulties in dissociating connectivity fingerprints of neighbouring voxels. That is, the nature of probabilistic tractography will contribute a certain spatial smoothness to connectivity fingerprints. Therefore, these results are agnostic regarding the spatial frequency of the signal that is contributing to the predictive power of the connectivity model—it might be spread across multiple centimetres rather than among directly neighbouring voxels. The second limitation relates to chapters 4-7. By using a simple reinforcement learning task it is impossible to tell whether behaviour is driven by model-based or model-free influences, or both. Despite the use of a model-free algorithm in this chapter I make no claim regarding the origin of these values in terms of model-based or model-free systems. After further development of the imaging approach in this chapter it would be useful to dissociate cortical contributions to model-free and model-based components of striatal function, respectively.

9 Preparing for selective inhibition within frontostriatal loops

9.1 Abstract

In previous chapters I discussed mechanisms of adaptive action selection. A complementary component to selecting the right action is inhibiting the actions that are not appropriate to a situation. Here I examine the behavioural and neural basis of selective inhibition focusing on the role of preparation. In 18 healthy human participants I manipulated the extent to which they could prepare for selective inhibition of specific actions by providing or withholding information on what actions might need to be stopped. I show that, on average, information improves both speed and selectivity of inhibition. BOLD data shows that preparation for selective inhibition engages the inferior frontal gyrus, supplementary motor area and striatum. Examining inter-individual differences, I find the benefit of proactive control to speed and selectivity of inhibition trade off against each other, such that an improvement in stopping speed leads to a deterioration of selectivity of inhibition, and vice versa. This trade-off is implemented through engagement of the dorsolateral prefrontal cortex and putamen. The results suggest proactive selective inhibition is implemented within frontostriatal structures, and I now provide evidence that a speed-selectivity trade-off might underlie a range of findings reported previously.

9.2 Introduction

The prefrontal cortex is thought to represent goals that are subsequently imposed on the motor system (Koechlin et al., 2003). Such executive control also involves the inhibition of actions that are misaligned with current goals, for example when overriding habits or impulsive responses (Isoda and Hikosaka, 2011). Failures of executive control, and in particular its expression during inhibition, are thought to be common in disorders such as addiction (Ersche et al., 2012) and attention deficit hyperactivity disorder (Casey et al., 1997).

Response inhibition is often studied using the stop-signal task (SST), which requires the inhibition of an action following an unpredictable stop signal (Logan et al., 1984). This type of inhibition has been referred to as 'global' because all actions are inhibited, and 'reactive' because no information is used to prepare for inhibition (Aron and Verbruggen, 2008).

The antipodes of reactive and global inhibition are proactive and selective inhibition, respectively (Aron, 2011). 'Proactive' refers to the use of information from the environment that helps prepare an upcoming stop response. 'Selective' refers to the inhibition of only a subset of all ongoing actions. When selective inhibition is executed without preparation, i.e. reactively, it causes interference with all ongoing actions (Coxon et al., 2007; Aron and Verbruggen, 2008; Coxon et al., 2009). This suggests inhibition is implemented by a global 'brake' followed by re-initiation of the remaining action. One framework suggests such a global stop involves the subthalamic nucleus (STN) in the hyperdirect pathway, whereas selective inhibition engages a more action-specific indirect pathway of the basal ganglia (Aron, 2011). However, others have found the IFG, SMA/pre-SMA and entire basal ganglia are involved in preparing for global inhibition (Chikazoe et al., 2009; Jahfari et al., 2012; Zandbelt et al., 2012).

Behavioural and transcranial magnetic stimulation studies have shown that preparation reduces interference between the inhibitory process and the remaining actions, potentially mediated by selective suppression of action representations in primary motor cortex (Aron and Verbruggen, 2008; Mars et al., 2009; Claffey et al., 2010; Neubert et al., 2010; Cai et al., 2011a; Majid et al., 2012; Majid et al., 2013). However, there is to my knowledge no characterization of the full neural network underlying prepared versus unprepared selective inhibition. Such data might extend models of inhibition beyond a current emphasis on reactive global inhibition (Aron, 2011; Schall and Godlove, 2012).

I investigated proactive selective inhibition in healthy human adults by manipulating the information provided about the potential stop target. I hypothesised that preparation would reduce interference caused by an inhibitory process upon the remaining response, and such an improvement in selectivity might lead to a deterioration in the speed of inhibition, essentially posing a speed-selectivity trade-off in inhibition (Aron and Verbruggen, 2008). An existing framework predicts such an improvement in selectivity, rather than speed, reflects greater engagement of an indirect relative to hyperdirect basal ganglia pathway, and involvement of dorsolateral prefrontal cortex (dlPFC) rather than rIFG (Aron, 2011). Thus, this model predicts that proactive selective inhibition will engage striatum and dlPFC, but not STN and rIFG (Aron, 2011).

9.3 Methods

9.3.1 Participants

Nineteen healthy adults participated in the experiment. I excluded one participant because a brain mask could not be created for all functional scans due to movement in the scanner, leaving 18 participants for further analysis (11 females; age range 19-25 years; mean = 21.2, SD = 2.1 years). Fifteen participants were classified as right-handed and three as ambidextrous (Oldfield, 1971). All participants had normal or corrected-to-normal vision, no history of psychiatric or neurological disorder, and provided written informed consent for the experiment, which was approved by the Research Ethics Committee at University College London (UK).

9.3.2 Experimental design

I modified a task previously used to study prepared and unprepared selective action inhibition (Figure 9.1; Aron and Verbruggen, 2008). In brief, on each trial participants responded to a Go signal with either both middle or both index fingers, depending on whether the top or bottom circles displayed on a screen were filled, respectively. On some trials a red cross was presented over one of the two filled circles of the Go signal after a stop-signal delay (SSD). The stop signal indicated that the response with the corresponding finger should be withheld, whereas the other finger should still press down as fast as possible.

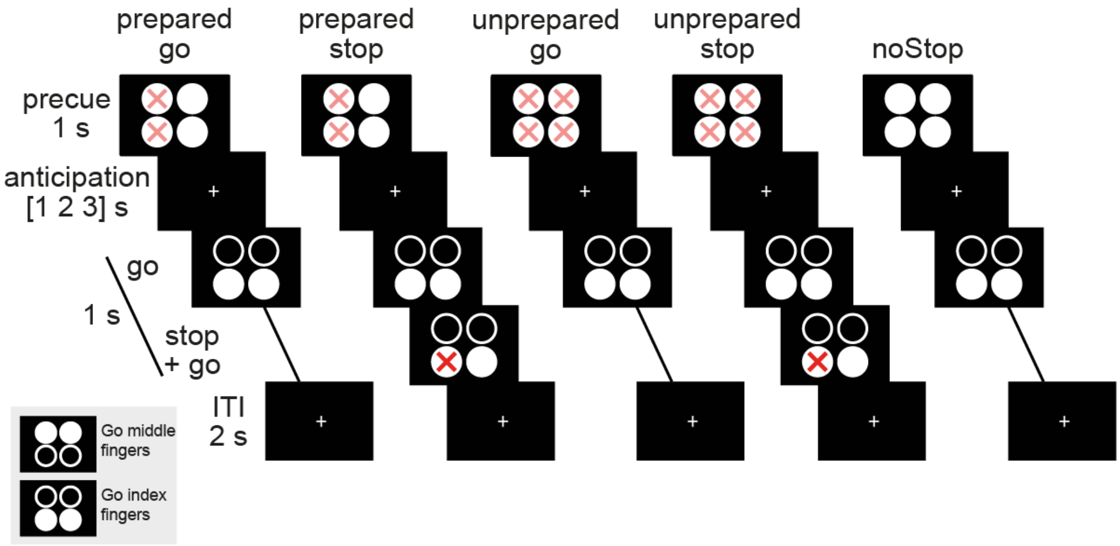


Figure 9.1: Proactive selective inhibition task. The task was designed to study the influence of prior information about inhibition targets on behaviour and neural responses. Responses were made using both index or middle fingers, the four circles on the screen corresponding to fingers on a keypad as indicated by the lower left inset. At the start of each trial faded red crosses cued the participant about the potential locations of a stop signal. In the Prepared condition this would be either both the left or right circles (left or right hand, respectively); in the Unprepared condition the stop signal could appear over any

of the four circles; in the noStop condition there would never be a stop signal. After a jittered anticipation period a Go cue was presented, consisting of 4 circles with either the top or bottom two filled. The fingers corresponding to the filled circles had to press down as fast as possible. In the Prepared and Unprepared condition a stop signal was presented on 30% of trials after a staircased stop signal delay. The location of the stop signal always followed the restrictions set by the cue (i.e. the cued side in Prepared condition, or either side in Unprepared condition). Participants had to stop the finger corresponding to the stop signal, but still go with the finger corresponding to the filled circle without the stop signal. No feedback was provided.

To specifically study preparation for selective inhibition each Go signal was preceded by one of four cues showing the potential locations of the stop signal for that trial. In the Unprepared condition the cue indicated the stop signal could appear anywhere by showing faded red crosses over all four circles, precluding participants from setting up a selective inhibitory representation. In the Prepared condition the cue indicated that the stop signal could only appear on the left or only on the right by showing faded red crosses only over the left or right circles, respectively. Previous work has shown that participants use such cues to set up an inhibitory process specific to the actions that need to be inhibited (Claffey et al., 2010; Majid et al., 2012). In both Unprepared and Prepared conditions the overall probability of a stop signal occurring was 30%. To balance the factor of Information (with levels Unprepared and Prepared), 40% of trials were Unprepared, 20% of trials Prepared-left and 20% of trials Prepared-right. The remaining 20% of trials were noStop trials: the cue consisted of 4 filled white circles with no faded red crosses, indicating that no stop signal would be presented on that trial. This control condition can reveal strategic slowing (Jahfari et al., 2012), but was not used in the imaging analysis as their frequency was not matched with the Information conditions. The design was fully counterbalanced over index and middle fingers.

Trial timings took the following form: the cue was presented for 1 s, followed by 1, 2 or 3 s of anticipation (intervals with probability 0.4, 0.2, and 0.4, respectively). The Go signal appeared on the screen for 1 s, and in 30% of trials was overlaid by a stop signal after a SSD. The Go stimulus remained on the screen for 1 s regardless of button presses, after which a 2 s ITI started. No feedback was provided.

Participants completed 100 trials per block (10 minutes) in the scanner, for 4 blocks during one session. Participants were then taken from the scanner and given 45 to 90 minutes of rest. During this time they were asked to wait in a waiting room and I provided no feedback on their performance. They then underwent another 4 blocks for a total of 800 trials per participant over 80 minutes of functional imaging. This yields 96 stop trials in both the Unprepared and Prepared condition, in line with the number of stop trials in previous studies using the stop signal task (e.g. Aron and Poldrack, 2006; Li et al., 2006; Chikazoe et al., 2009). Trial order was randomised for every block.

I used four independent SSD staircases, one for each of the cue-stop signal combinations (Unprepared left stop, Unprepared right stop, Prepared left stop, Prepared right stop). The SSD became longer after a successful stop, and shorter after a failed stop, in 50 ms steps. This tracking procedure yields a p(stopSuccess) of approximately 0.5, which is optimal for estimation of the stop-signal reaction time (SSRT, see below; Verbruggen and Logan, 2009b; Congdon et al., 2012). The staircases started at values determined during a training session 1 to 7 days before the scanning session.

During that training session, the participant first learned how to respond to Go cues (10 trials) and stop signals (20 trials), and then performed 2 blocks of 100 trials on the full task. Trial-by-trial feedback up until halfway through the first full training block aided instruction. Feedback consisted of success and error messages and a warning when the left and right buttons on Go trials were pressed more than 70 ms apart (asynchronous response). The 4 SSD staircases started off at 100 ms for the last full block, and the participant's last SSD in each staircase became the starting SSD for the scanning session. I instructed participants to use the cue to prepare for the Go signal, and explained it would be impossible to stop every time the stop signal appeared. I also emphasised that responding fast would be more important than correctly stopping on every stop trial. These instructions aim to prevent a 'waiting' strategy which invalidates assumptions of the horse race model used to calculate the SSRT (Logan, 1994).

This experiment was realised using Cogent 2000 developed by the Cogent 2000 team at the Wellcome Trust Centre for Neuroimaging and the Institute of Cognitive Neuroscience, and John Romaya developed Cogent Graphics at the Laboratory of Neurobiology at the Wellcome Department of Imaging Neuroscience.

9.3.3 Behavioural data analysis

Our analyses followed recommendations from the literature (Logan, 1994; Band et al., 2003; Verbruggen and Logan, 2009b; Congdon et al., 2012). I excluded participants if: p(stop) for any of the 4 SSD staircases was lower than 0.25, or higher than 0.75; proportion of correct Go trials (cued fingers pressed down within 70 ms of each other) following any of the 4 cues was below 0.7. All 18 participants passed these criteria. I further computed, for each condition, measures of the Go distribution and number of errors. To validate assumptions of the independent race model I computed stopFail RT as a function of SSD, and z-scored relative finishing time (ZRFT) calculated as (Go_{mean} – SSD – SSRT)/Go_{SD}, evaluated at SSDs of 150-500 ms in 50 ms steps (Logan et al., 1984; Verbruggen and Logan, 2009b).

Two key behavioural measures characterizing stopping in this task are 1) the stop signal reaction time (SSRT), which represents the speed of inhibition, and 2) interference, which represents the inverse selectivity of inhibition. I computed SSRT using the quantile method (Band et al., 2003; Congdon et al., 2012). For each condition (Prepared left, Prepared right, Unprepared) all Go RTs were arranged in descending order. The RT corresponding to the participant’s probability of successfully stopping in that condition was selected (e.g. for a p(stop) of 0.45 I selected the RT 45% down the ordered list), and I subtracted the mean SSD to yield the SSRT. SSRT in the Prepared condition was averaged across left and right cues as an estimate of the time it takes for the participant to inhibit an upcoming motor response after presentation of the stop signal.

I calculated interference of inhibition, or inverse selectivity, as RT on stopSuccess trials minus RT on Go trials for each condition separately (Aron and Verbruggen, 2008). Recall that in all stop trials, participants had to stop one finger and still press down with the other as fast as possible. A positive value in this measure of interference indicates that responses were slower when the participant had to stop a finger compared to Go trials. Interference in the Prepared condition was averaged across left and right cues. I then compared these values across conditions to observe changes in interference, i.e. selectivity, with experimental condition.

I observed that the benefit of Information on SSRT and interference trade off, such that participants seem to focus on improving either speed or selectivity of inhibition, but not both. I therefore computed a measure of this trade-off as ($SSRT_{Unprepared}$ - $SSRT_{Prepared}$) - ($Interference_{Unprepared}$ - $Interference_{Prepared}$). This means that a high trade-off represents a focus on improvement of SSRT with Information, whereas a low value represents a focus on improvement of interference (i.e. selectivity) with Information. Furthermore, I observed that the trade-off is not static over time (see Results: correlation between trade-off in first half versus second half of experiment, r = -10, p = .68), suggesting use of information to prepare selective inhibition is not necessarily homogenous across the entire duration of an experiment.

In order to characterise the brain correlates of this fluctuating use of information for proactive selective inhibition I calculated the magnitude of the trade-off for each trial *t* and used it as a parametric modulator in the fMRI analysis. This trial-by-trial trade-off estimate was calculated using a running average from RT data from trial *t*-75 to *t*+75 (i.e. sliding window of 150 trial width). For trial 1 to 75, the window spanned [1 t+75], and for trial 725 to 800 the window spanned [t-75 800]. Although a smaller width would provide a more fine-grained estimate of the trade-off, this has to be balanced against the number of data points used to calculate the interference and SSRT. With a window size of 150 trials, each Information condition contains 18 stop trials in each window on average, which is sufficient for reliable estimation of the SSRT (Congdon et al., 2012). Using this dynamic measure we could then interrogate neural signatures of this trade-off as expressed during Go and Stop trials in the task.

Behavioural analyses were performed in Matlab (The Mathworks Inc) and SPSS 19 (IBM). I used two-tailed permutation tests with 10^4 draws for paired tests (or 10^7 draws for p-values < .001), analysis of variance (ANOVA) to test for interactions, and 1-sample t-tests to compare outcomes to zero.

9.3.4 MRI data acquisition and preprocessing

I performed magnetic resonance imaging (MRI) on a 3-Tesla Siemens Trio magnetic resonance scanner (Siemens, Erlangen, Germany). Functional data were acquired over 8 runs, each run consisting of 208 whole-brain 3D EPI volumes with spatial resolution = 2.3 x 2.3 x 2.3 mm^3, 80 slices, echo time (TE) = 32.84 ms, volume repetition time (TR) = 2.96 s (Lutti et al., 2013). Parallel imaging (GRAPPA image reconstruction; Griswold et al., 2002), acceleration factor 2 along the partition-encoding direction) was used to speed-up the acquisition of each image volume. Acquisition of dummy volumes to allow for longitudinal magnetization to reach steady-state and of the GRAPPA reconstruction kernel was implemented prior to the acquisition of image data. I acquired B0 field maps for each session using a double echo FLASH sequence (matrix size = 64 x 64; 64 slices; spatial resolution = 3 x 3 x 3 mm^3; gap = 1 mm; short TE = 10 ms; long TE = 12.46 ms; TR = 1020 ms) to correct EPI images for distortion in the B0 field (Weiskopf et al., 2006). Field maps were estimated from the phase difference between the short and long TE using the FieldMap toolbox for SPM (Hutton et al., 2002). Cardiac rate was recorded using an MRI-compatible pulse oximeter (Model 8600 F0, Nonin Medical), and respiration was monitored using a pneumatic belt positioned around the abdomen. I processed these data as described in the literature (Hutton et al., 2011) and included them as regressors of no interest in all first level GLM models (see below). I acquired proton density (PD)-weighted, T1-weighted, and magnetization transfer (MT)-weighted images at 1x1x1 mm^3 resolution for each participant using multi-echo 3D FLASH (Helms et al., 2008b). Fully quantitative maps of the MR parameters MT, R1, PD and R2* were extracted from the acquired data as described in the Methods (chapter 2, also see Helms et al., 2008b). A B1-map was acquired using a 3D SE/STE EPI method (Lutti et al., 2012) to correct for the effects of inhomogeneous radio-frequency excitation on the quantitative maps.

I analysed the MRI data in SPM8 as described in chapter 2. Functional data were smoothed using either a 4 or 10 mm full width at half maximum Gaussian kernel. I used two smoothing levels to optimise sensitivity to widespread activations as well as focused sub-cortical activations in e.g. pallidum and STN.

9.3.5 FMRI data analysis

The preprocessed images were analysed in an event-related design using a general linear model (GLM) with 15 explanatory variables (EVs) of interest. I modelled 12 EVs as stick regressors at time of Information cue onset. Of these, four EVs described correct Go trials (Unprepared, Prepared left, Prepared right, noStop), and 8 EVs described stop trials, crossing information (Unprepared, Prepared), stop-signal side (left, right), and outcome (stopSuccess or stopFail). A further 3 regressors were added at time of the imperative cue (i.e. the go signal): one for all Go trials, one for all stopSuccess trials, and one for all stopFail trials. As all these imperative cues are identical between Information conditions, I did not separately model the information regressors at the imperative cue. As such, the 12 regressors modelled at the time of the precue capture BOLD during both the anticipation epoch and the action execution epoch. I opted for a fast-event related design with a large number of trials, foregoing the opportunity to dissociate activity from these two epochs in each trial.

As described I obtained a measure of the speed-selectivity trade-off, i.e. the extent to which the participant uses Information to improve the speed or selectivity of inhibition, for each trial. I hypothesised that a focus on speeded inhibition would result in a differential engagement of a stopping network compared to a focus on selective inhibition (Aron, 2011). I modelled the trade-off as a parametric modulator on the stick events of each of the 12 regressors at the time of precue, allowing me to examine how each of these events was modulated by the speed-selectivity trade-off. I decided to test for effects of trade-off on Go trials as well as Stop trials in view of the fact that the trade-off is a dynamic state measure, such that participants focus relatively more on speed or selectivity across trials. Although such focus only manifests on stop trials, the participant cannot dissociate stop from go trials until the stop-signal is presented. Thus, it can be reasonably expected that changes in trade-off are also reflected in proactive control during Go trials. I added the following nuisance regressors: 2 regressors for error trials with or without a response, respectively, 6 movement regressors produced by the realignment procedure, 14 physiological regressors for cardiac and respiratory variables (Hutton et al., 2011), and 7 block regressors covering run 1 to 7, respectively. The 8th block was subsumed in the constant of the design matrix. The GLM thus contained a total of 57 regressors over 1664 volumes per participant, and each GLM was estimated separately for each participant for the 4 and 10 mm smoothed images. All EVs (but not physiological regressors) were convolved with a canonical haemodynamic response function (Friston et al., 1995).

My primary interest was a comparison of Go trials between Unprepared and Prepared conditions. Such a contrast elucidates the implementation of proactive control without contamination by the execution of stops as both conditions are equal in terms of motor execution. Furthermore these trials are matched for (violations of) expectations related to stop signal probability (Zandbelt et al., 2012). To obtain group statistics each participant's contrast image was entered into a second level random-effects analysis using one-sample t-tests across participants. I used 10 mm smoothed images for whole-brain analyses, and corrected for multiple comparisons with cluster-level correction at p < .05 (initial threshold at p < .001 uncorrected). I further used a region-of-interest (ROI) approach to examine four areas for which I had strong *a priori* hypotheses regarding their involvement in proactive selective control (Aron, 2011; Jahfari et al., 2012; Zandbelt et al., 2012): the right STN (unthresholded probabilistic ROI as created by Forstmann et al., 2012) and the right caudate, right putamen and right pallidum from the Automated Anatomical Labeling (AAL) atlas. I chose right-lateralised ROIs based on previous work (Jahfari et al., 2012).

For analysis of the parametric modulators I used a similar approach. First I used a whole-brain analysis at cluster-level corrected p < .05 (initial threshold at p < .001 uncorrected). Second, I extracted parameter estimates from functional ROIs resulting from the Prepared > Unprepared Go contrast at 4 mm smoothing, thresholded at p < .01 uncorrected, and masked by anatomical ROIs. In addition to the anatomical ROIs described above, I also examined the right IFG (as defined by inferior operculum in the AAL atlas) and left SMA/pre-

SMA from the AAL atlas (Tzourio-Mazoyer et al., 2002). These two regions were included based on their activation in Prepared versus Unprepared Go trials (Figure 9.5). In all cases where parameter estimates were extracted from ROIs I used MarsBar (Brett et al., 2002) on the 4 mm smoothed images to minimise inclusion of signal not originating from the ROI itself.

9.4 Results

9.4.1 Accuracy and SSD staircase procedure

Go trials were matched between the Unprepared and Prepared condition for overall accuracy (mean (SD) proportion correct Go trials: Unprepared = .87 (.01), Prepared = .86 (.01), noStop = .87 (.01); Unprepared versus Prepared, p = .39). The SSD staircase procedure ensured p(stop) remained close to 0.50 for both Information conditions. I observed a small, but significant, increase in p(stop) for the Prepared compared to Unprepared condition (mean (SD) p(stop): Unprepared = .52 (.01), Prepared = .53 (.01), p = .01). This suggests participants gradually slowed down responses in Go trials over the course of the experiment (Logan, 1994), marginally more so during Prepared compared to Unprepared trials. Note that this difference does not impact on the applicability of the race model (see below).

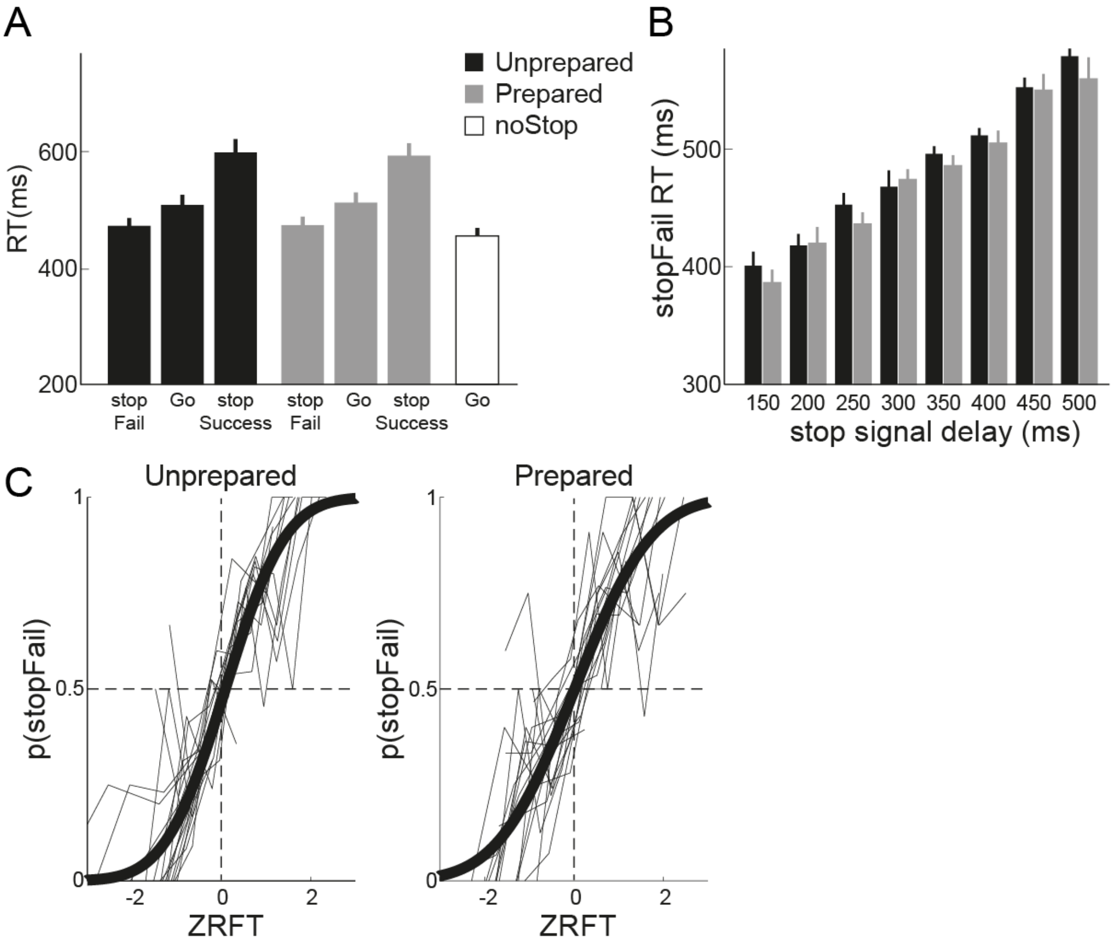


Figure 9.2: Reaction time data satisfies race model assumptions. (A) Reaction times were faster in stopFail trials compared to Go trials for both the Unprepared and Prepared condition. I further observed that noStop trials were faster than Go trials in both Information conditions, indicative of strategic slowing. (B) StopFail RT increased linearly with SSD, as predicted by the independent race model. (C) The z-scored relative finishing time (ZRFT) indicates the finishing time of a Stop and Go process, with higher values indicating a late finishing time for the stop process relative to the go process. Each participant is represented by a thin grey line. A cumulative Gaussian was fit for each participant, and the bold black lines were generated by averaging parameter fits over participants. Both Unprepared (left) and Prepared (right) conditions show that as ZRFT increases, the probability of stopFail increases. When the ZRFT is 0 the probability of stopSuccess and stopFail was close to 0.5, as predicted by the independent race model. Error bars represent SEM.

Previous work has emphasised that participants show a general slowing (rather than slowing over time) when faced with a potential stop versus noStop (Chikazoe et al., 2009; Verbruggen and Logan, 2009a; Jahfari et al., 2010; Jahfari et al., 2012; Zandbelt et al., 2012). In keeping with this I observed a significantly higher RT compared to noStop for both the Unprepared ($p = 8.9 \times 10^{-6}$) and Prepared ($p = 5.4 \times 10^{-6}$) Go conditions consistent with such strategic slowing, but at the same time I found no evidence for a difference between Information conditions ($\text{Go}_{\text{Unprepared}}$ versus $\text{Go}_{\text{Prepared}}$, p = .24). Participants also committed more asynchronous (two fingers > 70ms apart) Go responses in the Prepared compared to Unprepared condition (mean (SD) proportion asynchronous Go: Unprepared = .03 (.005), Prepared = .05 (.006), noStop = .02 (.005); Unprepared versus Prepared, p = .02). An increase in asynchronous errors suggests a higher degree of lateral asymmetry in action preparation. To avoid contamination in other analyses, asynchronous responses were treated as errors and discarded from further analysis. For the remaining trials I computed the frequency of left responses leading right responses, and vice versa, for Prepared-left, Prepared-right, and Unprepared trials. I observed no difference in frequency of these events between Prepared-left and Unprepared (chi-squared test with Yates' correction, χ^2 (1) = 0.06, p = .81) or Prepared-right and Unprepared (χ^2 (1) = .03, p = .86). This suggests excluding asynchronous trials successfully removed the asymmetry in action execution.

9.4.2 Selective inhibition satisfies independent race model assumptions

A dominant model in the inhibition literature is the independent race model (Logan et al., 1984). It is unclear, however, whether selective inhibition is accurately described by this class of model. It has been suggested that more complex models such as an interactive race model are required when assumptions of the independent race model are not met (Boucher et al., 2007; Verbruggen and Logan, 2009b; Schall and Godlove, 2012). I show that in this task, selective inhibition conforms to all assumptions of the independent race model, both for the Prepared as well as the Unprepared conditions (Figure 9.2).

Three basic conditions must be met for the independent race model to be valid. First, trials in which a stop signal occurs but the participant fails to stop ('stopFail') should represent the fast half of the RT distribution. Thus, stopFail RTs must be faster than Go RTs within the same condition. I observed that this was the case for both the Unprepared ($p = 8.5 \times 10^{-6}$) and Prepared ($p = 9.3 \times 10^{-6}$) condition (Figure 9.2A). Furthermore this difference between stopFail and Go was not significantly different between Unprepared and Prepared conditions (repeated-measures ANOVA interaction: $F(1, 17) = 1.1$, $p = .31$). The second condition for the independent race model is that stopFail RTs should increase as the SSD increases due to the stop process finishing later. I observed such a linear increase in stopFail RT with SSD for both the Unprepared (Figure 9.2B; linear regression for each participant, mean (SD) over population: intercept = 368 (74) ms, slope = 0.30 (0.25); 1-sample t-test on slope, $t(17) = 5.0$, $p = 5.5 \times 10^{-5}$) and Prepared (intercept = 347 (86) ms, slope = 0.38 (0.32), $t(17) = 5.2$, $p = 3.6 \times 10^{-5}$) condition. There was no difference between the Unprepared and Prepared condition in the intercept ($p = .27$) or slope ($p = .21$). Thus, stopFail RT significantly increased linearly with SSD, with no evidence for any difference between the Unprepared and Prepared condition. The third condition is that the p(stopFail) should predictably change with the z-scored relative finishing times (ZRFT) of the Stop and Go process (see Methods for calculation). A negative

ZRFT indicates that the stop process finished earlier than the Go process and the participant should thus have a low p(stopFail); a positive ZRFT indicates the Go process finished before the Stop process and the participant should be likely to erroneously respond, i.e. have a high p(stopFail). When the ZRFT is zero, the model predicts that both processes finished simultaneously and p(stopFail) should be 0.5.

I plot p(stopFail) as a function of z-scored relative finishing time (ZRFT) and fit a cumulative Gaussian to each participant's data individually. The bold curves in Figure 9.2C represent the average of each participant's fits. The mean of either bold curve was not significantly different from zero (Unprepared: 1-sample t-test, t(17) = 1.2, p = .25; Prepared: t(17) < 1), nor were the means significantly different between conditions (p = .14). The SD of the curve was larger in the Prepared than Unprepared condition (p = .02), indicating a significant decrease in the accuracy with which the ZRFT of the Stop and Go process predicts the outcome of the race in the Prepared compared to Unprepared condition. Nonetheless, the analyses presented here confirm that selective inhibition conforms to all assumptions of the independent race model regardless of proactive control. I can then use the race model to calculate parameters for further behavioural and fMRI analyses.

9.4.3 Modulation of speed and selectivity of inhibition by preparation

The previous analyses show that the Prepared and Unprepared condition are matched across a range of characteristics including accuracy, reaction times and strategic slowing, and that behaviour in this task can be modelled using the independent race model. From this model I derived the SSRT, and found that SSRT improved with prior knowledge of the hand that needs to be stopped, i.e.

when selective inhibition could be prepared (Figure 9.3A; Unprepared versus Prepared SSRT: p = .02). Preparation also improved interference, i.e. the selectivity of inhibition (Figure 9.3B; Unprepared versus prepared interference: p = .02). Thus, at the group level, both speed and selectivity of inhibition improved with preparation.

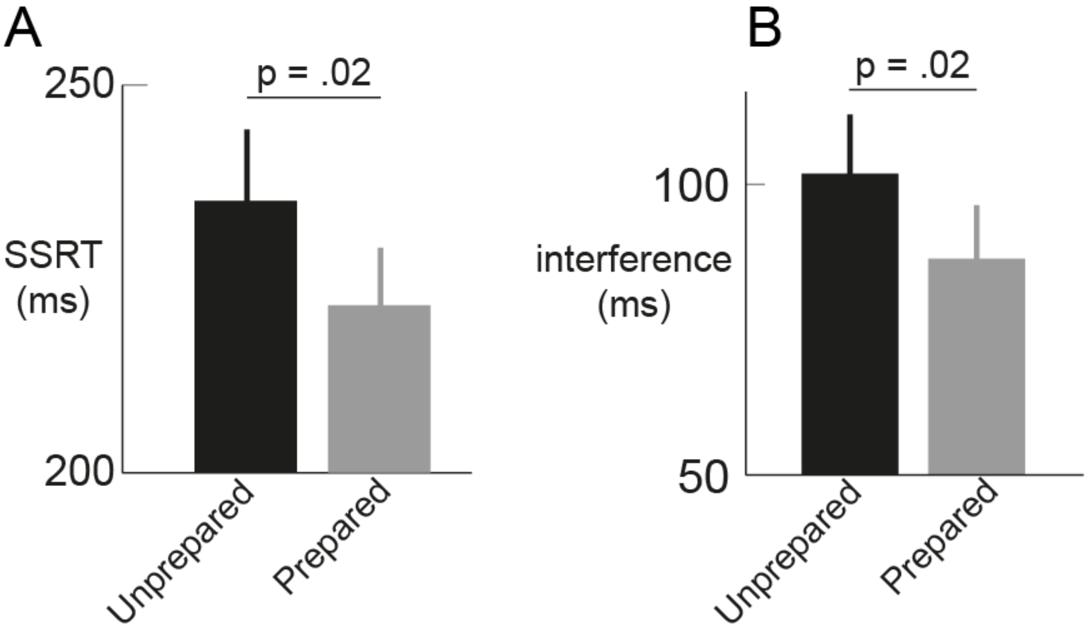


Figure 9.3: Information improves the speed and selectivity of inhibition. (A) Stop signal reaction time (SSRT), representing the speed of inhibition, is faster in the Prepared compared to Unprepared condition. (B) The interference between the inhibition process and the remaining action is reduced in the Prepared compared to Unprepared condition. P-values are from permutation tests, and error bars represent SEM.

This result contrasts with a previous report showing that preparation reduces interference but paradoxically lengthens the SSRT (Aron and Verbruggen, 2008). As in the current study, the task used in the latter required participants respond to a stop signal by stopping one finger while at the same time pressing down with their other finger as fast as possible. Consequently, I asked whether these two task requirements draw on some shared resource and trade off against each other. Specifically, when participants are provided with information they may prepare for a fast stop (i.e. improved SSRT), a fast remaining response (i.e. improved interference), or both. I show that at the group level preparation favours both speed and selectivity (Figure 9.3). However, when looking at inter-individual differences, I found preparation trades off speed against selectivity: the benefit of information to SSRT is negatively correlated with the benefit of information to interference (Figure 9.4A; r = -.70, p = .001). For example, some participants show SSRT improvements in the Prepared compared to Unprepared condition, but show no improvement in interference (Figure 9.4A, lower right). Other participants reduced their interference in the Prepared compared to Unprepared condition, but did not improve their stopping speed (Figure 9.4A, upper left). To quantify this trade-off I calculated a summary measure ($\text{SSRT}_{\text{Unprepared}}$ - $\text{SSRT}_{\text{Prepared}}$) - ($\text{Interference}_{\text{Unprepared}}$ - $\text{Interference}_{\text{Prepared}}$) for each participant. This trade-off is high when preparation is used to improve speed, and low when preparation is used to improve selectivity (i.e. to reduce interference).

A key observation in the study was that this trade-off measure based on the entire dataset (i.e. ~80 minutes of time on task) is not necessarily fully representative of a participant’s trade-off at any time point in the experiment, as shown by the lack of correlation between a participant’s trade-off calculated separately for the first compared to the second half of the experiment (Figure 9.4B; r = -10, p = .68). However, when I examined how participants *changed* their behaviour from the first to the second half of the experiment, I again observed a speed-selectivity trade-off: those participants that improved on

SSRT from the first to second half deteriorated on interference, and vice versa (Figure 9.4C; r = -.75, p = .0003). These results suggest the speed-selectivity trade-off is dynamic, and that an estimate of the trade-off based on a participant's entire dataset might not accurately describe the trade-off at any given point in the experiment. I therefore calculated a trade-off for each trial using a running average over RT data (see Methods for details; Figure 9.4D). This dynamic measure of the speed-selectivity trade-off was then used to interrogate the entire neuronal data on how proactive selective control is instantiated in the brain, providing the key test to identify regions that promote slow and selective versus fast and global inhibition.

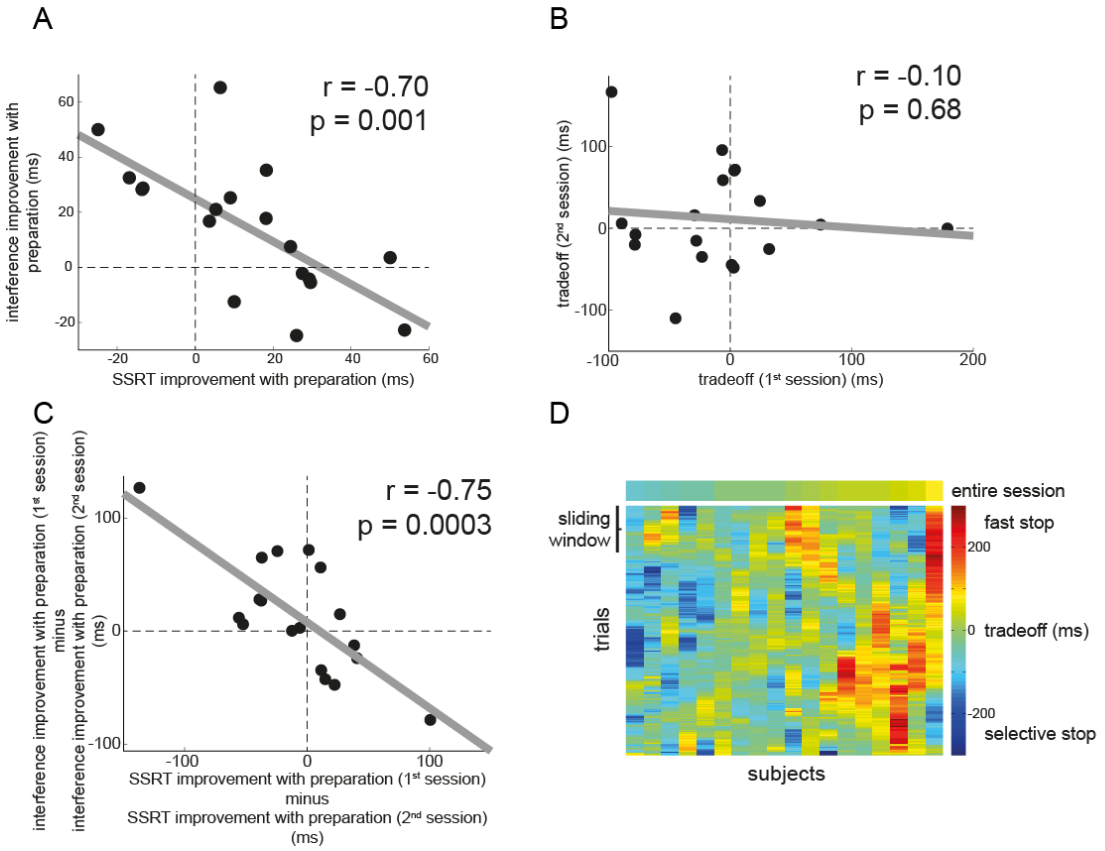


Figure 9.4: Preparation for selective inhibition trades off improvements in speed against selectivity. (A) Each black dot represents a participant. I observed a negative correlation between the effect of Information on SSRT and the effect of

Information on Interference: the more participants used the Prepared cue to stop fast, the less they used the Prepared cue to reduce interference, and vice versa. (B) I calculated a relative measure of speed-selectivity trade-off as ($SSRT_{Unprepared}$ - $SSRT_{Prepared}$) - ($Interference_{Unprepared}$ - $Interference_{Prepared}$). However, there was no correlation between a participant's trade-off in the 1st compared to 2nd session. This suggests that a single trade-off measure does not adequately describe a participant's trade-off during the entire experiment. (C) Despite this instability over time, the way in which a participant's behaviour changed over time was again governed by the speed-selectivity trade-off: participants that, from their 1st to 2nd session, increased use of the Prepared cue to stop fast decreased use of the Prepared cue to stop selectively. (D) To examine BOLD responses that might reflect this trade-off I calculated a trade-off measure for each trial based on RT data from a window around that trial. Each column represents a participant, each row represents a trial. A measure of trade-off over the entire experiment (top) ignores variance that is evident in trial-by-trial estimates of the trade-off (bottom).

9.4.4 BOLD responses in Unprepared versus Prepared Go trials

I first compared BOLD responses between Unprepared and Prepared correct Go trials. Note any effects in this contrast can be attributed to anticipation and preparation for selective inhibition without being confounded by differences in the actual stopping process such as SSRT and interference. Moreover, behaviour was matched between these Go conditions in terms of motor demands, accuracy and RTs (see e.g. Figure 9.2).

I observed a number of regions that responded more strongly to a Prepared compared to Unprepared cue (Figure 9.5), including the right IFG, left SMA/pre-SMA, bilateral dorsal premotor cortex (PMd), and bilateral parietal cortex (Figure 9.5A; $p < .05$ cluster-level corrected). To complement this voxel-based analysis, and given the strong a priori hypothesis for involvement of the right basal ganglia in response inhibition (Zandbelt and Vink, 2010; Aron, 2011;

Jahfari et al., 2012), I also performed a hypothesis-driven anatomical ROI analysis (Figure 9.5B; alpha = .013, Bonferroni-corrected for 4 ROIs). This showed that right putamen had a greater BOLD response to the Prepared compared to the Unprepared cue (p = .005), with only weak evidence for involvement of the right pallidum (p = .02), and no significant effects observed for right STN (p = .08) or right caudate nucleus (p = .17) (Figure 9.5B). Together this suggests that proactive selective inhibition engages a set of regions also involved in global and reactive inhibition (Aron, 2011), with the notable exception of STN and caudate nucleus (cf. Majid et al., 2013) where I observed a null effect. I observed no clusters in the brain that responded more strongly during Unprepared compared to Prepared Go trials.

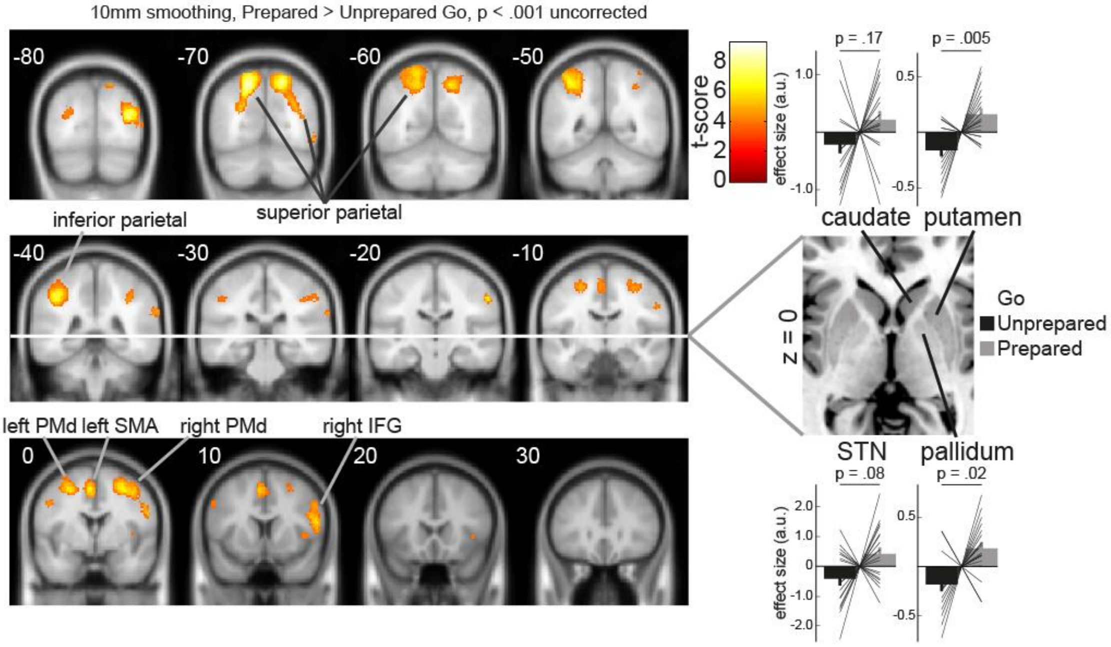


Figure 9.5: Prepared > Unprepared Go trials. (A) I observed stronger BOLD response to Prepared compared to Unprepared Go trials in bilateral dorsal premotor cortex (PMd), left SMA/pre-SMA, right IFG, and bilateral parietal structures. All maps were thresholded at p < .001 uncorrected (shown here for illustration purposes) and statistical significance was assessed at p < .05

cluster-level corrected threshold. Results were projected on coronal slices of the MNI 152T1 template (left = left) using xjView (www.alivelearn.net/xjview). (B) I extracted beta coefficients for the Prepared and Unprepared Go regressors from four anatomical ROIs and performed pair-wise permutation tests. The putamen and pallidum showed significantly stronger responses during Prepared compared to Unprepared Go trials, whereas the STN and caudate showed no such effect. The coefficients are mean-centred for each participant for visualization purposes. The axial slice was taken from the 'ch2' template in MRIcron (Rorden and Brett, 2000). Error bars represent SEM.

9.4.5 Speed-selectivity trade-off in frontostriatal regions

To understand how the brain uses information to promote the speed or selectivity of stopping I used a metric of the dynamic trade-off (Figure 9.4D) as parametric modulator on each of the EVs of interest. Note that the trade-off is a relative measure, such that a low trade-off could be due to slower and more selective inhibition across Prepared trials, or conversely speeded and more non-selective inhibition across Unprepared trials. Thus, any brain region that promotes fast but non-selective inhibition would show a negative coefficient for Unprepared, or positive coefficient for Prepared stop trials. Vice versa, regions that promote selective but slow inhibition would show a positive coefficient for Unprepared, or negative coefficient for Prepared stop trials. Crucially, for either type of region, the coefficients should be different between Information conditions. Thus, a two-tailed contrast of parametric modulators Trade-off$_{\text{Unprepared-stop}}$ versus Trade-off$_{\text{Prepared-stop}}$ identifies the regions that putatively modulate the speed versus selectivity of inhibition. Post-hoc t-tests can then confirm that the coefficients are significantly different from zero. I did not observe any such effect for the parametric modulator on Go trials (data not shown), and I therefore only report findings on Unprepared versus Prepared

Stop trials. I combined stopSuccess and stopFail trials in order to sample the entire distribution of responses; they comprise the fast and slow part of the Go distribution, respectively, such that the subset of trials that falls in either category is a function of SSRT, and thus of trade-off. By selecting all stop trials I prevent a confounding effect of general RT.

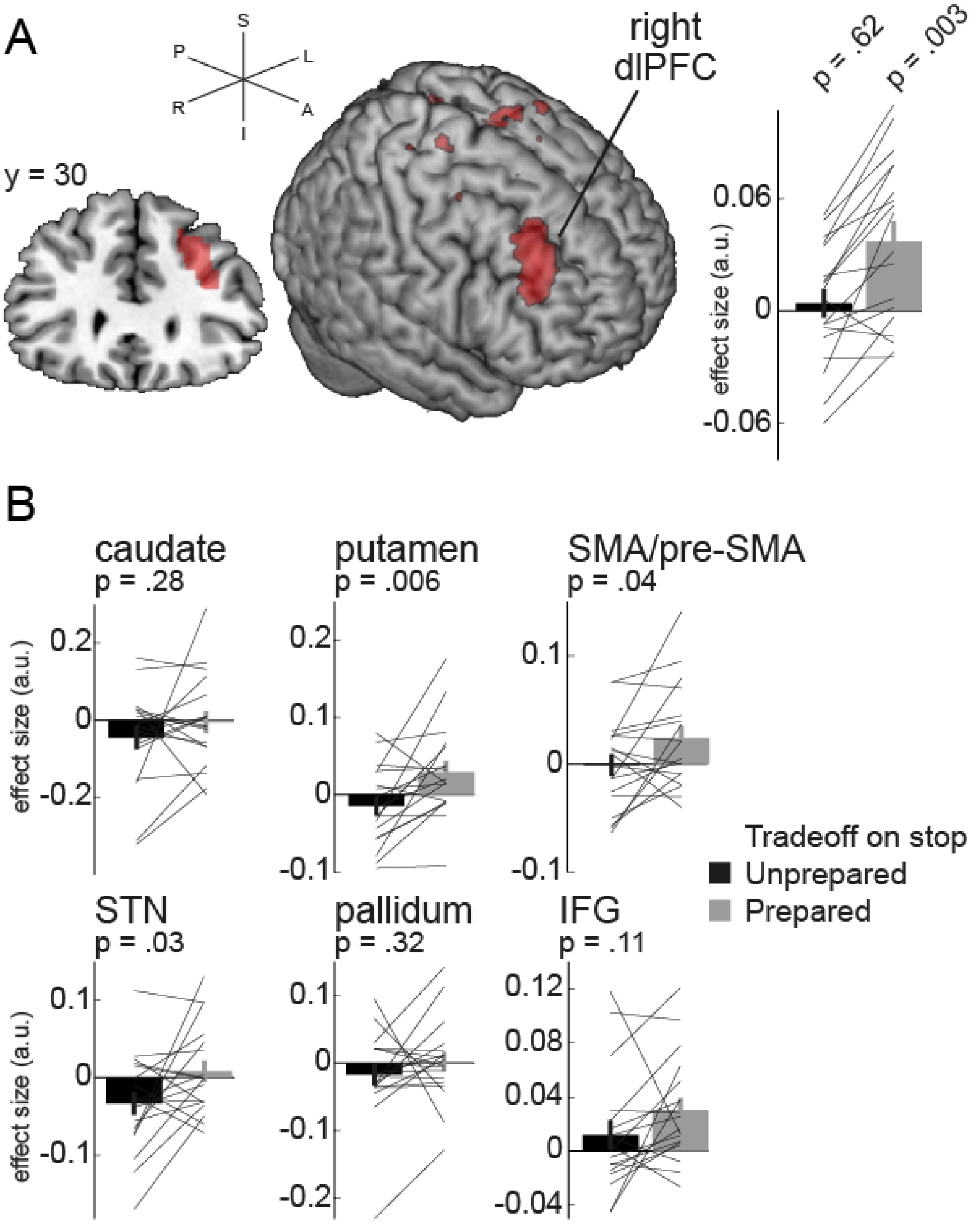


Figure 9.6: Expression of the speed-selectivity trade-off during stop trials. (A) The trade-off reflects a dynamic focus on speeded or selective inhibition in Prepared relative to Unprepared stop trials. I asked whether any regions modulated their activity to reflect this trade-off. Based on how I defined the trade-off (see Methods) one would predict that regions underlying speeded (but non-selective) inhibition have a negative coefficient in the Unprepared, or a positive coefficient in the Prepared condition. In a whole-brain analysis I

observed a cluster in the Trade-off$_{\text{Prepared-stop}}$ > Trade-off$_{\text{Unprepared-stop}}$ contrast corresponding to right dlPFC (left). Extraction of the coefficients showed that this cluster specifically promoted speeded (but non-selective) inhibition during Prepared, but not Unprepared, inhibition (right). Results visualised in MRIcron at p < .001 uncorrected. (B) The putamen, SMA/pre-SMA and STN showed the same effect as the cluster in Figure 9.6A, linking them to speeded inhibition. Of these, the SMA/pre-SMA and putamen mediated fast inhibition only in Prepared trials, whereas the STN mediated fast inhibition only in Unprepared trials*. Error bars indicate SEM.*

I observed only one cluster that survived multiple comparison correction in the Trade-off$_{\text{Prepared-stop}}$ > Trade-off$_{\text{Unprepared-stop}}$ with a peak activation at MNI coordinates [34, 30, 24] (Figure 9.6A; p < .05 cluster-level corrected), a cluster that incorporates the middle frontal gyrus, i.e. dlPFC. I extracted parameter estimates from the significant cluster (defined at p < .001 uncorrected) to test whether the cluster's activity reflects the speed-selectivity trade-off during Unprepared or Prepared inhibition, or both (Figure 9.6A, right). Whereas the trade-off did not modulate activity during Unprepared trials (1-sample t-test, t(17) < 1), the cluster was significantly positive in the Prepared condition (t(17) = 3.63, p = .002) suggesting increased activity in this region leads to speeded but non-selective inhibition during Prepared trials.

To further explore this effect I reasoned that the trade-off could be expressed within voxels sensitive to information in brain regions implicated in the implementation of stopping (i.e. rIFG, SMA/pre-SMA, caudate, putamen, pallidum, STN; Jahfari et al., 2011; Zandbelt et al., 2012). To test this hypothesis I built functional ROIs by thresholding the Prepared > Unprepared Go contrast at p < 0.01 and constraining these to an anatomical mask of each region (see Methods for details on ROI construction). As shown in Figure 9.6 I found evidence for an effect of Information on the trade-off in the putamen (p = .006) when using Bonferroni correction for 6 tests (alpha = .008). A similar pattern, but only significant at uncorrected threshold of p < .05, was found in the right STN (p = .03) and left SMA/pre-SMA (p = .04). All three showed the signature of regions that promote speeded but non-selective inhibition. To understand in what condition each region contributed most strongly I tested the individual coefficients against zero (all 1-sample t-tests with 17 DOF) to reveal the putamen (Unprepared, t = 1.2, p = .25; Prepared, t = 2.1, p = .05) and SMA/pre-SMA (Unprepared, t < 1; Prepared, t = 2.0, p = .06) mediated speeded but non-selective inhibition which was most pronounced when information was available. In contrast, the STN promoted speeded but non-selective inhibition only in the Unprepared condition (Unprepared, t = 2.2, p = .05; Prepared, t < 1). I did not identify any regions that promoted selective but slow inhibition. Together, this provides tentative evidence that the speed-selectivity trade-off was driven by changes towards a focus on speed implemented by different neural structures depending on the availability of prior information: the dlPFC, the putamen, and the SMA/pre-SMA when information was available (proactive inhibition) and the STN when no information was provided (reactive inhibition).

9.5 Discussion

These data show that participants trade off speed and selectivity in stopping when performing proactive selective inhibition, an effect implemented through engagement of dorsolateral prefrontal cortex and striatum. These two regions, contrary to predictions, promote speeded rather than selective inhibition. Provision of information to prepare selective inhibition recruits a set of brain regions implicated in the implementation of inhibition, including the SMA/pre-SMA, IFG and putamen.

A recent model of response inhibition describes action inhibition along two axes: global-selective (i.e. whether all or only a subset of actions are stopped) and reactive-proactive (i.e. the extent of preparation for inhibition; Aron, 2011). My task examined selective inhibition in a proactive versus reactive context by providing or withholding from participants specific information about the target of inhibition, respectively. This extends findings on global inhibition along a reactive-proactive scale (Verbruggen and Logan, 2009a; Jahfari et al., 2012; Zandbelt et al., 2012) and selective inhibition in the reactive domain (Coxon et al., 2007, 2009; Ko and Miller, 2013). In comparisons to these different types of inhibition one unresolved question is whether selective inhibition is sufficiently similar to global inhibition, such that it too can be analysed using the independent race model or might require a more elaborate interactive race model (Boucher et al., 2007; Verbruggen and Logan, 2009b; Schall and Godlove, 2012). Here I confirm that both proactive and reactive selective inhibition satisfy all assumptions of the independent race model. Given the current debate, however, it would be best to consider the validity of the independent race model on a study-by-study basis.

Applying the race model to these data, I observed proactive control improved speed (i.e. SSRT) and selectivity (i.e. interference) of inhibition compared to reactive control. This partly contrasts with results from a recent series of behavioural and transcranial magnetic stimulation studies that show preparation reduces interference, as found here, but either leads to a deterioration (Aron and Verbruggen, 2008) or does not affect (Claffey et al., 2010; Majid et al.,

2012) the SSRT. These seemingly contradictory results can potentially be explained by my finding that each participant is trading off SSRT with interference (i.e. speed and selectivity). To illustrate this point, this participant cohort contains subgroups that improve in selectivity but deteriorate in SSRT (as in Aron and Verbruggen, 2008; Figure 9.4A, left top quadrant); improve on both characteristics (Figure 9.4A, right top quadrant); and improve SSRT but deteriorate in selectivity (Figure 9.4A, right lower quadrant). Such a trade-off suggests that global inhibition is fast whereas selective inhibition is slow. Furthermore I show that the trade-off can change over time. It is an open question what exactly drives these changes in trade-off. Previous work has shown that participants can flexibly adjust their Go versus Stopping speed to optimise rewards (Leotti and Wager, 2010), and one might expect that participants can similarly adjust their trade-off when incentivised to do so. In addition to such top-down control, experimental factors are likely to influence the speed-selectivity trade-off, such as the probability of a stop-signal occurring, the dynamics of the SSD staircasing procedure, or the nature of the instructions and feedback provided to the participants. The results presented here show that this trade-off exists, and new experiments could usefully explore the factors that affect it.

The imaging analysis tested a number of predictions from the action inhibition framework suggested by Aron (2011). Briefly, the model suggests that reactive selective inhibition engages the IFG whereas proactive selective inhibition engages the dlPFC in association with the indirect pathway of the basal ganglia (but excluding the STN). However, I observed that the IFG, but not dlPFC, is more active during proactive selective control. Additionally, this form of anticipation engaged other regions previously implicated in reactive response inhibition itself, including SMA/pre-SMA and striatum. This is reminiscent of findings that stop-signal probability (a manipulator of proactive control) positively correlates with activity in this inhibition network (Jahfari et al., 2012). A notable difference with my study is that I find activity despite keeping the stop-signal probability equal between Unprepared and Prepared conditions, thus preventing a confound where the rIFG responds to the violation of an expectation rather than the preparation for inhibition itself (Zandbelt et al., 2012). As such, the increase in activity is likely to reflect additional processing required for proactive selective inhibition, which might involve for example attentional processes or the maintenance of inhibitory set, processes that can only be uncovered by targeted experimental designs that are not suitable to test my current hypothesis (e.g. Li et al., 2006; Zandbelt and Vink, 2010). A shortcoming of the fast event-related design was that the results remain inconclusive regarding the exact component processes that each of these areas subserves (see Ridderinkhof et al., 2004; Neubert et al., 2013), or how changes in activity in these areas relate to changes in performance. I also could not dissociate neural activity from the anticipation epoch from activity during the response epoch, such that the reported changes in neural activity might span either or both of these time windows. Despite these limitations my analysis shows that neural processing associated with prior information occurs within the known pathways of inhibition, which include IFG, SMA/pre-SMA and striatum, rather than in the dlPFC as suggested previously (Aron, 2011). Further research is required to ascertain the specific timing and cognitive processes implemented within the dlPFC during proactive inhibition.

The involvement of this ‘classic’ stopping network in proactive selective control contrasts with the absence of evidence for engagement of STN during proactive control, in line with a priori predictions (Aron, 2011). The STN has been widely implicated in the execution of global inhibition (e.g. Aron and Poldrack, 2006; Frank et al., 2007a; Eagle et al., 2008) and more recently, in the preparation for global inhibition (Jahfari et al., 2012). My results suggest that, at least with regards to preparation, the STN is not involved in selective inhibition. As noted earlier, selective inhibition might circumvent this pathway, and its associated global inhibitory effect, by inhibiting a specific motor command exclusively through the indirect pathway of the striatum (Baker et al., 2010b; Aron, 2011; Majid et al., 2013). I observed that the putamen is engaged in proactive selective control, whereas I observed a null effect for the STN in the same contrast (but note this null effect does not prove a lack of involvement). Regarding the striatum, I present evidence for involvement of the putamen, but not caudate nucleus, whereas I note other recent work has implicated both structures in proactive selective control (Majid et al., 2013). Given that the putamen, more so than the caudate, is a fundamental component of a basal ganglia motor loop (Alexander et al., 1986), I suggest that the putamen plays a pivotal role in implementing selective response inhibition. A closer inspection of the electrophysiology of the striatum might provide insights into the differential roles of putamen and caudate nucleus (as in Schmidt et al., 2013).

Activation within right dlPFC and right putamen most strongly reflected a speed-selectivity trade-off during stop trials: activations in these regions positively correlated with a focus on speeded rather than selective inhibition when information about which response to inhibit was available. This finding suggests that the dlPFC, together with the striatum, process available information to prioritise and prepare the speed of inhibition for an action. This role of the dlPFC in setting and prioritizing among future action goals resonates with recent findings suggesting that speed-accuracy trade-off (SAT) often observed in perception and action (Schouten and Bekker, 1967) is resolved within fronto-basal ganglia pathways (Forstmann et al., 2008; van Veen et al., 2008; Bogacz et al., 2010). Specifically, activity in the dlPFC and basal ganglia positively correlates with a focus on speeded rather than accurate responses. Note, however, these two types of trade-off are not identical: I find that Information affects the speed and selectivity of inhibition, but not the RT or accuracy of responses. However, similar to the SAT (van Veen et al., 2008) it might be a change in baseline firing rate that governs whether participants emphasise one or the other. These findings implicate dlPFC and striatum in selective inhibition as suggested by the Aron model, but it also suggests a refinement in which the putamen and dlPFC are more, rather than less, active when focusing on speed over selectivity. On the other hand, the data suggest that the STN is engaged when fast inhibition is prioritised over selectivity when no information is provided. In these circumstances selective stopping could not be prepared and had to be executed on line. These findings are compatible with the notion that the STN is engaged in fast but non-selective inhibition (Coxon et al., 2009).

The finding that frontostriatal circuits mediate proactive control raises a number of questions. Firstly, activity in frontostriatal circuits associated with proactive control might reflect either targeted inhibition or enhancement of specific actions. For example, low interference in this task could be caused by inhibition of the action that needs to be stopped, or enhancement of the actions that still need to be executed. These functions might be subserved by indirect and direct pathways respectively, but disentangling these different neuronal populations in human fMRI is a major challenge due to their likely anatomical overlap (Gerfen and Surmeier, 2011). In fact recent evidence indicates that the direct and indirect pathway are simultaneously active during action initiation thus suggesting their concurrent activation is required for an execution of a complex motor plan (Cui et al., 2013). Secondly, fMRI is not well suited to understanding the temporal dynamics of proactive control—a more suitable approach might be neurostimulation (Mars et al., 2009; Neubert et al., 2010) or electrophysiological recording (Isoda and Hikosaka, 2008, 2011).

A recent surge of interest in response inhibition that goes beyond all-out, reactive stopping motivated us to examine the role of preparation in selective inhibition. My data reveal that the opportunity to prepare for inhibition poses a trade-off between either faster or more selective inhibition. This trade-off is expressed in frontostriatal structures commonly associated with the preparation for, and execution of, response inhibition and allows adjustments of behaviour mandated by current context.

10 Proactive and reactive response inhibition across the lifespan

10.1 Abstract

In the previous chapter I described the neural mechanisms of preparatory inhibitory control, revealing a considerable overlap between prefrontal and striatal regions mediating outright response inhibition and preparation for response inhibition, including right inferior frontal gyrus (IFG), premotor areas and striatum. In this chapter I take a different approach and ask how reactive and proactive control change across demographics, including age, gender, education and also measures of depression. Specifically, if the neural structures underlying proactive and reactive control overlap, and these structures deteriorate with age, then this begs the question do reactive and proactive control decline similarly with age? To answer this I used an almost identical response inhibition task as before, but delivered using a smartphone-based platform that allowed me to test a very large community sample (n = 12,496). As in chapter 9 I examine proactive control as the change in stop-signal reaction time (SSRT) when participants are provided with advance information about the upcoming trial compared to when they are not, whereas reactive control is defined as the SSRT when no such advance information is provided. As predicted, reactive control declines with natural aging, and the rate of decline was greater in men than women (~10 ms versus ~8 ms per decade of adult life). Surprisingly, the benefit of preparation, i.e. proactive control, did not change over the lifespan and interestingly women showed greater proactive control at all ages compared to men. Together these results suggest that reactive and proactive inhibitory control at least partially rely on separate neural substrates that are differentially sensitive to age-related change.

10.2 Introduction

Humans frequently need to exert rapid reactive control over their actions, such as stopping their car when an animal unexpectedly jumps on to the road. However, humans can also use informative cues and contexts to implement proactive control (Gollwitzer, 1999; Aron, 2011; Braver, 2012), as when keeping one's foot close to the brake after passing a warning sign for a potential deer on the road. In chapter 2 I argued that proactive control provides a more ecologically interesting framework for understanding both everyday behaviour and impulse control disorders (Aron, 2011; Schall and Godlove, 2012). However, the dominant paradigm in the inhibition literature—the stop-signal task—only measures reactive control (Logan et al., 1984; Verbruggen and Logan, 2008). This task has provided a detailed understanding of how fronto-basal ganglia loops subserve reactive control (Aron and Poldrack, 2006; Schmidt et al., 2013) and how age-related decline in these pathways is associated with impaired reactive control (Coxon et al., 2012).

The neural basis of proactive control seems to show extensive overlap with that of reactive control as I noted in chapter 9 (and as observed by others, e.g. Jahfari et al., 2011; Majid et al., 2013), with perhaps a more prominent role for the striatum and dorsolateral prefrontal cortex in proactive control (chapter 9 and Zandbelt et al., 2012). In this chapter I focus on age-related decline and ask whether the similarity in the neural substrates that underlie reactive and proactive control means that both types of inhibition will decline at a similar rate over the lifespan. Age-related volume reductions are particularly pronounced in frontal regions and occur at a more rapid rate in men than women (Gur et al., 1991; Cowell et al., 1994; Murphy et al., 1996; Coffey et al., 1998). Men are also more likely to develop neurodegenerative disease early in life (Miech et al., 2002; Raber et al., 2004). I therefore hypothesised that age-related decline would be more pronounced in men than women for both reactive and proactive inhibitory control.

To acquire a large and comprehensive sample I collected data through The Great Brain Experiment, a smartphone app with experiments presented under the cover of games (Brown et al., 2014). The app also recorded educational attainment and, for a subset of players, a measure of depressive symptoms. Although depression is not thought to be related to reactive inhibitory control (Lau et al., 2007; Lipszyc and Schachar, 2010; Sjoerds et al., 2014), these studies have relatively small numbers of participants and additionally did not test for a possible relationship between proactive control and depressive symptoms. Lastly, this chapter examines the reliability of response inhibition measures acquired through smartphones and establishes whether assumptions of the race model underlying the calculation of stop-signal reaction time (SSRT) hold for these data, as performed in chapter 9 (Logan et al., 1984; Verbruggen and Logan, 2009b; Congdon et al., 2012).

10.3 Methods

10.3.1 Participants

All participants were recruited through The Great Brain Experiment (www.thegreatbrainexperiment.com, Brown et al., 2014), a smartphone application (app) that is freely available for download in the App Store on iTunes for iOS users, and Google Play for Android users. Between March 11th 2012 and April 3rd 2014 a total of 29,740 participants of at least 18 years old submitted 71,981 datasets ('plays') for the game "Am I Impulsive?". Upon starting the app for the first time participants provided informed consent.

Participants were asked for their age (<18, 18-24, 25-29, 30-39, 40-49, 50-51, 60-69, 70+ years old), gender (male or female), location, education (GCSE or equivalent, a-level or equivalent, degree, post-graduate qualification), life satisfaction (0-10 in steps of 1). As players were only known by an anonymous unique identifier (UID) assigned upon consent no identifiable data were stored (e.g. no IP addresses, email addresses, initials, dates of birth, and so forth). Ethical approval for this experiment was obtained from the UCL Research Ethics Committee. Participants could uninstall the app at any time, stop submitting data, or could request their data to be deleted from the server. These requests were made through the app and preserved anonymity.

10.3.2 Task

The design of the task was highly similar to the task in chapter 9, and indeed I obtained the same SSRT measures from both tasks. A critical difference was that in the task reported in this chapter, the participants were not asked to respond as fast as possible to a Go cue; rather, they were asked to respond within a certain time window after onset of the trial (cf. Coxon et al., 2007). This has implications for the study of the speed-selectivity trade-off: by using a response window the remaining response after a successful stop is not to be executed as quickly as possible, and as such there is no trade-off any more. This is one of the reasons for focusing exclusively on SSRT (at the exclusion of selectivity) in this chapter.

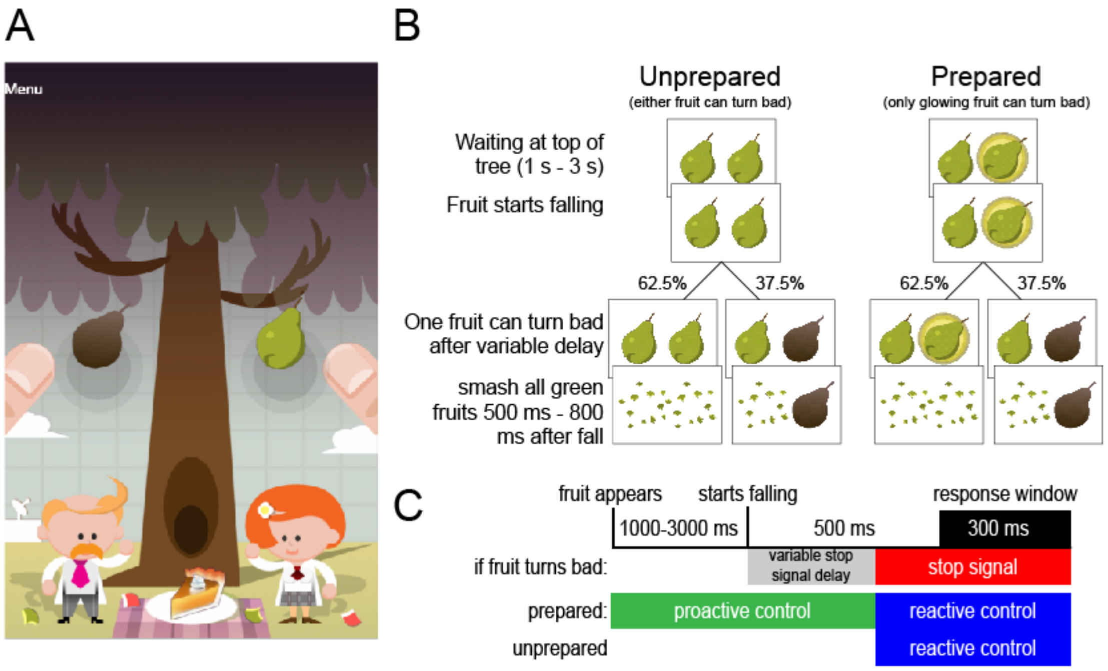


Figure 10.1: Task design for each 2-minute game. (A) The game required participants to smash fruits as they passed over the grey circles. (B) On 37.5% of trials one of the fruits turned brown in mid-flight, prompting the participant to quickly withhold their response only on that side. On Unprepared trials either fruit could turn bad (the example in B shows the right fruit turning bad). On Prepared trials one of the fruits glowed, indicating that only that fruit could turn bad. (C) This information could be used to employ proactive control, and I quantified proactive control as the improvement in performance in Prepared over Unprepared trials.

In this implementation of the task, participants tapped the left and right side of their smartphone or tablet screen to smash two falling fruits (Figure 10.1A). A single trial consisted of the fruits hanging at the top of the screen for 1-3 s (uniformly distributed), followed by the fruits falling down the vertical axis of the screen. When these passed over two circles indicating the response window, spanning 500 to 800 ms following onset of the fall, participants were required to tap both sides of the screen. Out of 32 trials in a single play of the game, a random draw of 12 trials (37.5%) were ‘selective stop trials’ on which one of the fruits turned brown, indicating the corresponding side of the screen should not be tapped. On 16 out of 32 trials (‘Prepared’ condition) a glowing circle around one of the fruits indicated to the participant that this fruit alone might turn brown (which it would do in 6 out of 16 trials, i.e. 37.5%). On the other 16 trials (‘Unprepared’ condition) neither fruit glowed meaning either fruit might turn bad (Figure 10.1B). The stop trials were always ‘selective’, i.e. never did both fruits turn bad in the same trial. As in chapter 9, Prepared trials thus contained extra information concerning the action that might require stopping, allowing the participant to prepare and exert proactive control (Figure 10.1C). The number of ms between the start of the fall and the fruit turning brown is the stop signal delay (SSD). I used separate staircases for the SSD in Prepared and Unprepared stop trials. The staircases started at 300 ms, moved by 50 ms up or down following correct or incorrect stops, respectively, and were reset at each play of 32 trials. Taken together, there were 4 types of trials: Unprepared go, Unprepared stop, Prepared go, and Prepared stop (but no ‘No Stop’ trials as in the lab version of this task). Reactive control was calculated as the SSRT in the Unprepared condition, whereas proactive control was calculated as the *difference* in SSRT between Unprepared and Prepared trials (i.e. the improvement in SSRT with information; see below). All trials types were fully counterbalanced over events on the left and right side of the screen. The order of trials was randomised for each play. Feedback consisted of one of the following statements: ‘You touched too soon!’; ‘You touched too late!’; ‘Touch the fruit inside the circles!’ (in case no buttonpress was detected); ‘Don’t touch the bad fruit!'. A single play of the game took approximately 2 minutes, and was preceded by a short instruction screen.

10.3.3 Participant exclusion

Only completed plays that were immediately sent to the server over an active internet connection were stored. I first discarded data from participants with no correct Go or Stop trials, no failed or successful stop trials in either Prepared or Unprepared trials, or an SSRT that was smaller than or equal to 0 (see below for SSRT estimation). This left us with 22,098 out of a total of 29,740 participants (74%). Unless noted otherwise, I then performed all analyses on participants that submitted 2 plays or more to allow for reliable estimation of the SSRT (12,496 out of 22,098 participants played 2+ games, 57%; Congdon et al., 2012). Data collected over multiple plays by a single participant were concatenated. In all regressions I added the number of submitted games as a nuisance regressor.

10.3.4 Beck Depression Inventory (BDI)

The app contained a messaging system allowing researchers to contact participants through their UID alone. A link to an online version of the BDI-2 (Beck et al., 1961; Beck et al., 1996) was sent to all participants. In addition to a standard set of 21 questions to measure current levels of depression, I asked 5 optional questions about depression history, number of depressive episodes, duration of depression over lifespan, history of anti-depressant medications, and occurrence of depression in immediate family. These data were then matched to the task data based on their UID. In case of duplicate submissions only the first submitted questionnaire was analysed. In the sample of participants (>1 play) the response rate was 4% (509 participants).

10.3.5 Estimation of SSRT

Using the same method as in chapter 9 I calculated SSRT for the Unprepared and Prepared condition separately. In this study, SSRT-Unprepared corresponds to reactive control, and the percentage improvement between Unprepared and Prepared, i.e. 100*(SSRT-Prepared - SSRT-Unprepared)/SSRT-Unprepared, corresponds to proactive control.

For the purpose of SSRT estimation, Go RTs were defined as the first detected button press in the trial (ignoring when the opposite side was pressed). All Go RTs, including those outside the correct response window, were arranged in descending order. I chose to include such trials to more accurately capture the true RT distribution. I excluded 18 participants whose proactive control was larger than 100% or smaller than -100%.

10.3.6 Estimation of selectivity

I estimated the selectivity of inhibition by calculating $RT_{stopSuccess}$ - RT_{Go}. This represents the slowdown in the remaining response when executing a stop compared to when no stop is required. Although the race model predicts a positive value (because successful stop trials are assumed to represent slow response trials), individual variation can capture differences in selectivity of the inhibition process (Coxon et al., 2007, 2009). However, the reliability analysis (see Results) showed that this measure cannot be estimated accurately from the data, so in this chapter I focus on SSRT rather than selectivity.

10.3.7 Statistical analysis

To test the hypotheses I performed linear regressions in R (R Development Core Team, 2008; R Core Team, 2011). I regressed the various dependent variables against models that always included nuisance variables including the number of submitted games, proportion of correct Go trials (both location and timing of press), and operating system (iOS or Android). As regressors of interest I used, across multiple models, scalar variables of age and BDI (both mean-centred; each participant's age set to the centre of their age bin), and factors gender and education. For analyses that included education I excluded participants 18-24 years old as a large proportion might not yet have finished their education. I explicitly did not enter all variables and their interactions into a single large model, as this leads to many effects for which I had no hypotheses or likely interpretations. I used the R-package 'doBy' for post-hoc contrasts (Højsgaard, 2012). For reliability analysis and plotting I used MatLab R2012a (The MathWorks, Inc.). To calculate split-half reliability I always compared even to odd plays (as performed in Williams et al., 1999). Note that I report effect sizes and 95% confidence intervals in ms rather than p-values, as the large sample size makes p-values less informative (Kline and Association, 2004).

10.4 Results

10.4.1 Reliability of SSRT and selectivity estimates

SSRT and selectivity have been used as markers of inhibitory control (e.g. Aron and Verbruggen, 2008; Smittenaar et al., 2013a). As these are summary measures derived for each participant, their reliability depends on the amount of available data. For example, 52% of participants submitted only a single play, which is about 2 minutes' worth of data. Indeed, for participants who played more than once, the intra-class correlation between the first and second play is below 0.4 for both SSRT and selectivity, which is classified as 'poor' (see Figure 10.2; Cicchetti, 2001). Including twice the number of trials leads to a 'good' reliability for SSRT-Unprepared and SSRT-Prepared, whereas the reliability of the selectivity measures remains poor (Figure 10.2). Given these results I excluded those participants who only submitted a single game, and focused the analyses on SSRT rather than selectivity.

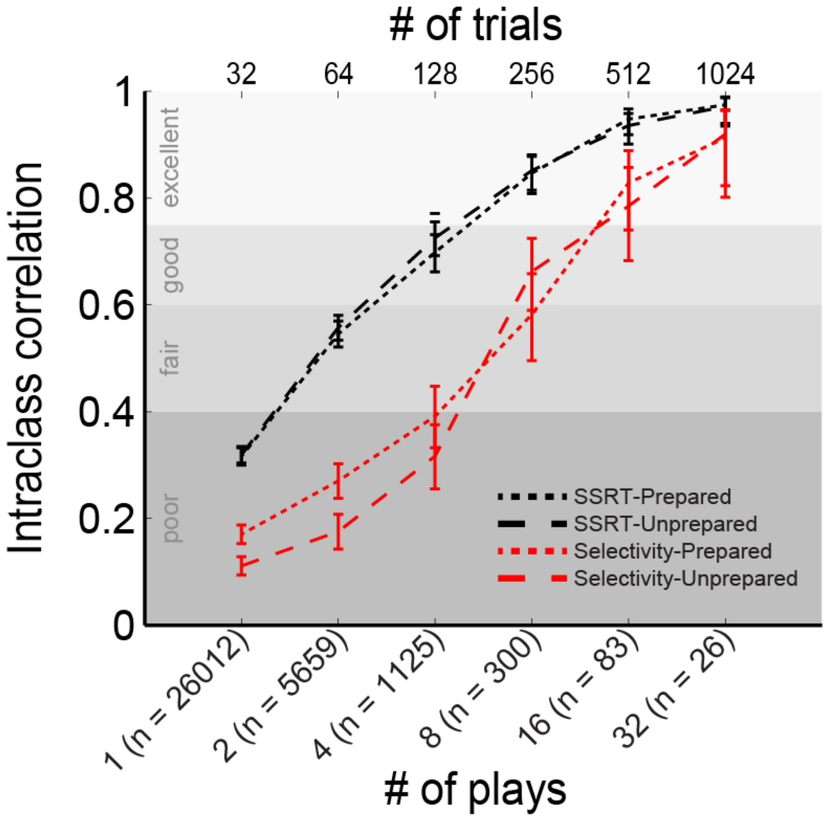


Figure 10.2: Reliability of SSRT and selectivity depends on the number of available plays per participant. The SSRT represents the duration of the inhibitory process, i.e. the speed of inhibition (see Methods). Selectivity represents the slowing that occurs on the concurrent response when an action is inhibited. Both measures can be estimated for the Prepared and Unprepared conditions separately. Here I use the intra-class correlation (ICC) to quantify the reliability of these measures as a function of the number of trials that are used in the estimation. If SSRT and selectivity are estimated from the first game only and compared to the second game (# of plays = 1 in the figure), reliability is poor for all measures (following criteria from Cicchetti, 2001). As more games are used for estimation reliability increases, although reliable estimation of selectivity requires approximately 4 times as much data as reliable estimation of SSRT as shown by the rightward shift of the selectivity curves compared to

SSRT curves. Error bars represent 95% CI on the ICC, n = number of participants for whom sufficient plays were available.

10.4.2 Applicability of the independent horse race model

Despite good reliabilities, the relatively small number of trials included in the calculation of some of the SSRTs might make the independent horse race model unsuitable (Logan, 1994; Verbruggen and Logan, 2009b). Participants with 2 or more submitted plays had at least 12 stop trials and 20 go trials in the Prepared and Unprepared condition each. This is half the number of trials per condition that Congdon et al. (2012) showed were needed to yield a reliable estimate of SSRT. To ascertain the applicability of the horse race model to these data I checked a range of assumptions of the model (Figure 10.3). Firstly, on trials where the participant fails to stop, the RT tends to be faster than the average Go RT, in line with the prediction that failed stop trials represent the fast part of the Go RT distribution (Figure 10.3A-B; Prepared: mean RT_{Go} - $RT_{stopFail}$ = 22.8 ± 0.6 ms; Unprepared: mean RT_{Go} - $RT_{stopFail}$ = 25.3 ± 0.6 ms). Secondly, the later the stop-signal was presented, i.e. the longer the stop-signal delay, the lower the probability of stopping (Figure 10.3C) and the higher the reaction time on failed stop trials (Figure 10.3D). Together these results suggest an independent horse race model is applicable to data from both the Unprepared and Prepared conditions in participants with 2 or more submitted plays.

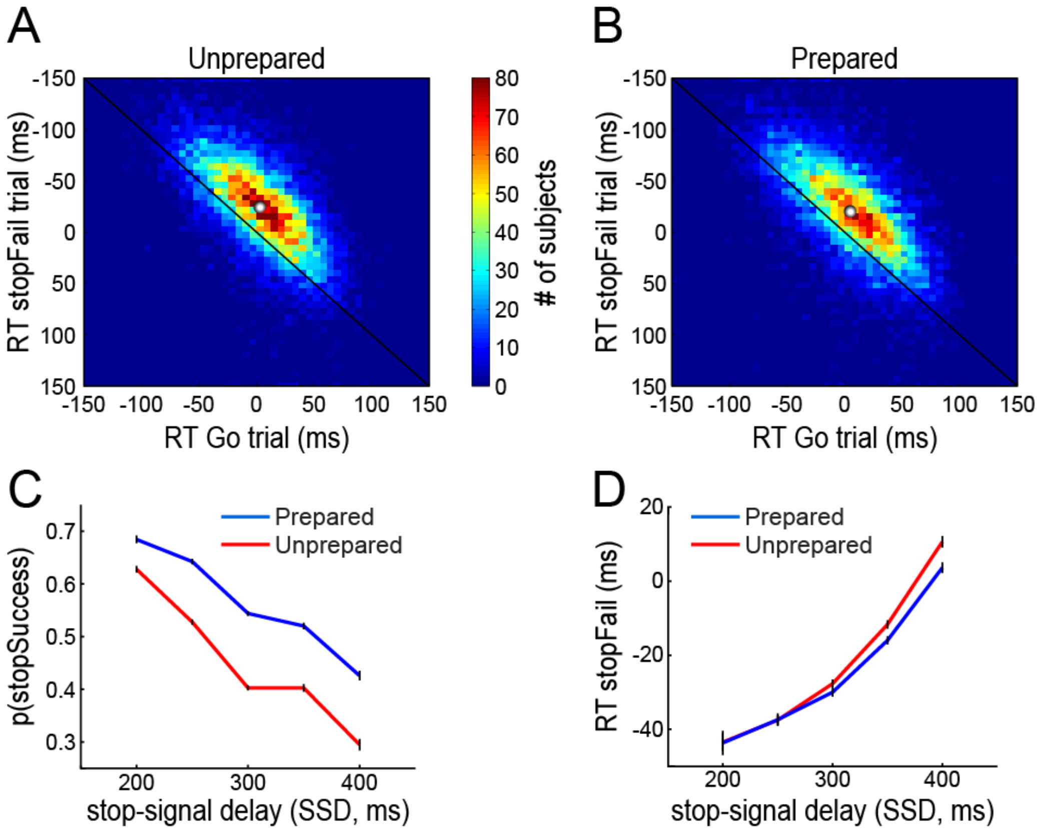


Figure 10.3: The Unprepared and Prepared condition satisfy assumptions of the race model used to calculate SSRT. The horse race model (Verbruggen and Logan, 2009b) assumes that a stopping process with fixed duration is set off as soon as the stop signal is presented. This process then catches up with the Go process only if the stop was initiated far enough in advance of the Go response, i.e. 1) if the Go process happened to be slow on that trial and/or 2) if the stop-signal was presented early. Confirming the first prediction, in stopFail trials (where the participant erroneously responds and thus fails to stop) the reaction times are on average faster than in Go trials, in both the Unprepared (A) and Prepared (B) condition. Confirming the second prediction, the later the stop-signal was presented (i.e. the later the fruit turned brown) the lower the chance of stopping successfully (C). Lastly, if the SSD is large the stop process cannot catch up even with slow Go responses; this predicts that the average stopFail RT will go up with larger SSDs, which is indeed the case (D). All RTs are shown relative to the centre of the response window, which was at 650 ms after the start of the fall. SSDs are relative to the start of the fall. Error bars indicate 95% confidence intervals. White dots in A and B indicate population means.

10.4.3 Preparation improves speed of inhibition

As expected, preparation (proactive control) improved the SSRT (Unprepared - Prepared, 31.9 ± 1.1 ms improvement). To examine how the between-participant factors affected inhibitory control I examined SSRT-Unprepared as a measure of reactive control, and the improvement from Unprepared to Prepared as a measure of proactive control.

10.4.4 Demographics of proactive and reactive control

In a regression of SSRT-Unprepared on age, gender, age-by-gender and three nuisance variables (see Methods), I observed 18-24 year old women are 9.97 ± 1.99 ms slower than men. However, there is also a significant age-by-gender interaction on this measure of reactive control (Figure 10.4A, red lines). That is, whereas men on average deteriorate by 10.1 ± 1.22 ms per decade, women do so only at 8.2 ± 1.05 ms per decade (1.9 ± 1.57 ms per decade slower than men). Reactive inhibitory control thus declines more slowly in women than men.

For proactive control (the difference between Unprepared and Prepared SSRT), 18-24 year old women showed a larger improvement with preparation (Figure 10.4B; 1.95 ± 0.57 % point difference). This effect of gender on proactive control, if anything, became stronger with older age (the gender difference increased by 0.46 ± 0.45 % point per decade). Although Figure 10.4B suggests a trend whereby proactive control increases with age, this was not significant (-0.01 ± 0.03 ms) and therefore presumably captured by the nuisance regressors. Together, these results show that reactive, but not proactive, control deteriorates with age, but more so in men than women (Figure 10.4A). Furthermore, women experience greater benefits from proactive control across all ages (Figure 10.4B).

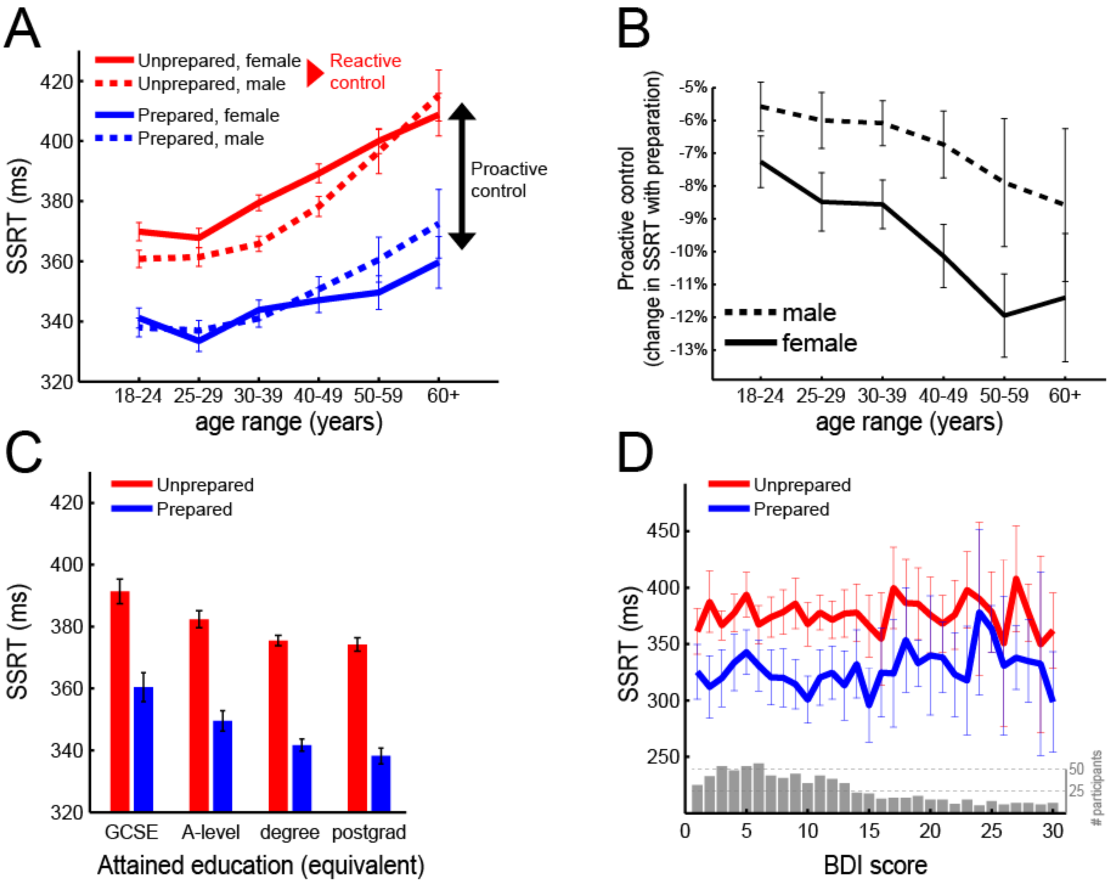


Figure 10.4: Demographics of proactive and reactive control. (A) SSRT-Unprepared, which measures the speed of reactive control, increases with age. However, this age-related decline is more rapid in men than women. Proactive control is quantified as the difference between Unprepared and Prepared SSRT, i.e. the amount by which inhibition is improved through preparation (difference between red and blue line). (B) In proactive control strikingly different pattern is observed. Relative to performance in Unprepared trials, women improve more with preparation across all ages, with a slight increase in this improvement with age. Although the lines seem to have a negative slope, there was no evidence for an effect of age on proactive control. The y-axis represents the improvement in SSRT between Unprepared and Prepared as % of SSRT-Unprepared, such that more negative values indicate greater benefit of preparation on the speed of inhibition. (C) Higher attained education in participants aged 25 or over is not only associated with better reactive control (reduction in SSRT-Unprepared), but also with a better proactive control (larger difference between red and blue bar). (D) The BDI scores were distributed as

shown in the grey histogram. No relationship with SSRT-Unprepared or proactive control was apparent. All error bars indicate 95% CI. BDI = Beck Depression Inventory; GCSE = general certificate of secondary education; A-level = general certificate of education advanced level.

For education I also examined SSRT-Unprepared and the difference between SSRT-Unprepared and SSRT-Prepared to distinguish reactive versus proactive control, respectively (Figure 10.4C). Progressively higher levels of education were associated with better SSRT-Unprepared (compared to GCSE, in ms: a-level, -4.8 ± 4.1; degree, -7.1 ± 3.7; postgrad, -8.4 ± 4.0). Similarly, proactive control increased with education (compared to GCSE, in % points: a-level, -1.6 ± 1.2; degree, -2.6 ± 1.1; postgrad, -3.3 ± 1.2). Together, this shows that a higher level of education is associated with better reactive as well as proactive control.

Lastly I explored possible relationships between depression indices and SSRT (Figure 10.4D). In a regression identical to the age-by-gender regression above, but with BDI scores added as a predictor, I observed no relationship between the BDI score and SSRT-Unprepared (change in SSRT with every point on BDI scale, 0.05 ± 0.39 ms) or proactive control (0.04 ± 0.14 % points).

10.5 Discussion

This chapter describes a dataset acquired through smartphones in which I examined inhibitory control in a strongly heterogeneous group of participants in terms of their age, gender, education and depressive symptoms. I showed that these data conform to assumptions of the widely used independent horse race model, which was used to estimate the SSRT in the Prepared and Unprepared conditions. These measures can be reliably estimated from the data even with only ~4 minutes of data (64 trials). I show that reactive control, operationalised as SSRT in the Unprepared condition, deteriorates with age, and this occurs faster in men than women. Proactive control, operationalised as the change in SSRT in the Prepared compared to the Unprepared condition, did not significantly change with age. Moreover, proactive control was stronger in women compared to men at all ages. Additionally, higher levels of education are associated with better reactive as well as proactive inhibitory control, whereas depression shows no relationship with either measure of inhibitory self-control.

There are many benefits as well as limitations to the use of smartphones in cognitive science (Dufau et al., 2011). As in lab-based (Henrich et al., 2010) and online (Chandler et al., 2014) cognitive research a sampling bias is assumed which is compounded by the cross-sectional approach as in lab studies on aging (e.g. Salthouse, 2009). Broad sampling of the wider population is a first step to understanding how our models of cognition apply beyond the common context of Western university students. For example, another game from The Great Brain Experiment showed that a computational model of subjective well-being developed in the lab could be used to predict well-being in the broader population (Rutledge et al., 2014). Smartphones, given their ubiquity not only in the West but worldwide (Bicheno, 2012), provide a cost-efficient and societally engaging way of achieving this goal.

These results extend previous research on inhibition in multiple ways. I replicate the finding that reactive control deteriorates with age (Williams et al., 1999; Bedard et al., 2002; Coxon et al., 2012). Unlike these previous studies I observed an age-by-gender interaction. This adds to an existing controversy on differential cognitive decline in men and women, with some reports finding faster decline in men (on a cognitive battery; Maylor et al., 2007), and others finding no differential decline (in spatial ability; Willis and Schaie, 1988; Driscoll et al., 2005), and yet others finding women decline faster (on simple and choice RT; Der and Deary, 2006). This cognitive heterogeneity is surprising given the more rapid age-related neural decline consistently found in men compared to women (Gur et al., 1991; Cowell et al., 1994; Murphy et al., 1996; Coffey et al., 1998; Good et al., 2002), including in prefrontal regions critical for inhibitory control (chapter 9 and Majid et al., 2013). Future work, for example on white-matter connectivity rather than brain volume (Coxon et al., 2012), might shed more light on the neural underpinnings of age- and gender-related decline and maintenance of cognitive function. It would be particularly relevant to understand how proactive control performance is maintained in the face of neural decline, and how this aligns with theories of proactive control in aging (Braver, 2012; Lindenberger and Mayr, 2014).

In contrast to this gender-by-age effect on reactive control, proactive control was greater (i.e. SSRT improved more with preparation) in women compared to men at all ages. That is, whereas advance information benefits the speed of inhibition as observed previously (Aron and Verbruggen, 2008; Jahfari et al., 2012; Smittenaar et al., 2013a), women improve more than men at all ages. This suggests that preparatory benefits are not simply a function of baseline performance whereby worse performers show larger improvements.

Education is often controlled for (e.g. Monterosso et al., 2005) or measured but not reported on (e.g. Bedard et al., 2002; Rucklidge and Tannock, 2002) in the response inhibition literature. To the best of my knowledge there have been no previous reports showing education is associated with better reactive and proactive inhibitory control. Attention-deficit/hyperactivity disorder (ADHD) is characterised by impaired response inhibition (Casey et al., 1997) as well as poor educational attainment (Loe and Feldman, 2007). Even adults who show symptoms and/or behavioural indications of ADHD during early childhood, but who are never formally diagnosed, show poor education attainment (Lambert, 1988). This suggests that impulsivity traits contribute to educational performance, in addition to potential effects whereby poor education leads to impaired self-control.

The approach presented in this chapter shows the power of harnessing large-scale public participation in psychology and academic research (Bonney et al., 2014). By transforming often tedious laboratory tasks into engaging games (Kim et al., 2014), researchers can now engage the public in research at vast scales without compromising their experimental designs.

11 General discussion

This dissertation describes the use of computational modelling, neuromodulation and neuroimaging to investigate the neural and behavioural correlates of reward-guided and inhibitory action control.

In chapter 2 I reviewed evidence that reinforcement learning can be parsed into model-free and model-based components. I went on to study how levodopa (chapter 5) and transcranial neurostimulation (chapters 6 and 7) can modulate these specific components of reinforcement learning. In chapter 8 I used high-resolution imaging to investigate the functional neuroanatomy of frontostriatal reward learning networks.

In the second part of this thesis I studied proactive and reactive inhibitory control which, like reward learning, requires adjustments to action based on uncertain information from the environment. In chapter 9 I studied the frontostriatal networks that subserve proactive control over response inhibition.

In the final study of this thesis I studied proactive and reactive inhibitory control in a much larger and more diverse sample than is available for laboratory based studies through a smartphone application, revealing how inhibitory control varies by age and gender (chapter 10).

Having discussed the conclusions and implications of each individual study in their respective chapters, in this final discussion I touch upon two issues that are of a more general nature; firstly whether model-based and model-free control can really be considered distinct and secondly; how recent advances in human neuroimaging might drive a more anatomically grounded view of striatal function. I examine these in light of previous studies in the field as well as my own work presented throughout this thesis.

11.1 How distinct are model-based and model-free control?

Throughout this thesis I deliberately drew a clear-cut distinction between model-based and model-free influences over behaviour. It is more likely, however, that these two forms of control represent opposite poles of a continuous spectrum. This continuous view of control is reflected in the many machine learning algorithms that borrow elements from both classes of model (Sutton, 1990; Sutton and Barto, 1998). Furthermore a range of recent findings in cognitive neuroscience support the notion of a spectrum of control (Daw et al., 2011; Gershman et al., 2014).

For example, in chapter 2 I introduced dopaminergic and ventral striatal reward prediction errors as a fundamentally model-free, stimulus-response learning mechanism. Indeed, a purely model-based controller would have no use for stimulus-response reward prediction error in most naturalistic environments. It came as a surprise, then, when Daw et al. (2011) reported that ventral striatal BOLD responses can best be explained by modelling not only a model-free but also a model-based component to the reward prediction error (also see Deserno et al., 2015). Equally surprising was the finding that the lesions of the ventral striatum abolish model-based reward identity learning (McDannald et al., 2011). A more formal departure from fully segregated systems is the Dyna algorithm (Sutton, 1990). Here, a model of the environment replays events and action sequences to train an instrumental learning mechanism, e.g. by replaying events through models of the environment represented in the hippocampus to generate reward learning signals in the striatum (as suggested by Johnson and Redish, 2005). Behavioural evidence of such model-based instruction of a model-free system has now been observed in humans (Gershman et al., 2014).

In their task a model-based and model-free system individually predicted that participants would make one choice, whereas a model-based system training a model-free actor would predict an alternative choice. The latter choices were indeed observed, a result that was replicated in follow up experiments (Gershman et al., 2014).

Taken together, the above studies provide compelling evidence that model-based and model-free control exist along a continuum. The question then is not *if* there are multiple forces driving behaviour, but rather *at what point* during a decision and across the timespan of learning do models of the world exert their influence. How does this literature on dual processes relate to the findings from this thesis?

As I alluded to in the discussion of chapter 5, it seems that we need to re-evaluate some historical work on instrumental learning. In a critical paper Collins and Frank (2012) showed that even simple, 1-step reinforcement learning problems such as the one in chapter 8 are now known to engage both model-free and model-based mechanisms. This means than when behaviour on such a task is altered by Parkinson's disease (Frank et al., 2004; Bodi et al., 2009), dopaminergic drugs (Pessiglione et al., 2006), psychosis (Murray et al., 2008), schizophrenia (Koch et al., 2010), genetic traits (Frank et al., 2007b), ageing (Chowdhury et al., 2013), adolescent development (van den Bos et al., 2012) or addiction (Redish, 2004) it is difficult to tell whether this is due to a change in a model-free or model-based component of choice, or indeed in their interaction. Chapter 5 showed the surprising result that levodopa in healthy participants had no effect whatsoever on model-free learning, but boosted model-based control. In a 1-step learning task such a dopamine-induced boost in model-based control might well be expressed through faster learning rates in a (model-free) temporal difference reinforcement learning model, leading to potentially erroneous conclusions, for example that dopamine is impacting on model-free prediction errors.

In conclusion, more work is needed to understand the intricate organization of decision-making, probably through the development of tasks that finely place behavioural control on a spectrum rather than categorise it into discrete forms. Until we have such an understanding, tasks that can provide more insightful and nuanced accounts of behaviour should be favoured over more basic tasks that until now have provided the foundation for our study of reinforcement learning.

11.2 The specificity of 'frontostriatal' systems

Corticostriatal systems are important for reward learning and decision-making, but also for virtually every other domain of cognitive neuroscience. Indeed, the study of frontostriatal systems has strong momentum at present —about 20% of all papers on the topic have been published in the last 2 years. But despite three decades of work on functionally segregated loops through the basal ganglia (Alexander et al., 1986; Averbeck et al., 2014), many functional studies in humans still refer to the striatum as a unitary structure. Indeed, given that the striatum to some extent represents a microcosm of the cortex, a study concluding that a task 'activated the striatum' is akin to stating a task 'activated the cortex' if further anatomical specificity is omitted.

The remarkable topography of the striatum provides an opportunity to understand what specific corticostriatal loops are involved in cognitive and motor functions. As such, the anatomical rigour applied to, for example, prefrontal cortex (Rushworth et al., 2011; Haber and Behrens, 2014) or the visual system (Lund, 1988; Zeki et al., 1991) could also be leveraged in the striatum. Studies that do divide the striatum into parts often limit these efforts to three areas: the ventral striatum, putamen and caudate (e.g. chapter 9 and O'Doherty et al., 2004). In reality, the functional and anatomical division is much more subtle (Graybiel and Ragsdale, 1978; Averbeck et al., 2014; van den Bos et al., 2014).

In chapter 8 I took a data-driven approach to understanding how local differences in corticostriatal connectivity relate to local differences in functional responses. The underlying idea is that function in the striatum is not easily defined along anatomical landmarks, as might be the case in motor or visual cortex. Indeed, the border between nucleus accumbens, caudate and putamen is at best highly ambiguous. Using the connectivity fingerprint of individual voxels is a promising method to define functional zones within the striatum (Draganski et al., 2008; Saygin et al., 2012; Averbeck et al., 2014). As methods for diffusion imaging and probabilistic tractography improve (Glasser et al., 2013; Sotiropoulos et al., 2013; Van Essen et al., 2013) we can expect to see more studies beginning to explain computations and representations across the brain in terms of anatomical connectivity. This promises to complement the large number of studies of functional connectivity (Fox and Raichle, 2007; Bullmore and Sporns, 2009; Fries, 2009; Friston and Dolan, 2010), such as resting state fMRI which has already been used to segment the striatum (Di Martino et al., 2008; Robinson et al., 2009; Barnes et al., 2010; Helmich et al., 2010; Choi et al., 2012). Indeed, whereas functional connectivity has been convincingly used to study frontostriatal interactions in the competition between model-based and model-free control (Wunderlich et al., 2012b; Lee et al., 2014), a similarly rigorous approach using structural connectivity has to the best of my knowledge not been applied.

In summary, the argument put forward is that not all striatal activations are created equal. By using a combination of functional and anatomical connectivity methods we will hopefully develop a finer scalpel to study the origin of corticostriatal activations and their role in cognition.

References

Abe H, Lee D (2011) Distributed coding of actual and hypothetical outcomes in the orbital and dorsolateral prefrontal cortex. Neuron 70:731-741.

Adams CD, Dickinson A (1981) Instrumental responding following reinforcer devaluation. The Quarterly journal of experimental psychology 33:109-121.

Ahsan RL, Allom R, Gousias IS, Habib H, Turkheimer FE, Free S, Lemieux L, Myers R, Duncan JS, Brooks DJ (2007) Volumes, spatial extents and a probabilistic atlas of the human basal ganglia and thalamus. NeuroImage 38:261-270.

Aizman O, Brismar H, Uhlén P, Zettergren E, Levey AI, Forssberg H, Greengard P, Aperia A (2000) Anatomical and physiological evidence for D1 and D2 dopamine receptor colocalization in neostriatal neurons. Nature Neuroscience 3:226-230.

Alexander GE, Crutcher MD (1990) Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends in Neurosciences 13:266-271.

Alexander GE, DeLong MR, Strick PL (1986) Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu Rev Neurosci 9:357-381.

Andersson JL, Skare S, Ashburner J (2003) How to correct susceptibility distortions in spin-echo echo-planar images: application to diffusion tensor imaging. Neuroimage 20:870-888.

Andersson JL, Hutton C, Ashburner J, Turner R, Friston K (2001) Modeling geometric deformations in EPI time series. NeuroImage 13:903-919.

Aron AR (2011) From reactive to proactive and selective control: developing a richer model for stopping inappropriate responses. Biol Psychiatry 69:e55-68.

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

Aron AR, Verbruggen F (2008) Stop the presses: dissociating a selective from a global mechanism for stopping. Psychol Sci 19:1146-1153.

Ashburner J, Friston KJ (2005) Unified segmentation. NeuroImage 26:839-851.

Attwell D, Iadecola C (2002) The neural basis of functional brain imaging signals. Trends in Neurosciences 25:621-625.

Attwell D, Buchan AM, Charpak S, Lauritzen M, MacVicar BA, Newman EA (2010) Glial and neuronal control of brain blood flow. Nature 468:232-243.

Averbeck BB, Lehman J, Jacobson M, Haber SN (2014) Estimates of projection overlap and zones of convergence within Frontal-Striatal Circuits. The Journal of Neuroscience 34:9497-9505.

Baddeley A (2012) Working memory: theories, models, and controversies. Annual review of psychology 63:1-29.

Badry R, Mima T, Aso T, Nakatsuka M, Abe M, Fathi D, Foly N, Nagiub H, Nagamine T, Fukuyama H (2009) Suppression of human cortico-motoneuronal excitability during the Stop-signal task. Clinical Neurophysiology 120:1717-1723.

Baker JM, Rorden C, Fridriksson J (2010a) Using transcranial direct-current stimulation to treat stroke patients with aphasia. Stroke 41:1229-1236.

Baker KB, Lee JY, Mavinkurve G, Russo GS, Walter B, DeLong MR, Bakay RA, Vitek JL (2010b) Somatotopic organization in the internal segment of the globus pallidus in Parkinson's disease. Experimental neurology 222:219-225.

Balleine BW, Dickinson A (1998) Goal-directed instrumental action: contingency and incentive learning and their cortical substrates. Neuropharmacology 37:407-419.

Balleine BW, O'Doherty JP (2010) Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology 35:48-69.

Band GP, van der Molen MW, Logan GD (2003) Horse-race model simulations of the stop-signal procedure. Acta psychologica 112:105-142.

Bari A, Robbins TW (2013) Inhibition and impulsivity: behavioral and neural basis of response control. Progress in Neurobiology 108:44-79.

Barker AT, Jalinous R, Freeston IL (1985) Non-invasive magnetic stimulation of human motor cortex. The Lancet 325:1106-1107.

Barkley RA (1997) Behavioral inhibition, sustained attention, and executive functions: constructing a unifying theory of ADHD. Psychological bulletin 121:65.

Barnes KA, Cohen AL, Power JD, Nelson SM, Dosenbach YB, Miezin FM, Petersen SE, Schlaggar BL (2010) Identifying basal ganglia divisions in individuals using resting-state functional connectivity MRI. Frontiers in systems neuroscience 4.

Barto AG, Sutton RS (1982) Simulation of anticipatory responses in classical conditioning by a neuron-like adaptive element. Behavioural Brain Research 4:221-235.

Baruzzi A, Contin M, Riva R, Procaccianti G, Albani F, Tonello C, Zoni E, Martinelli P (1987) Influence of meal ingestion time on pharmacokinetics of orally administered levodopa in parkinsonian patients. Clinical neuropharmacology 10:527-537.

Basser PJ, Mattiello J, LeBihan D (1994) MR diffusion tensor spectroscopy and imaging. Biophysical journal 66:259-267.

Basser PJ, Pajevic S, Pierpaoli C, Duda J, Aldroubi A (2000) In vivo fiber tractography using DT-MRI data. Magnetic Resonance in Medicine 44:625-632.

Bates D, Maechler M, Bolker B (2012) lme4: Linear mixed-effects models using S4 classes.

Bay NS-Y, Bay B-H (2010) Greek anatomist herophilus: the father of anatomy. Anatomy & cell biology 43:280-283.

Beck AT, Steer RA, Ball R, Ranieri WF (1996) Comparison of Beck Depression Inventories-IA and-II in psychiatric outpatients. Journal of personality assessment 67:588-597.

Beck AT, Ward CH, Mendelson ML, Mock JE, Erbaugh JK (1961) An inventory for measuring depression. Archives of general psychiatry 4:561-571.

Beckstead RM, Domesick VB, Nauta WJ (1993) Efferent connections of the substantia nigra and ventral tegmental area in the rat. In: Neuroanatomy, pp 449-475: Springer.

Bedard A-C, Nichols S, Barbosa JA, Schachar R, Logan GD, Tannock R (2002) The development of selective inhibitory control across the life span. Developmental neuropsychology 21:93-111.

Behrens T, Johansen-Berg H, Woolrich M, Smith S, Wheeler-Kingshott C, Boulby P, Barker G, Sillery E, Sheehan K, Ciccarelli O (2003a) Non-invasive mapping of connections between human thalamus and cortex using diffusion imaging. Nature neuroscience 6:750-757.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007a) Learning the value of information in an uncertain world. Nat Neurosci 10:1214-1221.

Behrens TE, Hunt LT, Woolrich MW, Rushworth MF (2008) Associative learning of social value. Nature 456:245-249.

Behrens TE, Berg HJ, Jbabdi S, Rushworth MF, Woolrich MW (2007b) Probabilistic diffusion tractography with multiple fibre orientations: What can we gain? Neuroimage 34:144-155.

Behrens TE, Woolrich MW, Jenkinson M, Johansen-Berg H, Nunes RG, Clare S, Matthews PM, Brady JM, Smith SM (2003b) Characterization and propagation of uncertainty in diffusion-weighted MR imaging. Magn Reson Med 50:1077-1088.

Bellman R (1956) Dynamic programming and Lagrange multipliers. Proceedings of the National Academy of Sciences of the United States of America 42:767.

Berridge KC (2007) The debate over dopamine’s role in reward: the case for incentive salience. Psychopharmacology 191:391-431.

Bestmann S, Ruff CC, Blankenburg F, Weiskopf N, Driver J, Rothwell JC (2008) Mapping causal interregional influences with concurrent TMS–fMRI. Experimental Brain Research 191:383-402.

Bicheno S (2012) Global Smartphone Installed Base Forecast by Operating System for 88 Countries: 2007 to 2017. Strategy Analytics.

Bodi N, Keri S, Nagy H, Moustafa A, Myers CE, Daw N, Dibo G, Takats A, Bereczki D, Gluck MA (2009) Reward-learning and the novelty-seeking personality: a between-and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain 132:2385-2395.

Bogacz R, Gurney K (2007) The basal ganglia and cortex implement optimal decision making between alternative actions. Neural computation 19:442-477.

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

Boggio PS, Ferrucci R, Rigonatti SP, Covre P, Nitsche M, Pascual-Leone A, Fregni F (2006) Effects of transcranial direct current stimulation on working memory in patients with Parkinson's disease. J Neurol Sci 249:31-38.

Boggio PS, Rigonatti SP, Ribeiro RB, Myczkowski ML, Nitsche MA, Pascual-Leone A, Fregni F (2008) A randomized, double-blind clinical trial on the efficacy of cortical direct current stimulation for the treatment of major depression. Int J Neuropsychopharmacol 11:249-254.

Boggio PS, Bermpohl F, Vergara AO, Muniz AL, Nahas FH, Leme PB, Rigonatti SP, Fregni F (2007) Go-no-go task performance improvement after anodal transcranial DC stimulation of the left dorsolateral prefrontal cortex in major depression. Journal of affective disorders 101:91-98.

Bond A, Lader M (1974) The use of analogue scales in rating subjective feelings. British Journal of Medical Psychology 47:211-218.

Bonney R, Shirk JL, Phillips TB, Wiggins A, Ballard HL, Miller-Rushing AJ, Parrish JK (2014) Citizen science. Next steps for citizen science. Science 343:1436-1437.

Boorman ED, Behrens TE, Woolrich MW, Rushworth MF (2009) How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 62:733-743.

Boucher L, Palmeri TJ, Logan GD, Schall JD (2007) Inhibitory control in mind and brain: an interactive race model of countermanding saccades. Psychol Rev 114:376-397.

Boureau YL, Dayan P (2011) Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 36:74-97.

Boynton GM, Engel SA, Glover GH, Heeger DJ (1996) Linear systems analysis of functional magnetic resonance imaging in human V1. The Journal of Neuroscience 16:4207-4221.

Braver TS (2012) The variable nature of cognitive control: a dual mechanisms framework. Trends Cogn Sci 16:106-113.

Brett M, Anton J-L, Valabregue R, Poline J-B (2002) Region of interest analysis using the MarsBar toolbox for SPM 99. NeuroImage 16:S497.

Brown HR, Zeidman P, Smittenaar P, Adams RA, McNab F, Rutledge RB, Dolan RJ (2014) Crowdsourcing for cognitive science--the utility of smartphones. PLoS One 9:e100662.

Brown P (2003) Oscillatory nature of human basal ganglia activity: relationship to the pathophysiology of Parkinson's disease. Movement Disorders 18:357-363.

Bryson Jr AE (1996) Optimal control-1950 to 1985. Control Systems, IEEE 16:26-33.

Bullmore E, Sporns O (2009) Complex brain networks: graph theoretical analysis of structural and functional systems. Nature Reviews Neuroscience 10:186-198.

Button KS, Ioannidis JP, Mokrysz C, Nosek BA, Flint J, Robinson ES, Munafo MR (2013) Power failure: why small sample size undermines the reliability of neuroscience. Nat Rev Neurosci 14:365-376.

Cai W, Oldenkamp CL, Aron AR (2011a) A proactive mechanism for selective suppression of response tendencies. J Neurosci 31:5965-5969.

Cai X, Kim S, Lee D (2011b) Heterogeneous Coding of Temporally Discounted Values in the Dorsal and Ventral Striatum during Intertemporal Choice. Neuron 69:170-182.

Calabresi P, Picconi B, Tozzi A, Di Filippo M (2007) Dopamine-mediated regulation of corticostriatal synaptic plasticity. Trends in Neurosciences 30:211-219.

Calabresi P, Picconi B, Tozzi A, Ghiglieri V, Di Filippo M (2014) Direct and indirect pathways of basal ganglia: a critical reappraisal. Nature Neuroscience 17:1022-1030.

Callaghan MF, Freund P, Draganski B, Anderson E, Cappelletti M, Chowdhury R, Diedrichsen J, Fitzgerald TH, Smittenaar P, Helms G, Lutti A, Weiskopf N (2014) Widespread age-related differences in the human brain microstructure revealed by quantitative magnetic resonance imaging. Neurobiology of aging 35:1862-1872.

Casey B, Castellanos FX, Giedd JN, Marsh WL, Hamburger SD, Schubert AB, Vauss YC, Vaituzis AC, Dickstein DP, Sarfatti SE (1997) Implication of right frontostriatal circuitry in response inhibition and attention-deficit/hyperactivity disorder. Journal of the American Academy of Child & Adolescent Psychiatry 36:374-383.

Chandler J, Mueller P, Paolacci G (2014) Nonnaïveté among Amazon Mechanical Turk workers: Consequences and solutions for behavioral researchers. Behavior research methods 46:112-130.

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

Choi EY, Yeo BTT, Buckner RL (2012) The organization of the human striatum estimated by intrinsic functional connectivity.

Chowdhury R, Guitart-Masip M, Lambert C, Dayan P, Huys Q, Düzel E, Dolan RJ (2013) Dopamine restores reward prediction errors in old age. Nature Neuroscience 16:648-653.

Cicchetti DV (2001) Methodological commentary the precision of reliability and validity estimates re-visited: distinguishing between clinical and statistical significance of sample size requirements. Journal of Clinical and Experimental Neuropsychology 23:695-700.

Civier O, Bullock D, Max L, Guenther FH (2013) Computational modeling of stuttering caused by impairments in a basal ganglia thalamo-cortical circuit involved in syllable selection and initiation. Brain and language 126:263-278.

Claffey MP, Sheldon S, Stinear CM, Verbruggen F, Aron AR (2010) Having a goal to stop action is associated with advance control of specific motor representations. Neuropsychologia 48:541-548.

Clatworthy PL, Lewis SJ, Brichard L, Hong YT, Izquierdo D, Clark L, Cools R, Aigbirhio FI, Baron J-C, Fryer TD (2009) Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. The Journal of Neuroscience 29:4690-4696.

Coffey CE, Lucke JF, Saxton JA, Ratcliff G, Unitas LJ, Billig B, Bryan RN (1998) Sex differences in brain aging: a quantitative magnetic resonance imaging study. Archives of neurology 55:169-179.

Cohen J (1992) A power primer. Psychological bulletin 112:155.

Cohen JY, Haesler S, Vong L, Lowell BB, Uchida N (2012) Neuron-type-specific signals for reward and punishment in the ventral tegmental area. Nature 482:85-88.

Collins AG, Frank MJ (2012) How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. European Journal of Neuroscience 35:1024-1035.

Congdon E, Mumford JA, Cohen JR, Galvan A, Canli T, Poldrack RA (2012) Measurement and reliability of response inhibition. Frontiers in psychology 3:37.

Cools R (2006) Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease. Neurosci Biobehav Rev 30:1-23.

Cools R (2011) Dopaminergic control of the striatum for high-level cognition. Curr Opin Neurobiol 21:402-407.

Cools R, Ivry RB, D'Esposito M (2006) The human striatum is necessary for responding to changes in stimulus relevance. J Cogn Neurosci 18:1973-1983.

Cools R, Barker RA, Sahakian BJ, Robbins TW (2003) L-Dopa medication remediates cognitive inflexibility, but increases impulsivity in patients with Parkinson's disease. Neuropsychologia 41:1431-1441.

Cools R, Stefanova E, Barker RA, Robbins TW, Owen AM (2002) Dopaminergic modulation of high-level cognition in Parkinson's disease: the role of the prefrontal cortex revealed by PET. Brain: A Journal of Neurology 125:584-594.

Cools R, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M (2008) Working memory capacity predicts dopamine synthesis capacity in the human striatum. Journal of Neuroscience 28:1208-1208.

Corbit LH, Balleine BW (2003) The role of prelimbic cortex in instrumental conditioning. Behav Brain Res 146:145-157.

Cotzias GC, Papavasiliou PS, Gellene R (1969) Modification of Parkinsonism—chronic treatment with L-dopa. New England Journal of Medicine 280:337-345.

Courville AC, Daw ND, Touretzky DS (2006) Bayesian theories of conditioning in a changing world. Trends in Cognitive Sciences 10:294-300.

Cowell PE, Turetsky BI, Gur RC, Grossman RI, Shtasel D, Gur R (1994) Sex differences in aging of the human frontal and temporal lobes. The Journal of Neuroscience 14:4748-4755.

Coxon JP, Stinear CM, Byblow WD (2007) Selective inhibition of movement. J Neurophysiol 97:2480-2489.

Coxon JP, Stinear CM, Byblow WD (2009) Stop and go: the neural basis of selective movement prevention. J Cogn Neurosci 21:1193-1203.

Coxon JP, Van Impe A, Wenderoth N, Swinnen SP (2012) Aging and inhibitory control of action: cortico-subthalamic connection strength predicts stopping performance. J Neurosci 32:8401-8412.

Cragg SJ, Rice ME (2004) DAncing past the DAT at a DA synapse. Trends in Neurosciences 27:270-277.

Croxson PL, Johansen-Berg H, Behrens TE, Robson MD, Pinsk MA, Gross CG, Richter W, Richter MC, Kastner S, Rushworth MF (2005) Quantitative investigation of connections of the prefrontal cortex in the human and macaque using probabilistic diffusion tractography. The Journal of Neuroscience 25:8854-8866.

Cui G, Jun SB, Jin X, Pham MD, Vogel SS, Lovinger DM, Costa RM (2013) Concurrent activation of striatal direct and indirect pathways during action initiation. Nature.

Curtis CE, Lee D (2010) Beyond working memory: the role of persistent activity in decision making. Trends in cognitive sciences 14:216-222.

D'Ardenne K, McClure SM, Nystrom LE, Cohen JD (2008) BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. Science 319:1264-1267.

Daw ND (2011) Trial-by-trial data analysis using computational models. Decision making, affect, and learning: Attention and performance XXIII 23:3-38.

Daw ND, Niv Y, Dayan P (2005) Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 8:1704-1711.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876-879.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ (2011) Model-based influences on humans' choices and striatal prediction errors. Neuron 69:1204-1215.

Dayan P, Niv Y (2008) Reinforcement learning: the good, the bad and the ugly. Current Opinion in Neurobiology 18:185-196.

de Wit S, Barker RA, Dickinson AD, Cools R (2011) Habitual versus goal-directed action control in Parkinson disease. J Cogn Neurosci 23:1218-1229.

de Wit S, Corlett PR, Aitken MR, Dickinson A, Fletcher PC (2009) Differential engagement of the ventromedial prefrontal cortex by goal-directed and habitual behavior toward food pictures in humans. J Neurosci 29:11330-11338.

de Wit S, Watson P, Harsay HA, Cohen MX, van de Vijver I, Ridderinkhof KR (2012a) Corticostriatal connectivity underlies individual differences in the balance between habitual and goal-directed action control. J Neurosci 32:12066-12075.

de Wit S, Standing HR, Devito EE, Robinson OJ, Ridderinkhof KR, Robbins TW, Sahakian BJ (2012b) Reliance on habits at the expense of goal-directed control following dopamine precursor depletion. Psychopharmacology (Berl) 219:621-631.

DeLong MR (1990) Primate models of movement disorders of basal ganglia origin. Trends Neurosci 13:281-285.

Der G, Deary IJ (2006) Age and sex differences in reaction time in adulthood: results from the United Kingdom Health and Lifestyle Survey. Psychology and aging 21:62.

Deserno L, Huys QJ, Boehme R, Buchert R, Heinze H-J, Grace AA, Dolan RJ, Heinz A, Schlagenhauf F (2015) Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proceedings of the National Academy of Sciences:201417219.

Destrieux C, Fischl B, Dale A, Halgren E (2010) Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. Neuroimage 53:1-15.

Di Martino A, Scheres A, Margulies DS, Kelly A, Uddin LQ, Shehzad Z, Biswal B, Walters JR, Castellanos FX, Milham MP (2008) Functional connectivity of human striatum: a resting state FMRI study. Cerebral Cortex 18:2735-2747.

Dichter GS, Damiano CA, Allen JA (2012) Reward circuitry dysfunction in psychiatric and neurodevelopmental disorders and genetic syndromes: animal models and clinical findings. J Neurodev Disord 4:19.

Dickinson A (1985) Actions and habits: the development of behavioural autonomy. Philosophical Transactions of the Royal Society of London B, Biological Sciences 308:67-78.

Dickinson A, Nicholas D, Adams CD (1983) The effect of the instrumental training contingency on susceptibility to reinforcer devaluation. The Quarterly Journal of Experimental Psychology 35:35-51.

Dolan RJ, Dayan P (2013) Goals and habits in the brain. Neuron 80:312-325.

Doll BB, Simon DA, Daw ND (2012) The ubiquity of model-based reinforcement learning. Current Opinion in Neurobiology 22:1075-1081.

Donders FC (1868) Over de snelheid van psychische processen. Onderzoekingen gedaan in het Physiologisch Laboratorium der Utrechtse Hoogeschool:39.

Doya K (1999) What are the computations of the cerebellum, the basal ganglia and the cerebral cortex? Neural Networks 12:961-974.

Doya K, Samejima K, Katagiri K-i, Kawato M (2002) Multiple model-based reinforcement learning. Neural computation 14:1347-1369.

Draganski B, Ashburner J, Hutton C, Kherif F, Frackowiak RSJ, Helms G, Weiskopf N (2011) Regional specificity of MRI contrast parameter changes in normal ageing revealed by voxel-based quantification (VBQ). NeuroImage 55:1423-1434.

Draganski B, Kherif F, Kloppel S, Cook PA, Alexander DC, Parker GJ, Deichmann R, Ashburner J, Frackowiak RS (2008) Evidence for segregated and integrative connectivity patterns in the human Basal Ganglia. J Neurosci 28:7143-7152.

Driscoll I, Hamilton DA, Yeo RA, Brooks WM, Sutherland RJ (2005) Virtual navigation in humans: the impact of age, sex, and hormones on place learning. Hormones and Behavior 47:326-335.

Dufau S, Dunabeitia JA, Moret-Tatay C, McGonigal A, Peeters D, Alario FX, Balota DA, Brysbaert M, Carreiras M, Ferrand L, Ktori M, Perea M, Rastle K, Sasburg O, Yap MJ, Ziegler JC, Grainger J (2011) Smart phone, smart science: how the use of smartphones can revolutionize research in cognitive science. PLoS One 6:e24974.

Duzel E, Bunzeck N, Guitart-Masip M, Wittmann B, Schott BH, Tobler PN (2009) Functional imaging of the human dopaminergic midbrain. Trends Neurosci 32:321-328.

Eagle DM, Baunez C, Hutcheson DM, Lehmann O, Shah AP, Robbins TW (2008) Stop-signal reaction-time task performance: role of prefrontal cortex and subthalamic nucleus. Cereb Cortex 18:178-188.

Ersche KD, Jones PS, Williams GB, Turton AJ, Robbins TW, Bullmore ET (2012) Abnormal brain structure implicated in stimulant drug addiction. Science 335:601-604.

Evenden JL (1999) Varieties of impulsivity. Psychopharmacology 146:348-361.

Everett G, Borcherding J (1970) L-DOPA: Effect of concentrations of dopamine, norepinephrine, and serotonin in brains of mice. Science 169:963-963.

Everitt BJ, Robbins TW (2005) Neural systems of reinforcement for drug addiction: from actions to habits to compulsion. Nat Neurosci 8:1481-1489.

Fabbrini G, Juncos J, Mouradian M, Serrati C, Chase T (1987) Levodopa pharmacokinetic mechanisms and motor fluctuations in Parkinson's disease. Annals of neurology 21:370-376.

Fallon JH, Moore RY (1978) Catecholamine innervation of the basal forebrain IV. Topography of the dopamine projection to the basal forebrain and neostriatum. Journal of Comparative Neurology 180:545-579.

Faul F, Erdfelder E, Lang AG, Buchner A (2007) G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. Behavior research methods 39:175-191.

Faul F, Erdfelder E, Buchner A, Lang AG (2009) Statistical power analyses using G*Power 3.1: tests for correlation and regression analyses. Behavior research methods 41:1149-1160.

Faure A, Haberland U, Condé F, El Massioui N (2005) Lesion to the nigrostriatal dopamine system disrupts stimulus-response habit formation. The Journal of Neuroscience 25:2771-2780.

Featherstone R, McDonald R (2004) Dorsal striatum and stimulus-response learning: lesions of the dorsolateral, but not dorsomedial, striatum impair acquisition of a stimulus-response-based instrumental discrimination task, while sparing conditioned place preference learning. Neuroscience 124:23-31.

Fecteau S, Knoch D, Fregni F, Sultani N, Boggio P, Pascual-Leone A (2007a) Diminishing risk-taking behavior by modulating activity in the prefrontal cortex: a direct current stimulation study. J Neurosci 27:12500-12505.

Fecteau S, Pascual-Leone A, Zald DH, Liguori P, Theoret H, Boggio PS, Fregni F (2007b) Activation of prefrontal cortex by transcranial direct current stimulation reduces appetite for risk during ambiguous decision making. J Neurosci 27:6212-6218.

Feredoes E, Heinen K, Weiskopf N, Ruff C, Driver J (2011) Causal evidence for frontal involvement in memory target maintenance by posterior brain areas during distracter interference of visual working memory. Proc Natl Acad Sci U S A 108:17510-17515.

Fischl B (2012) FreeSurfer. Neuroimage 62:774-781.

Fischl B, van der Kouwe A, Destrieux C, Halgren E, Ségonne F, Salat DH, Busa E, Seidman LJ, Goldstein J, Kennedy D (2004) Automatically parcellating the human cerebral cortex. Cerebral cortex 14:11-22.

Foerde K, Braun EK, Shohamy D (2012) A trade-off between feedback-based learning and episodic memory for feedback events: evidence from Parkinson's disease. Neuro-degenerative diseases 11:93-101.

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

Forstmann BU, Keuken MC, Jahfari S, Bazin PL, Neumann J, Schafer A, Anwander A, Turner R (2012) Cortico-subthalamic white matter tract strength predicts interindividual efficacy in stopping a motor response. Neuroimage 60:370-375.

Fox MD, Raichle ME (2007) Spontaneous fluctuations in brain activity observed with functional magnetic resonance imaging. Nature Reviews Neuroscience 8:700-711.

Frank MJ (2005) Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Nonmedicated Parkinsonism. Journal of Cognitive Neuroscience 17:51-72.

Frank MJ, O'Reilly RC (2006) A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behavioral Neuroscience 120:497-517.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: Cognitive reinforcement learning in parkinsonism. Science 306:1940-1943.

Frank MJ, Samanta J, Moustafa A, Sherman SJ (2007a) Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science (New York, NY) 318:1309-1312.

Frank MJ, Moustafa A, Haughey HM, Curran T, Hutchison KE (2007b) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proceedings of the National Academy of Sciences of the United States of America 104:16311-16316.

Fregni F, Boggio PS, Nitsche M, Bermpohl F, Antal A, Feredoes E, Marcolin MA, Rigonatti SP, Silva MTA, Paulus W (2005a) Anodal transcranial direct current stimulation of prefrontal cortex enhances working memory. Experimental Brain Research 166:23-30.

Fregni F, Boggio PS, Mansur CG, Wagner T, Ferreira MJ, Lima MC, Rigonatti SP, Marcolin MA, Freedman SD, Nitsche MA (2005b) Transcranial direct current stimulation of the unaffected hemisphere in stroke patients. Neuroreport 16:1551-1555.

Freitas C, Mondragón-Llorca H, Pascual-Leone A (2011) Noninvasive brain stimulation in Alzheimer's disease: systematic review and perspectives for the future. Experimental gerontology 46:611-627.

Freund T, Powell J, Smith A (1984) Tyrosine hydroxylase-immunoreactive boutons in synaptic contact with identified striatonigral neurons, with particular reference to dendritic spines. Neuroscience 13:1189-1215.

Fries P (2009) Neuronal gamma-band synchronization as a fundamental process in cortical computation. Annual Review of Neuroscience 32:209-224.

Friston K, Jezzard P, Turner R (1994) Analysis of functional MRI time-series. Human Brain Mapping 1:153-171.

Friston K, Kilner J, Harrison L (2006) A free energy principle for the brain. Journal of Physiology-Paris 100:70-87.

Friston KJ, Dolan RJ (2010) Computational and dynamic models in neuroimaging. Neuroimage 52:752-765.

Friston KJ, Fletcher P, Josephs O, Holmes A, Rugg M, Turner R (1998) Event-related fMRI: characterizing differential responses. NeuroImage 7:30-40.

Friston KJ, Holmes AP, Poline J, Grasby P, Williams S, Frackowiak RS, Turner R (1995) Analysis of fMRI time-series revisited. NeuroImage 2:45-53.

Fuster J (2008) The prefrontal cortex: Academic Press.

Gallagher M, McMahan RW, Schoenbaum G (1999) Orbitofrontal cortex and representation of incentive value in associative learning. J Neurosci 19:6610-6614.

Gazzaniga MS (2004) The cognitive neurosciences: MIT press.

Georgiou-Karistianis N, Sritharan A, Asadi H, Johnston L, Churchyard A, Egan G (2011) Diffusion tensor imaging in Huntington’s disease reveals distinct patterns of white matter degeneration associated with motor and cognitive deficits. Brain imaging and behavior 5:171-180.

Gerfen C, Bolam J (2010) The neuroanatomical organization of the basal ganglia. Handbook of basal ganglia structure and function 20:3-28.

Gerfen CR, Surmeier DJ (2011) Modulation of striatal projection systems by dopamine. Annu Rev Neurosci 34:441-466.

Gerfen CR, Engber TM, Mahan LC, Susel Z, Chase TN, Monsma FJ, Jr., Sibley DR (1990) D1 and D2 dopamine receptor-regulated gene expression of striatonigral and striatopallidal neurons. Science 250:1429-1432.

Gershman SJ, Markman AB, Otto AR (2014) Retrospective revaluation in sequential decision making: A tale of two systems. Journal of Experimental Psychology: General 143:182.

Gläscher J, Daw N, Dayan P, O'Doherty JP (2010) States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-Based and Model-Free Reinforcement Learning. Neuron 66:585-595.

Glaser D, Friston KJ (2004) Variance components. Human brain function:781-793.

Glasser MF, Sotiropoulos SN, Wilson JA, Coalson TS, Fischl B, Andersson JL, Xu J, Jbabdi S, Webster M, Polimeni JR (2013) The minimal preprocessing pipelines for the Human Connectome Project. NeuroImage 80:105-124.

Glimcher PW, Fehr E (2013) Neuroeconomics: Decision making and the brain: Academic Press.

Gollwitzer PM (1999) Implementation intentions: strong effects of simple plans. American psychologist 54:493.

Good CD, Johnsrude IS, Ashburner J, Henson RN, Fristen K, Frackowiak RS (2002) A voxel-based morphometric study of ageing in 465 normal adult human brains. In: Biomedical Imaging, 2002. 5th IEEE EMBS International Summer School on, p 16 pp.: IEEE.

Graybiel AM, Ragsdale CW (1978) Histochemically distinct compartments in the striatum of human, monkeys, and cat demonstrated by acetylthiocholinesterase staining. Proceedings of the National Academy of Sciences 75:5723-5726.

Greenhouse I, Oldenkamp CL, Aron AR (2012) Stopping a response has global or nonglobal effects on the motor system depending on preparation. Journal of Neurophysiology 107:384-392.

Griswold MA, Jakob PM, Heidemann RM, Nittka M, Jellus V, Wang J, Kiefer B, Haase A (2002) Generalized autocalibrating partially parallel acquisitions (GRAPPA). Magn Reson Med 47:1202-1210.

Guitart-Masip M, Duzel E, Dolan R, Dayan P (2014) Action versus valence in decision making. Trends in Cognitive Sciences 18:194-202.

Guitart-Masip M, Huys QJ, Fuentemilla L, Dayan P, Duzel E, Dolan RJ (2012) Go and no-go learning in reward and punishment: interactions between affect and effect. NeuroImage 62:154-166.

Guitart-Masip M, Fuentemilla L, Bach DR, Huys QJ, Dayan P, Dolan RJ, Duzel E (2011) Action dominates valence in anticipatory representations in the human striatum and dopaminergic midbrain. J Neurosci 31:7867-7875.

Gur RC, Mozley PD, Resnick SM, Gottlieb GL, Kohn M, Zimmerman R, Herman G, Atlas S, Grossman R, Berretta D (1991) Gender differences in age effect on brain atrophy measured by magnetic resonance imaging. Proceedings of the National Academy of Sciences 88:2845-2849.

Gurney K, Prescott TJ, Redgrave P (2001) A computational model of action selection in the basal ganglia. I. A new functional anatomy. Biol Cybern 84:401-410.

Haber S (2010) Integrative networks across basal ganglia circuits. Handbook of basal ganglia structure and function 20:409-427.

Haber SN (2003) The primate basal ganglia: parallel and integrative networks. Journal of chemical neuroanatomy 26:317-330.

Haber SN, Behrens TE (2014) The Neural Network Underlying Incentive-Based Learning: Implications for Interpreting Circuit Disruptions in Psychiatric Disorders. Neuron 83:1019-1039.

Haber SN, Fudge JL, McFarland NR (2000) Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 20:2369-2382.

Haber SN, Kim KS, Mailly P, Calzavara R (2006) Reward-related cortical inputs define a large striatal region in primates that interface with associative cortical connections, providing a substrate for incentive-based learning. J Neurosci 26:8368-8376.

Hall CN, Reynell C, Gesslein B, Hamilton NB, Mishra A, Sutherland BA, O’Farrell FM, Buchan AM, Lauritzen M, Attwell D (2014) Capillary pericytes regulate cerebral blood flow in health and disease. Nature 508:55-60.

Handwerker DA, Ollinger JM, D'Esposito M (2004) Variation of BOLD hemodynamic responses across subjects and brain regions and their effects on statistical analyses. NeuroImage 21:1639-1651.

Hartman DS, Civelli O (1997) Dopamine receptor diversity: molecular and pharmacological perspectives. In: Progress in Drug Research/Fortschritte der Arzneimittelforschung/Progrès des recherches pharmaceutiques, pp 173-194: Springer.

Haydon PG, Carmignoto G (2006) Astrocyte control of synaptic transmission and neurovascular coupling. Physiological reviews 86:1009-1031.

Hazy TE, Frank MJ, O'Reilly RC (2006) Banishing the homunculus: making working memory work. Neuroscience 139:105-118.

Helmich RC, Derikx LC, Bakker M, Scheeringa R, Bloem BR, Toni I (2010) Spatial remapping of cortico-striatal connectivity in Parkinson's disease. Cerebral Cortex 20:1175-1186.

Helms G, Dechent P (2009) Increased SNR and reduced distortions by averaging multiple gradient echo signals in 3D FLASH imaging of the human brain at 3T. Journal of Magnetic Resonance Imaging 29:198-204.

Helms G, Dathe H, Dechent P (2008a) Quantitative FLASH MRI at 3T using a rational approximation of the Ernst equation. Magnetic Resonance in Medicine 59:667-672.

Helms G, Dathe H, Kallenberg K, Dechent P (2008b) High-resolution maps of magnetization transfer with inherent correction for RF inhomogeneity and T1 relaxation obtained from 3D FLASH MRI. Magn Reson Med 60:1396-1407.

Helms G, Draganski B, Frackowiak R, Ashburner J, Weiskopf N (2009) Improved segmentation of deep brain grey matter structures using magnetization transfer (MT) parameter maps. NeuroImage 47:194-198.

Henrich J, Heine SJ, Norenzayan A (2010) The weirdest people in the world? Behavioral and brain sciences 33:61-83.

Herwig U, Satrapi P, Schonfeldt-Lecuona C (2003) Using the international 10-20 EEG system for positioning of transcranial magnetic stimulation. Brain topography 16:95-99.

Hikosaka O (2007) Basal ganglia mechanisms of reward-oriented eye movement. Ann N Y Acad Sci 1104:229-249.

Hitchcott PK, Quinn JJ, Taylor JR (2007) Bidirectional modulation of goal-directed actions by prefrontal cortical dopamine. Cereb Cortex 17:2820-2827.

Højsgaard S (2012) The doBy package. The Newsletter of the R Project Volume 6/2, May 2006 1:47.

Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nature Neuroscience 1:304-309.

Hong S, Hikosaka O (2011) Dopamine-mediated learning and switching in cortico-striatal circuit explain behavioral changes in reinforcement learning. Frontiers in behavioral neuroscience 5:15.

Hoogendam JM, Ramakers GM, Di Lazzaro V (2010) Physiology of repetitive transcranial magnetic stimulation of the human brain. Brain stimulation 3:95-118.

Horvath JC, Forte JD, Carter O (2015) Quantitative Review Finds No Evidence of Cognitive Effects in Healthy Populations from Single-Session Transcranial Direct Current Stimulation (tDCS). Brain stimulation.

Houk JC, Wise SP (1995) Distributed modular architectures linking basal ganglia, cerebellum, and cerebral cortex: their role in planning and controlling action. Cereb Cortex 5:95-110.

Houk JC, Adams JL, Barto AG (1995) A model of how the basal ganglia generate and use neural signals that predict reinforcement. Models of information processing in the basal ganglia:249-270.

Howard RA (1960) Dynamic programming and markov processes.

Huang YZ, Edwards MJ, Rounis E, Bhatia KP, Rothwell JC (2005) Theta burst stimulation of the human motor cortex. Neuron 45:201-206.

Huettel SA, Song AW, McCarthy G (2004) Functional magnetic resonance imaging: Sinauer Associates Sunderland, MA.

Hutton C, Bork A, Josephs O, Deichmann R, Ashburner J, Turner R (2002) Image distortion correction in fMRI: a quantitative evaluation. NeuroImage 16:217-240.

Hutton C, Josephs O, Stadler J, Featherstone E, Reid A, Speck O, Bernarding J, Weiskopf N (2011) The impact of physiological noise correction on fMRI at 7T. NeuroImage 57:101-112.

Huys QJ, Eshel N, O'Nions E, Sheridan L, Dayan P, Roiser JP (2012) Bonsai trees in your head: how the pavlovian system sculpts goal-directed choices by pruning decision trees. PLoS Comput Biol 8:e1002410.

Iadecola C, Nedergaard M (2007) Glial regulation of the cerebral microvasculature. Nature Neuroscience 10:1369-1376.

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

Isoda M, Hikosaka O (2008) Role for subthalamic nucleus neurons in switching from automatic to controlled eye movement. J Neurosci 28:7209-7218.

Isoda M, Hikosaka O (2011) Cortico-basal ganglia mechanisms for overcoming innate, habitual and motivational behaviors. Eur J Neurosci 33:2058-2069.

Jahfari S, Stinear CM, Claffey M, Verbruggen F, Aron AR (2010) Responding with restraint: what are the neurocognitive mechanisms? J Cogn Neurosci 22:1479-1492.

Jahfari S, Waldorp L, van den Wildenberg WP, Scholte HS, Ridderinkhof KR, Forstmann BU (2011) Effective connectivity reveals important roles for both the hyperdirect (fronto-subthalamic) and the indirect (fronto-striatal-pallidal) fronto-basal ganglia pathways during response inhibition. J Neurosci 31:6891-6899.

Jahfari S, Verbruggen F, Frank MJ, Waldorp LJ, Colzato L, Ridderinkhof KR, Forstmann BU (2012) How preparation changes the need for top-down control of the basal ganglia when inhibiting premature actions. J Neurosci 32:10870-10878.

Jbabdi S, Lehman JF, Haber SN, Behrens TE (2013) Human and monkey ventral prefrontal fibers use the same organizational principles to reach their targets: tracing versus tractography. The Journal of Neuroscience 33:3190-3201.

Jbabdi S, Sotiropoulos SN, Savio AM, Graña M, Behrens TE (2012) Model-based analysis of multishell diffusion MR data for tractography: How to get over fitting problems. Magnetic Resonance in Medicine 68:1846-1855.

Jessup RK, O'Doherty JP (2011) Human dorsal striatal activity during choice discriminates reinforcement learning behavior from the gambler's fallacy. The Journal of Neuroscience 31:6296-6304.

Jin X, Tecuapetla F, Costa RM (2014) Basal ganglia subcircuits distinctively encode the parsing and concatenation of action sequences. Nature Neuroscience.

Johansen-Berg H, Behrens TE, Sillery E, Ciccarelli O, Thompson AJ, Smith SM, Matthews PM (2005) Functional–anatomical validation and individual variation of diffusion tractography-based segmentation of the human thalamus. Cerebral Cortex 15:31-39.

Johansen-Berg H, Behrens T, Robson M, Drobnjak I, Rushworth M, Brady J, Smith S, Higham D, Matthews P (2004) Changes in connectivity profiles define functionally distinct regions in human medial frontal cortex. Proceedings of the National Academy of Sciences of the United States of America 101:13335-13340.

Johnson A, Redish AD (2005) Hippocampal replay contributes to within session learning in a temporal difference reinforcement learning model. Neural Networks 18:1163-1171.

Jones DK, Knösche TR, Turner R (2013) White matter integrity, fiber count, and other fallacies: the do's and don'ts of diffusion MRI. NeuroImage 73:239-254.

Kable JW, Glimcher PW (2007) The neural correlates of subjective value during intertemporal choice. Nature Neuroscience 10:1625-1633.

Kahneman D (2011) Thinking, fast and slow: Macmillan.

Kawaguchi Y, Wilson CJ, Emson PC (1990) Projection subtypes of rat neostriatal matrix cells revealed by intracellular injection of biocytin. The Journal of Neuroscience 10:3421-3438.

Keefe JO, Nadel L (1978) The hippocampus as a cognitive map: Clarendon Press Oxford.

Kemp JM, Powell T (1971) The connexions of the striatum and globus pallidus: synthesis and speculation. Philosophical Transactions of the Royal Society of London B, Biological Sciences 262:441-457.

Keramati M, Dezfouli A, Piray P (2011) Speed/accuracy trade-off between the habitual and the goal-directed processes. PLoS Comput Biol 7:e1002055.

Keuken M, Bazin P-L, Crown L, Hootsmans J, Laufer A, Müller-Axt C, Sier R, van der Putten E, Schäfer A, Turner R (2014) Quantifying inter-individual anatomical variability in the subcortex using 7T structural MRI. NeuroImage 94:40-46.

Killcross S, Coutureau E (2003) Coordination of actions and habits in the medial prefrontal cortex of rats. Cerebral Cortex 13:400-408.

Kim H, Sul JH, Huh N, Lee D, Jung MW (2009) Role of striatum in updating values of chosen actions. The Journal of Neuroscience 29:14701-14712.

Kim JS, Greene MJ, Zlateski A, Lee K, Richardson M, Turaga SC, Purcaro M, Balkam M, Robinson A, Behabadi BF, Campos M, Denk W, Seung HS (2014) Space-time wiring specificity supports direction selectivity in the retina. Nature 509:331-336.

Kincses TZ, Antal A, Nitsche MA, Bartfai O, Paulus W (2004) Facilitation of probabilistic classification learning by transcranial direct current stimulation of the prefrontal cortex in the human. Neuropsychologia 42:113-117.

Klein-Flugge MC, Hunt LT, Bach DR, Dolan RJ, Behrens TE (2011) Dissociable reward and timing signals in human midbrain and ventral striatum. Neuron 72:654-664.

Klein-Flugge MC, Barron HC, Brodersen KH, Dolan RJ, Behrens TE (2013) Segregated encoding of reward-identity and stimulus-reward associations in human orbitofrontal cortex. J Neurosci 33:3202-3211.

Kline RB, Association AP (2004) Beyond significance testing: Reforming data analysis methods in behavioral research.

Ko JH, Monchi O, Ptito A, Bloomfield P, Houle S, Strafella AP (2008) Theta burst stimulation-induced inhibition of dorsolateral prefrontal cortex reveals hemispheric asymmetry in striatal dopamine release during a set-shifting task: a TMS-[(11)C]raclopride PET study. Eur J Neurosci 28:2147-2155.

Ko YT, Miller J (2013) Signal-related contributions to stopping-interference effects in selective response inhibition. Exp Brain Res:205-212.

Koch K, Schachtzabel C, Wagner G, Schikora J, Schultz C, Reichenbach JR, Sauer H, Schlösser RG (2010) Altered activation in association with reward-related trial-and-error learning in patients with schizophrenia. NeuroImage 50:223-232.

Koechlin E, Ody C, Kouneiher F (2003) The architecture of cognitive control in the human prefrontal cortex. Science 302:1181-1185.

Kolbinger HM, Höflich G, Hufnagel A, Müller HJ, Kasper S (1995) Transcranial magnetic stimulation (TMS) in the treatment of major depression—a pilot study. Human Psychopharmacology: Clinical and Experimental 10:305-310.

Kötter R (1994) Postsynaptic integration of glutamatergic and dopaminergic signals in the striatum. Progress in Neurobiology 44:163-196.

Kravitz AV, Kreitzer AC (2011) Optogenetic manipulation of neural circuitry in vivo. Curr Opin Neurobiol 21:433-439.

Kravitz AV, Tye LD, Kreitzer AC (2012) Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat Neurosci 15:816-818.

Kravitz AV, Freeze BS, Parker PR, Kay K, Thwin MT, Deisseroth K, Kreitzer AC (2010) Regulation of parkinsonian motor behaviours by optogenetic control of basal ganglia circuitry. Nature 466:622-626.

Lambert C, Zrinzo L, Nagy Z, Lutti A, Hariz M, Foltynie T, Draganski B, Ashburner J, Frackowiak R (2012) Confirmation of functional zones within the human subthalamic nucleus: patterns of connectivity and sub-parcellation using diffusion weighted imaging. Neuroimage 60:83-94.

Lambert NM (1988) Adolescent outcomes for hyperactive children: Perspectives on general and specific patterns of childhood risk for adolescent educational, social, and mental health problems. American psychologist 43:786.

Landauer TK (1969) Reinforcement as consolidation. Psychological Review 76:82.

Lappin JS, Eriksen CW (1966) Use of a delayed signal to stop a visual reaction-time response. Journal of Experimental Psychology 72:805.

Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. Journal of the Experimental Analysis of Behavior.

Lau B, Glimcher PW (2007) Action and outcome encoding in the primate caudate nucleus. The Journal of Neuroscience 27:14502-14514.

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451-463.

Lau MA, Christensen BK, Hawley LL, Gemar MS, Segal ZV (2007) Inhibitory deficits for negative information in persons with major depressive disorder. Psychological medicine 37:1249-1259.

Lauterbur PC (1973) Image formation by induced local interactions: examples employing nuclear magnetic resonance. Nature 242:190-191.

Le Bihan D, Breton E (1985) Imagerie de diffusion in-vivo par résonance magnétique nucléaire. Comptes-Rendus de l'Académie des Sciences 93:27-34.

Le Bihan D, Breton E, Lallemand D, Grenier P, Cabanis E, Laval-Jeantet M (1986) MR imaging of intravoxel incoherent motions: application to diffusion and perfusion in neurologic disorders. Radiology 161:401-407.

Le Bihan D, Mangin JF, Poupon C, Clark CA, Pappata S, Molko N, Chabriat H (2001) Diffusion tensor imaging: concepts and applications. Journal of Magnetic Resonance Imaging 13:534-546.

Lee D, Seo H (2007) Mechanisms of reinforcement learning and decision making in the primate dorsolateral prefrontal cortex. Ann N Y Acad Sci 1104:108-122.

Lee SW, Shimojo S, O'Doherty JP (2014) Neural computations underlying arbitration between model-based and model-free learning. Neuron 81:687-699.

Leh SE, Ptito A, Chakravarty MM, Strafella AP (2007) Fronto-striatal connections in the human brain: a probabilistic diffusion tractography study. Neuroscience letters 419:113-118.

Lei W, Jiao Y, Del Mar N, Reiner A (2004) Evidence for differential cortical input to direct pathway versus indirect pathway striatal projection neurons in rats. The Journal of Neuroscience 24:8289-8299.

Lenglet C, Abosch A, Yacoub E, De Martino F, Sapiro G, Harel N (2012) Comprehensive in vivo mapping of the human basal ganglia and thalamic connectome in individuals using 7T MRI. PloS One 7:e29153.

Leotti LA, Wager TD (2010) Motivational influences on response inhibition measures. Journal of experimental psychology Human perception and performance 36:430-447.

Lewis SJ, Slabosz A, Robbins TW, Barker RA, Owen AM (2005) Dopaminergic basis for deficits in working memory but not attentional set-shifting in Parkinson's disease. Neuropsychologia 43:823-832.

Li CS, Huang C, Constable RT, Sinha R (2006) Imaging response inhibition in a stop-signal task: neural correlates independent of signal monitoring and post-response processing. J Neurosci 26:186-192.

Liljeholm M, O'Doherty JP (2012) Contributions of the striatum to learning, motivation, and performance: an associative account. Trends Cogn Sci 16:467-475.

Lindenberger U, Mayr U (2014) Cognitive aging: is there a dark side to environmental support? Trends Cogn Sci 18:7-15.

Lipszyc J, Schachar R (2010) Inhibitory control and psychopathology: a meta-analysis of studies using the stop signal task. Journal of the International Neuropsychological Society 16:1064-1076.

Livet J, Weissman TA, Kang H, Draft RW, Lu J, Bennis RA, Sanes JR, Lichtman JW (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. Nature 450:56-62.

Lloyd K, Davidson L, Hornykiewicz O (1975) The neurochemistry of Parkinson's disease: effect of L-dopa therapy. Journal of Pharmacology and Experimental Therapeutics 195:453-464.

Loe IM, Feldman HM (2007) Academic and educational outcomes of children with ADHD. Journal of Pediatric Psychology 32:643-654.

Logan GD (1994) On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In: Inhibitory processes in attention, memory, and language (Dagenbach D, Carr TH, eds), p 461. San Diego, CA, US: Academic Press.

Logan GD, Cowan WB, Davis KA (1984) On the ability to inhibit simple and choice reaction time responses: a model and a method. Journal of experimental psychology Human perception and performance 10:276-291.

Logan GD, Schachar RJ, Tannock R (1997) Impulsivity and inhibitory control. Psychological Science 8:60-64.

Logothetis NK (2008) What we can do and what we cannot do with fMRI. Nature 453:869-878.

Logothetis NK, Pauls J, Augath M, Trinath T, Oeltermann A (2001) Neurophysiological investigation of the basis of the fMRI signal. Nature 412:150-157.

Loopuijt LD, Van der Kooy D (1985) Organization of the striatum: collateralization of its efferent axons. Brain research 348:86-99.

Lund JS (1988) Anatomical organization of macaque monkey striate visual cortex. Annual Review of Neuroscience 11:253-288.

Lutti A, Thomas DL, Hutton C, Weiskopf N (2013) High-resolution functional MRI at 3 T: 3D/2D echo-planar imaging with optimized physiological noise correction. Magnetic Resonance in Medicine 69:1657-1664.

Lutti A, Stadler J, Josephs O, Windischberger C, Speck O, Bernarding J, Hutton C, Weiskopf N (2012) Robust and fast whole brain mapping of the RF transmit field B1 at 7T. PLoS One 7:e32379.

Lyness RC, Alvarez I, Sereno MI, MacSweeney M (2014) Microstructural differences in the thalamus and thalamic radiations in the congenitally deaf. Neuroimage 100:347-357.

Madden GJ, Bickel WK (2010) Impulsivity: The behavioral and neurological science of discounting: American Psychological Association.

Maia TV, Frank MJ (2011) From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 14:154-162.

Majid DA, Cai W, Corey-Bloom J, Aron AR (2013) Proactive Selective Response Suppression Is Implemented via the Basal Ganglia. The Journal of Neuroscience 33:13259-13269.

Majid DS, Cai W, George JS, Verbruggen F, Aron AR (2012) Transcranial magnetic stimulation reveals dissociable mechanisms for global versus selective corticomotor suppression underlying the stopping of action. Cereb Cortex 22:363-371.

Mansfield P (1977) Multi-planar image formation using NMR spin echoes. Journal of Physics C: Solid State Physics 10:L55.

Markov A (1971) Extension of the limit theorems of probability theory to a sum of variables connected in a chain.

Markov N, Ercsey-Ravasz M, Gomes AR, Lamy C, Magrou L, Vezoli J, Misery P, Falchier A, Quilodran R, Gariel M (2012) A weighted and directed interareal connectivity matrix for macaque cerebral cortex. Cerebral Cortex:bhs270.

Mars RB, Klein MC, Neubert FX, Olivier E, Buch ER, Boorman ED, Rushworth MF (2009) Short-latency influence of medial frontal cortex on primary motor cortex during action selection under conflict. J Neurosci 29:6926-6931.

Marshall L, Mölle M, Hallschmid M, Born J (2004) Transcranial direct current stimulation during sleep improves declarative memory. The Journal of Neuroscience 24:9985-9992.

Mathiesen C, Caesar K, Akgören N, Lauritzen M (1998) Modification of activity-dependent increases of cerebral blood flow by excitatory synaptic activity and spikes in rat cerebellar cortex. The Journal of physiology 512:555-566.

Maylor EA, Reimers S, Choi J, Collaer ML, Peters M, Silverman I (2007) Gender and sexual orientation differences in cognition across adulthood: Age is kinder to women than to men regardless of sexual orientation. Archives of sexual behavior 36:235-249.

Mazziotta J, Toga A, Evans A, Fox P, Lancaster J, Zilles K, Woods R, Paus T, Simpson G, Pike B (2001) A probabilistic atlas and reference system for the human brain: International Consortium for Brain Mapping (ICBM). Philosophical Transactions of the Royal Society of London Series B: Biological Sciences 356:1293-1322.

McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. The Journal of Neuroscience 31:2700-2705.

McNab F, Klingberg T (2008) Prefrontal cortex and basal ganglia control access to working memory. Nat Neurosci 11:103-107.

Mehta MA, Gumaste D, Montgomery AJ, McTavish SF, Grasby PM (2005) The effects of acute tyrosine and phenylalanine depletion on spatial working memory and planning in healthy volunteers are predicted by changes in striatal dopamine levels. Psychopharmacology 180:654-663.

Merton P, Morton H (1980) Stimulation of the cerebral cortex in the intact human subject.

Mesulam M-M (1978) Tetramethyl benzidine for horseradish peroxidase neurohistochemistry: a non-carcinogenic blue reaction product with superior sensitivity for visualizing neural afferents and efferents. Journal of Histochemistry & Cytochemistry 26:106-117.

Miech R, Breitner J, Zandi P, Khachaturian A, Anthony J, Mayer L (2002) Incidence of AD may decline in the early 90s for men, later for women The Cache County study. Neurology 58:209-218.

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

Mink JW (1996) The basal ganglia: focused selection and inhibition of competing motor programs. Prog Neurobiol 50:381-425.

Minsky M (1954) Neural nets and the brain-model problem. Unpublished doctoral dissertation, Princeton University, NJ.

Mirenowicz J, Schultz W (1994) Importance of unpredictability for reward responses in primate dopamine neurons. trials 16:25.

Mishkin M, Malamut B, Bachevalier J (1984) Memories and habits: Two neural systems. Neurobiology of learning and memory:65-77.

Moeller FG, Barratt ES, Dougherty DM, Schmitz JM, Swann AC (2014) Psychiatric aspects of impulsivity.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 16:1936-1947.

Monterosso JR, Aron AR, Cordova X, Xu J, London ED (2005) Deficits in response inhibition associated with chronic methamphetamine abuse. Drug and alcohol dependence 79:273-277.

Mori S, Crain BJ, Chacko V, Van Zijl P (1999) Three-dimensional tracking of axonal projections in the brain by magnetic resonance imaging. Annals of neurology 45:265-269.

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. Nature Neuroscience 9:1057-1063.

Mull BR, Seyal M (2001) Transcranial magnetic stimulation of left prefrontal cortex impairs working memory. Clinical Neurophysiology 112:1672-1675.

Mulquiney PG, Hoy KE, Daskalakis ZJ, Fitzgerald PB (2011) Improving working memory: exploring the effect of transcranial random noise stimulation and transcranial direct current stimulation on the dorsolateral prefrontal cortex. Clinical Neurophysiology 122:2384-2389.

Murphy DG, DeCarli C, McIntosh AR, Daly E, Mentis MJ, Pietrini P, Szczepanik J, Schapiro MB, Grady CL, Horwitz B (1996) Sex differences in human brain morphometry and metabolism: an in vivo quantitative magnetic resonance imaging and positron emission tomography study on the effect of aging. Archives of general psychiatry 53:585-594.

Murray G, Corlett P, Clark L, Pessiglione M, Blackwell A, Honey G, Jones P, Bullmore E, Robbins T, Fletcher P (2008) Substantia nigra/ventral tegmental reward prediction error disruption in psychosis. Molecular Psychiatry 13:267-276.

Nakamura K, Santos GS, Matsuzaki R, Nakahara H (2012) Differential reward coding in the subdivisions of the primate caudate during an oculomotor task. The Journal of Neuroscience 32:15963-15982.

Neubert F-X, Mars RB, Rushworth MF (2013) Is there an inferior frontal cortical network for cognitive control and inhibition? In: Principles of frontal lobe function (2nd edition) (Stuss D, Knight R, eds), pp 332-352. Oxford: Oxford University Press.

Neubert FX, Mars RB, Buch ER, Olivier E, Rushworth MF (2010) Cortical and subcortical interactions during action reprogramming and their related white matter pathways. Proc Natl Acad Sci U S A 107:13240-13245.

Nitsche M, Paulus W (2000) Excitability changes induced in the human motor cortex by weak transcranial direct current stimulation. The Journal of physiology 527:633-639.

Nitsche MA, Paulus W (2001) Sustained excitability elevations induced by transcranial DC motor cortex stimulation in humans. Neurology 57:1899-1901.

Nitsche MA, Schauenburg A, Lang N, Liebetanz D, Exner C, Paulus W, Tergau F (2003) Facilitation of implicit motor learning by weak transcranial direct current stimulation of the primary motor cortex in the human. J Cogn Neurosci 15:619-626.

Nitsche MA, Cohen LG, Wassermann EM, Priori A, Lang N, Antal A, Paulus W, Hummel F, Boggio PS, Fregni F (2008) Transcranial direct current stimulation: State of the art 2008. Brain stimulation 1:206-223.

Norman DA, Shallice T (1986) Attention to action: Springer.

O'Doherty J, Dayan P, Schultz J, Deichmann R, Friston K, Dolan RJ (2004) Dissociable roles of ventral and dorsal striatum in instrumental conditioning. Science 304:452-454.

O’Doherty JP (2014) The problem with value. Neuroscience & Biobehavioral Reviews 43:259-268.

Ogawa S, Lee T, Kay A, Tank D (1990) Brain magnetic resonance imaging with contrast dependent on blood oxygenation. Proceedings of the National Academy of Sciences 87:9868-9872.

Ogawa S, Tank DW, Menon R, Ellermann JM, Kim SG, Merkle H, Ugurbil K (1992) Intrinsic signal changes accompanying sensory stimulation: functional brain mapping with magnetic resonance imaging. Proceedings of the National Academy of Sciences 89:5951-5955.

Oldfield RC (1971) The assessment and analysis of handedness: the Edinburgh inventory. Neuropsychologia 9:97-113.

Osher DE, Saxe RR, Koldewyn K, Gabrieli JD, Kanwisher N, Saygin ZM (2015) Structural Connectivity Fingerprints Predict Cortical Selectivity for Multiple Visual Categories across Cortex. Cerebral Cortex:bhu303.

Ostlund SB, Balleine BW (2005) Lesions of medial prefrontal cortex disrupt the acquisition but not the expression of goal-directed learning. The Journal of Neuroscience 25:7763-7770.

Ott DVM, Ullsperger M, Jocham G, Neumann J, Klein TA (2011) Continuous theta-burst stimulation (cTBS) over the lateral prefrontal cortex alters reinforcement learning bias. NeuroImage 57:617-623.

Otto AR, Gershman SJ, Markman AB, Daw ND (2013) The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci 24:751-761.

Owen AM (2005) Cognitive planning in humans: New insights from the Tower of London (TOL) task. The cognitive psychology of planning:135-151.

Papanikolaou N, Karampekios S (2008) 3D MRI Acquisition: Technique. In: Image Processing in Radiology, pp 15-26: Springer.

Pascual-Leone A, Amedi A, Fregni F, Merabet LB (2005) The plastic human brain cortex. Annu Rev Neurosci 28:377-401.

Passingham RE, Stephan KE, Kötter R (2002) The anatomical basis of functional localization in the cortex. Nature Reviews Neuroscience 3:606-616.

Patenaude B, Smith SM, Kennedy DN, Jenkinson M (2011) A Bayesian model of shape and appearance for subcortical brain segmentation. NeuroImage 56:907-922.

Paton JJ, Louie K (2012) Reward and punishment illuminated. Nature Neuroscience 15:807.

Pavlov IP (1906) The scientific investigation of the psychical faculties or processes in the higher animals. Science 24:613-619.

Pavlov IP (2003) Conditioned reflexes: Courier Dover Publications.

Pavlov IP, Anrep GVe (1960) Conditioned Reflexes: An Investigation of the Physiological Activity of the Cerebral Cortex; Translated [from the Russian] and Edited by GV Anrep: Dover Publications.

Penfield W, Boldrey E (1937) Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. Brain: A Journal of Neurology.

Penny W (2012) Comparing dynamic causal models using AIC, BIC and free energy. NeuroImage 59:319-330.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006) Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442:1042-1045.

Purves D, Augustine GJ, Fitzpatrick D, Katz LC, LaMantia A-S, McNamara JO, Williams SM (2001) Modulation of movement by the basal ganglia.

Pykett IL, Newhouse JH, Buonanno FS, Brady TJ, Goldman MR, Kistler JP, Pohost GM (1982) Principles of nuclear magnetic resonance imaging. Radiology 143:157-168.

R Core Team (2011) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2012. Open access available at: http://cranr-projectorg.

R Development Core Team (2008) R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing:1-1731.

Raber J, Huang Y, Ashford JW (2004) ApoE genotype accounts for the vast majority of AD risk and AD pathology. Neurobiology of aging 25:641-650.

Rangel A, Hare T (2010) Neural computations associated with goal-directed choice. Current Opinion in Neurobiology 20:262-270.

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nature Reviews Neuroscience 9:545-556.

Rao RP, Ballard DH (1999) Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. Nature Neuroscience 2:79-87.

Redgrave P, Rodriguez M, Smith Y, Rodriguez-Oroz MC, Lehericy S, Bergman H, Agid Y, Delong MR, Obeso Ja (2010) Goal-directed and habitual control in the basal ganglia: implications for Parkinson's disease. Nature Reviews Neuroscience 11:760-772.

Redish AD (2004) Addiction as a computational process gone awry. Science 306:1944-1947.

Reiner A, Hart NM, Lei W, Deng Y (2010) Corticostriatal projection neurons–dichotomous types and dichotomous functions. Frontiers in neuroanatomy 4.

Reis J, Schambra HM, Cohen LG, Buch ER, Fritsch B, Zarahn E, Celnik PA, Krakauer JW (2009) Noninvasive cortical stimulation enhances motor skill acquisition over multiple days through an effect on consolidation. Proceedings of the National Academy of Sciences 106:1590-1595.

Rescorla RA (1988) Pavlovian conditioning: It's not what you think it is. American psychologist 43:151.

Rescorla RA, Wagner AR (1972) A theory of Pavlovian conditioning: Variations in the effectiveness of reinforcement and nonreinforcement. Classical conditioning II: Current research and theory 2:64-99.

Richfield EK, Penney JB, Young AB (1989) Anatomical and affinity state comparisons between dopamine D< sub> 1</sub> and D< sub> 2</sub> receptors in the rat central nervous system. Neuroscience 30:767-777.

Ridderinkhof KR, van den Wildenberg WP, Segalowitz SJ, Carter CS (2004) Neurocognitive mechanisms of cognitive control: the role of prefrontal cortex in action selection, response inhibition, performance monitoring, and reward-based learning. Brain Cogn 56:129-140.

Robbins TW, Everitt BJ (2007) A role for mesencephalic dopamine in activation: commentary on Berridge (2006). Psychopharmacology 191:433-437.

Robinson S, Basso G, Soldati N, Sailer U, Jovicich J, Bruzzone L, Kryspin-Exner I, Bauer H, Moser E (2009) A resting state network in the motor control circuit of the basal ganglia. BMC Neuroscience 10:137.

Romo R, Schultz W (1990) Dopamine neurons of the monkey midbrain: contingencies of responses to active touch during self-initiated arm movements. Journal of Neurophysiology 63:592-606.

Rorden C, Brett M (2000) Stereotaxic display of brain lesions. Behavioural neurology 12:191-200.

Rossi S, Hallett M, Rossini PM, Pascual-Leone A (2009) Safety, ethical considerations, and application guidelines for the use of transcranial magnetic stimulation in clinical practice and research. Clinical Neurophysiology 120:2008-2039.

Rucklidge JJ, Tannock R (2002) Neuropsychological profiles of adolescents with ADHD: Effects of reading difficulties and gender. Journal of child psychology and psychiatry 43:988-1003.

Rummery GA, Niranjan M (1994) On-Line Q-Learning Using Connectionist Systems. In: Technical Report CUED/F-INFENG/TR 166: Cambridge University Engineering Department.

Rushworth MF, Noonan MP, Boorman ED, Walton ME, Behrens TE (2011) Frontal cortex and reward-guided learning and decision-making. Neuron 70:1054-1069.

Rutledge RB, Skandali N, Dayan P, Dolan RJ (2014) A computational and neural model of momentary subjective well-being. Proceedings of the National Academy of Sciences 111:12252-12257.

Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW (2009) Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. The Journal of Neuroscience: The Official Journal of the Society for Neuroscience 29:15104-15114.

Salmon DP, Butters N (1995) Neurobiology of skill and habit learning. Current Opinion in Neurobiology 5:184-190.

Salthouse TA (2009) When does age-related cognitive decline begin? Neurobiology of aging 30:507-514.

Samejima K, Doya K (2007) Multiple representations of belief states and action values in corticobasal ganglia loops. Annals of the New York Academy of Sciences 1104:213-228.

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337-1340.

Saygin ZM, Osher DE, Koldewyn K, Reynolds G, Gabrieli JD, Saxe RR (2012) Anatomical connectivity patterns predict face selectivity in the fusiform gyrus. Nat Neurosci 15:321-327.

Schall JD, Godlove DC (2012) Current advances and pressing problems in studies of stopping. Curr Opin Neurobiol 22:1012-1021.

Schmidt R, Leventhal DK, Mallet N, Chen F, Berke JD (2013) Canceling actions involves a race between basal ganglia pathways. Nature Neuroscience 16:1118-1124.

Scholz VH, Flaherty A, Kraft E, Keltner J, Kwong K, Chen Y, Rosen B, Jenkins B (2000) Laterality, somatotopy and reproducibility of the basal ganglia and motor cortex during motor tasks. Brain research 879:204-215.

Schouten JF, Bekker JA (1967) Reaction time and accuracy. Acta psychologica 27:143-153.

Schroeder C, Mehta A, Givre S (1998) A spatiotemporal profile of visual system activation revealed by current source density analysis in the awake macaque. Cerebral Cortex 8:575-592.

Schultz W, Apicella P, Ljungberg T (1993) Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. The Journal of Neuroscience 13:900-913.

Schultz W, Dayan P, Montague PR (1997) A Neural Substrate of Prediction and Reward. Science 275:1593-1599.

Schulz KF, Grimes DA (2002) Blinding in randomised trials: hiding who got what. The Lancet 359:696-700.

Schwabe L, Wolf OT (2009) Stress prompts habit behavior in humans. The Journal of Neuroscience 29:7191-7198.

Schwabe L, Wolf OT (2011) Stress-induced modulation of instrumental behavior: from goal-directed to habitual control of action. Behavioural Brain Research 219:321-328.

Sebold M, Deserno L, Nebe S, Schad DJ, Garbusow M, Hägele C, Keller J, Jünger E, Kathmann N, Smolka M (2014) Model-based and model-free decisions in alcohol dependence. Neuropsychobiology 70:122-131.

Simon DA, Daw ND (2011) Environmental statistics and the trade-off between model-based and TD learning in humans. In: Advances in neural information processing systems, pp 127-135.

Sjoerds Z, Van Den Brink W, Beekman A, Penninx B, Veltman D (2014) Response inhibition in alcohol-dependent patients and patients with depression/anxiety: a functional magnetic resonance imaging study. Psychological medicine 44:1713-1725.

Skinner BF (1938) The behavior of organisms: An experimental analysis.

Sloan H, Austin V, Blamire A, Schnupp JW, Lowe AS, Allers K, Matthews PM, Sibson NR (2010) Regional differences in neurovascular coupling in rat brain as determined by fMRI and electrophysiology. NeuroImage 53:399-411.

Smith SM (2002) Fast robust automated brain extraction. Human brain mapping 17:143-155.

Smith SM, Jenkinson M, Woolrich MW, Beckmann CF, Behrens TE, Johansen-Berg H, Bannister PR, De Luca M, Drobnjak I, Flitney DE (2004) Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23:S208-S219.

Smittenaar P, Guitart-Masip M, Lutti A, Dolan RJ (2013a) Preparing for selective inhibition within frontostriatal loops. J Neurosci 33:18087-18097.

Smittenaar P, FitzGerald TH, Romei V, Wright ND, Dolan RJ (2013b) Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron 80:914-919.

Somogyi P, Bolam J, Smith A (1981) Monosynaptic cortical input and local axon collaterals of identified striatonigral neurons. A light and electron microscopic study using the golgi-peroxidase transport-degeneration procedure. Journal of Comparative Neurology 195:567-584.

Sotiropoulos SN, Aganj I, Jbabdi S, Sapiro G, Lenglet C, Behrens TE (2011) Inference on constant solid angle orientation distribution functions from diffusion-weighted mri. Quebec, Canada, June.

Sotiropoulos SN, Jbabdi S, Xu J, Andersson JL, Moeller S, Auerbach EJ, Glasser MF, Hernandez M, Sapiro G, Jenkinson M (2013) Advances in diffusion MRI acquisition and processing in the Human Connectome Project. NeuroImage 80:125-143.

Stagg CJ, Nitsche MA (2011) Physiological basis of transcranial direct current stimulation. The Neuroscientist 17:37-53.

Stagg CJ, Lin RL, Mezue M, Segerdahl A, Kong Y, Xie J, Tracey I (2013) Widespread Modulation of Cerebral Perfusion Induced during and after Transcranial Direct Current Stimulation Applied to the Left Dorsolateral Prefrontal Cortex. J Neurosci 33:11425-11431.

Stalnaker TA, Calhoon GG, Ogawa M, Roesch MR, Schoenbaum G (2012) Reward prediction error signaling in posterior dorsomedial striatum is action specific. The Journal of Neuroscience 32:10296-10305.

Steiner H, Tseng KY (2010) Handbook of basal ganglia structure and function: a decade of progress: Academic Press.

Stephens PA, Buskirk SW, Hayward GD, Martinez Del Rio C (2005) Information theory and hypothesis testing: a call for pluralism. Journal of Applied Ecology 42:4-12.

Surmeier DJ, Reiner A, Levine MS, Ariano MA (1993) Are neostriatal dopamine receptors co-localized? Trends in Neurosciences 16:299-305.

Sutton RS (1990) Integrated architectures for learning, planning, and reacting based on approximating dynamic programming. In: Proceedings of the seventh international conference on machine learning, pp 216-224.

Sutton RS, Barto AG (1981) Toward a modern theory of adaptive networks: expectation and prediction. Psychological Review 88:135.

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, Mass.: MIT Press.

Swann N, Poizner H, Houser M, Gould S, Greenhouse I, Cai W, Strunk J, George J, Aron AR (2011) Deep brain stimulation of the subthalamic nucleus alters the cortical profile of response inhibition in the beta frequency band: a scalp EEG study in Parkinson's disease. The Journal of Neuroscience 31:5721-5729.

Tanji J, Shima K, Mushiake H (2007) Concept-based behavioral planning and the lateral prefrontal cortex. Trends in cognitive sciences 11:528-534.

Tekin S, Cummings JL (2002) Frontal–subcortical neuronal circuits and clinical neuropsychiatry: an update. Journal of psychosomatic research 53:647-654.

Thompson SP (1910) A physiological effect of an alternating magnetic field. Proceedings of the Royal Society of London Series B, Containing Papers of a Biological Character:396-398.

Thorndike EL (1898) Animal intelligence: An experimental study of the associative processes in animals. Psychological Monographs: General and Applied 2:i-109.

Thorndike EL (1911) Animal intelligence: Experimental studies: Macmillan.

Tolman EC (1932) Purposive behavior in animals and men: Univ of California Press.

Tolman EC (1948) Cognitive maps in rats and men. Psychological review 55:189.

Tolman EC, Honzik CH (1930) Introduction and removal of reward, and maze performance in rats. University of California Publications in Psychology.

Tran-Tu-Yen DA, Marchand AR, Pape JR, Di Scala G, Coutureau E (2009) Transient role of the rat prelimbic cortex in goal-directed behaviour. Eur J Neurosci 30:464-471.

Tricomi E, Balleine BW, O'Doherty JP (2009) A specific role for posterior dorsolateral striatum in human habit learning. European Journal of Neuroscience 29:2225-2232.

Tricomi EM, Delgado MR, Fiez JA (2004) Modulation of caudate activity by action contingency. Neuron 41:281-292.

Tsai HC, Zhang F, Adamantidis A, Stuber GD, Bonci A, de Lecea L, Deisseroth K (2009) Phasic firing in dopaminergic neurons is sufficient for behavioral conditioning. Science 324:1080-1084.

Tziortzi AC, Haber SN, Searle GE, Tsoumpas C, Long CJ, Shotbolt P, Douaud G, Jbabdi S, Behrens TE, Rabiner EA (2014) Connectivity-based functional analysis of dopamine release in the striatum using diffusion-weighted MRI and positron emission tomography. Cerebral Cortex 24:1165-1177.

Tzourio-Mazoyer N, Landeau B, Papathanassiou D, Crivello F, Etard O, Delcroix N, Mazoyer B, Joliot M (2002) Automated anatomical labeling of activations in SPM using a macroscopic anatomical parcellation of the MNI MRI single-subject brain. NeuroImage 15:273-289.

Uylings HB, Groenewegen HJ, Kolb B (2003) Do rats have a prefrontal cortex? Behav Brain Res 146:3-17.

Valentin VV, Dickinson A, O'Doherty JP (2007) Determining the neural substrates of goal-directed learning in the human brain. J Neurosci 27:4019-4026.

van den Bos W, Cohen MX, Kahnt T, Crone EA (2012) Striatum–medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. Cerebral Cortex 22:1247-1255.

van den Bos W, Rodriguez CA, Schweitzer JB, McClure SM (2014) Connectivity Strength of Dissociable Striatal Tracts Predict Individual Differences in Temporal Discounting. The Journal of Neuroscience 34:10298-10310.

Van Essen DC, Smith SM, Barch DM, Behrens TE, Yacoub E, Ugurbil K (2013) The WU-Minn human connectome project: an overview. NeuroImage 80:62-79.

Van Essen DC, Ugurbil K, Auerbach E, Barch D, Behrens T, Bucholz R, Chang A, Chen L, Corbetta M, Curtiss SW (2012) The Human Connectome Project: a data acquisition perspective. NeuroImage 62:2222-2231.

van Veen V, Krug MK, Carter CS (2008) The neural and computational basis of controlled speed-accuracy tradeoff during task performance. J Cogn Neurosci 20:1952-1965.

Verbruggen F, Logan GD (2008) Response inhibition in the stop-signal paradigm. Trends Cogn Sci 12:418-424.

Verbruggen F, Logan GD (2009a) Proactive adjustments of response strategies in the stop-signal paradigm. Journal of experimental psychology Human perception and performance 35:835-854.

Verbruggen F, Logan GD (2009b) Models of response inhibition in the stop-signal and stop-change paradigms. Neurosci Biobehav Rev 33:647-661.

Verstynen TD, Badre D, Jarbo K, Schneider W (2012) Microstructural organizational patterns in the human corticostriatal system. Journal of neurophysiology 107:2984-2995.

Vink M, Kahn RS, Raemaekers M, van den Heuvel M, Boersma M, Ramsey NF (2005) Function of striatum beyond inhibition and execution of motor responses. Human Brain Mapping 25:336-344.

Voon V, Pessiglione M, Brezing C, Gallea C, Fernandez H, Dolan RJ, Hallett M (2010) Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. Neuron 65:135-142.

Voon V, Derbyshire K, Rück C, Irvine M, Worbe Y, Enander J, Schreiber L, Gillan C, Fineberg N, Sahakian B (2014) Disorders of compulsivity: a common bias towards learning habits. Molecular Psychiatry.

Wagner A, Rescorla R (1972) Inhibition in Pavlovian conditioning: Application of a theory. In: Inhibition and learning.

Wagner T, Fregni F, Fecteau S, Grodzinsky A, Zahn M, Pascual-Leone A (2007) Transcranial direct current stimulation: a computer-based human model study. NeuroImage 35:1113-1124.

Wall NR, De La Parra M, Callaway EM, Kreitzer AC (2013) Differential innervation of direct-and indirect-pathway striatal projection neurons. Neuron 79:347-360.

Wassum K, Cely I, Maidment N, Balleine B (2009) Disruption of endogenous opioid activity during instrumental learning enhances habit acquisition. Neuroscience 163:770-780.

Watkins CJCH (1989) Learning from delayed rewards. In: University of Cambridge.

Watkins CJCH, Dayan P (1992) Q-learning. Machine Learning 8:279-292.

Wedeen VJ, Wang R, Schmahmann JD, Benner T, Tseng W, Dai G, Pandya D, Hagmann P, D'Arceuil H, de Crespigny AJ (2008) Diffusion spectrum magnetic resonance imaging (DSI) tractography of crossing fibers. NeuroImage 41:1267-1277.

Weiskopf N, Helms G (2008) Multi-parameter mapping of the human brain at 1mm resolution in less than 20 minutes. In: Proc. Intl. Soc. Magn. Reson. Med, p 2241.

Weiskopf N, Hutton C, Josephs O, Deichmann R (2006) Optimal EPI parameters for reduction of susceptibility-induced BOLD sensitivity losses: a whole-brain analysis at 3 T and 1.5 T. NeuroImage 33:493-504.

Weiskopf N, Suckling J, Williams G, Correia MM, Inkster B, Tait R, Ooi C, Bullmore ET, Lutti A (2013) Quantitative multi-parameter mapping of R1, PD*, MT, and R2* at 3T: a multi-center validation. Frontiers in neuroscience 7.

Whiteside SP, Lynam DR (2001) The five factor model and impulsivity: Using a structural model of personality to understand impulsivity. Personality and individual differences 30:669-689.

Wickens JR, Horvitz JC, Costa RM, Killcross S (2007) Dopaminergic mechanisms in actions and habits. The Journal of Neuroscience 27:8181-8183.

Williams BR, Ponesse JS, Schachar RJ, Logan GD, Tannock R (1999) Development of inhibitory control across the life span. Dev Psychol 35:205-213.

Willis SL, Schaie KW (1988) Gender differences in spatial ability in old age: Longitudinal and intervention findings. Sex Roles 18:189-203.

Wilson RC, Takahashi YK, Schoenbaum G, Niv Y (2014) Orbitofrontal cortex as a cognitive map of task space. Neuron 81:267-279.

Wise RA (2004) Dopamine, learning and motivation. Nat Rev Neurosci 5:483-494.

Wise RA, Bozarth MA (1987) A psychomotor stimulant theory of addiction. Psychological Review 94:469.

Witten IH (1977) An adaptive optimal controller for discrete-time Markov environments. Information and control 34:286-295.

Worsley KJ, Friston KJ (1995) Analysis of fMRI time-series revisited—again. NeuroImage 2:173-181.

Wunderlich K, Smittenaar P, Dolan RJ (2012a) Dopamine enhances model-based over model-free choice behavior. Neuron 75:418-424.

Wunderlich K, Dayan P, Dolan RJ (2012b) Mapping value based planning and extensively trained choice in the human brain. Nat Neurosci 15:786-791.

Xue G, Juan CH, Chang CF, Lu ZL, Dong Q (2012) Lateral prefrontal cortex contributes to maladaptive decisions. Proc Natl Acad Sci U S A 109:4401-4406.

Yin HH, Knowlton BJ (2006) The role of the basal ganglia in habit formation. Nature Reviews Neuroscience 7:464-476.

Yin HH, Knowlton BJ, Balleine BW (2004) Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning. Eur J Neurosci 19:181-189.

Yin HH, Ostlund SB, Knowlton BJ, Balleine BW (2005) The role of the dorsomedial striatum in instrumental conditioning. Eur J Neurosci 22:513-523.

Yu AJ, Dayan P (2005) Uncertainty, neuromodulation, and attention. Neuron 46:681-692.

Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31:1116-1128.

Zaghloul KA, Blanco JA, Weidemann CT, McGill K, Jaggi JL, Baltuch GH, Kahana MJ (2009) Human substantia nigra neurons encode unexpected financial rewards. Science 323:1496-1496.

Zandbelt BB, Vink M (2010) On the role of the striatum in response inhibition. PLoS One 5:e13848.

Zandbelt BB, Bloemendaal M, Neggers SF, Kahn RS, Vink M (2012) Expectations and violations: Delineating the neural network of proactive inhibitory control. Hum Brain Mapp:Epub ahead of print.

Zeki S, Watson J, Lueck C, Friston KJ, Kennard C, Frackowiak R (1991) A direct demonstration of functional specialization in human visual cortex. The Journal of Neuroscience 11:641-649.