UNIVERSITEIT GENT

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Hebbian learning and cognitive control: Modeling and empirical studies

Anand Ramamoorthy

Promotor: Prof. Dr. Tom Verguts

Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Psychologie

2013

UNIVERSITEIT GENT

FACULTEIT PSYCHOLOGIE EN PEDAGOGISCHE WETENSCHAPPEN

Hebbian learning and cognitive control: Modeling and empirical studies

Anand Ramamoorthy

Promotor: Prof. Dr. Tom Verguts

Proefschrift ingediend tot het behalen van de academische graad van Doctor in de Psychologie

2013

CONTENTS

ACKNOWLEDGEMENTS

I sincerely thank my doctoral supervisor Prof.Dr.Tom Verguts, members of the doctoral guidance committee, examination committee, friends, family and colleagues for all the support during my doctoral studies. I could not have done it without you all.

CHAPTER 1

INTRODUCTION

Learning and cognitive control are integral to adaptive behaviour in human beings. Learning allows us to constantly explore and understand our environment ("what is out there") as well as to refine our goals ("what to do"). Cognitive control allows us to achieve said goals. Here, we briefly discuss learning and cognitive control prior to outlining the program of research described in this thesis.

1.1 Learning

Learning could be defined as the acquisition of information that was not available to the learner prior to the act. Consider, for instance, a child learning how to write or how to produce a certain pattern of sounds that might eventually be construed as music. Although a distinction can be made between the acquisition of knowledge (e.g.: the structure and meaning of a word) and the skills associated with having said knowledge (e.g.: the motor skills needed to produce the word), in both instances, the brain acquires *information* that then gets applied in different ways.

Learning is a process that can take many forms. Learning can be through trial-and-error (Thorndike, 1911); for example, a child learning to walk for the first time. This is a rich source of learning, but as its name suggests, it is prone to errors and attendant costs. Some such costs could be dangerous for the learner. For example, trying out a wild berry without prior knowledge could lead to allergic responses and even anaphylactic shock. Learning by observation (Bandura, 1977) is a way to circumvent these costs. Humans can and do learn by observing others and this has been studied extensively in psychology and the social neurosciences (Bandura, 1989; Cross, Kraemer, Hamilton, Kelley & Grafton, 2009). A third and highly influential form of learning in humans is instruction-following, or, learning-by-being-told.

Humans learn in all these ways and sometimes combine different ways of learning to accelerate knowledge-acquisition. For instance, learning a form in the martial arts often involves observing the instructor, receiving verbal guidance, practicing the moves and eventually calibrating them according to feedback.

The phenomenon of learning has been studied extensively since Pavlov (1927) famously demonstrated classical conditioning in animals. In classical conditioning, a neutral stimulus (e.g. a tone) is paired with an unconditioned stimulus (e.g. air puff) that elicits an unconditioned response (e.g. eye-blink). Associative learning of the neutral stimulus leads to a conditioned response; in the given example eye-blinks are produced in response to the tone even in the absence of an air-puff. This demonstrates the power of simple associative learning in the brain. While this may seem far removed from the types of learning discussed above, a key idea explored in this thesis is that different forms of learning are fundamentally associative in nature. Associative learning refers to the linking of two occurrences or items by virtue of their being related in time. It reflects the brain's capacity to discern and discover correlations in the environment. An elegant and useful form of associative learning in neuronal or neural networks is Hebbian learning.

1.1.1 Hebbian learning

Donald Hebb observed in The Organisation of Behaviour (Hebb, 1949), that "The general idea is an old one, that any two cells or systems of cells that are repeatedly active at the same time will tend to become 'associated', so that activity in one facilitates activity in the other." The idea has come a long way from that pithy observation and has found much application in computational accounts of learning in the brain. Hebbian learning, in its various incarnations, finds application in computational neuroscience (Kempter, Gerstner & van Hemmen, 1999), modelling of cognition (Verguts & Notebaert, 2008) and artificial intelligence (Hinton, 1989).

Formally, Hebbian learning is expressed as a variant of the following general equation:

$$\Delta w_{ij} = \lambda x_i y_j$$

where Δw_{ij} is the change in synaptic connection strength (weight) between the i^{th} presynaptic neuron and j^{th} postsynaptic neuron, λ is the learning rate, x_i is the activation of the presynaptic neuron and y_j is the activation of the postsynaptic neuron.

Hebbian Learning is considered a biologically plausible learning mechanism, (Antonov, Antonova, Kandel & Hawkins, 2003; Kandel, Abrams, Bernier, Carew, Hawkins & Schwartz, 1983; Kelso, Ganong & Brown, 1986) and is understood to involve experience-based changes at synapses between neurons, with these changes encoding a change in connection strength. Hebbian learning is arguably the most realistic of all available learning rules in the study of neural networks as applied to cognition (see O'Reilly, 1998; 2001). In addition to simple associative processes, synaptic plasticity is also modulated by other factors, such as temporal aspects of spiking behaviour (Abbot & Nelson, 2000). One such factor, spike-timing-dependent-plasticity, allows for the instantiation of a competitive Hebbian process (Song, Miller & Abbot, 2000). Reward-modulated Hebbian learning rules implement temporal-difference learning (a type of reinforcement learning) (Rao & Sejnowski, 2001). More complex learning rules contain a Hebbian core; examples include predictive coding (Rao & Ballard, 1999) and the influential Bienenstock-Cooper-Munro rule (Bienenstock, Cooper & Munro, 1982; Cooper & Bear, 2012).

1.2 Cognitive control

Consider the example of shopping at a large supermarket. Shelves get rearranged with goods moving from one location to another, periodically. Shoppers learn the locations of their preferred brands/items only to have them changed within a few weeks. The first time someone encounters such a change, she/he is momentarily confused as the goal of reaching a certain object cannot be accomplished. Then she/he looks around, learns the new location and picks up the desired item. On subsequent trips to the supermarket, the shopper now has to override a previously acquired location map to navigate to the items they need correctly. *Cognitive control* refers to the ability to overcome such a prepotent response (for example, "go to the top shelf on Aisle 1") in favour of the appropriate response (for example, "go to the middle shelf

on Aisle 4"). This ability is a general one that applies across behaviours and circumstances and contributes to adaptive behaviour.

Recently, cognitive control has enjoyed much attention from psychologists and computational modelers alike, on account of its importance to human adaptive behaviour. Influential accounts of cognitive control include the conflict-monitoring model of Botvinick, Braver, Barch, Carter and Cohen (2001), the adaptation-by-binding model of Verguts and Notebaert (2008) and the prediction-of-response-outcome model of Alexander & Brown (2010).

Despite the generality of the very concept of cognitive control, previous models (e.g., Botvinick et al., 2001; Verguts & Notebaert, 2008), have tended to restrict their scope to specific circumstances (e.g., Stroop or Simon tasks). Here, we begin from simple, powerful computational concepts and develop models of learning and control that account for hitherto neglected aspects of cognitive control. We focus on two such instances below; instruction following and self-control in decision making.

1.2.1 Instruction following

Once described as an unsolved mystery by Monsell (1996), the human ability to immediately and accurately implement instructions and rules remains an open question in spite of recent empirical investigations into its neural correlates (Cole, Laurent & Stocco, 2013; Hartstra et al. 2011; Ruge & Wolfensteller, 2010). Instructions are verbal or symbolic statements that associate stimuli with responses or require behaviours contingent upon specific stimuli/conditions (e.g. "if you hear a beep, please press the right arrow key"). From parents teaching rules to their children, to manuals for repairing advanced machinery, instructions pervade human life thoroughly. The ability to hold, interpret and implement an instruction has seen few theoretical attempts to understand it. Virtually every experiment in psychology relies on this phenomenon. Every model of human adaptive behaviour and cognitive control assumes the existence of a system that utilizes task-demand information (context, rules, instructions etc.), yet few models explore this system *per se*.

1.2.2 Self-control

Second, we are interested in self-control, which is a form of cognitive control that is known to have significant implications for long-term success and well-being in the life of an individual (Casey et al. 2011). Self-control can manifest in diverse situations, such as overcoming a fear-response, overriding the impulse to act when the prudent course of action would be to wait and when temptations are overcome in favour of goals with more lasting benefits. When presented with a choice between two alternatives, differing in terms of reward-value over time (immediate, small, short term gains versus larger long-term gains requiring delays), how does the human brain resist the temptation of the more immediate reward in order to obtain a better one that would become available at a later point in time? And can this be conceptualized as just another instance of cognitive control?

1.3 The modeling framework

In this thesis, we present a unified computational framework applied to the phenomena described earlier; instruction following and self-control. In particular, they are treated as two instances of cognitive control. This framework conceptualizes complementary processing systems /pathways as instances along a computational tradeoff.

Complementary processing systems have been hypothesised since the advent of cognitive psychology. Schneider and Schiffrin (1977) posited the existence of automatic and controlled processing in the brain. Automatic processes are fast, unconscious and inflexible by virtue of their automaticity, whereas controlled processing is more deliberate, modulated by attention and more flexible (Kahneman, 2011). Dual-systems have been proposed to underlie several processes, including memory (Reyna, 2012), reasoning (Sloman, 1996) and learning (Ashby and Maddox, 2005; Ashby & Crossley, 2010).

Computational tradeoffs are a simple yet powerful way of understanding many aspects of adaptive behaviour in organisms. In the case of human behaviour, a very well known example is the speed-accuracy tradeoff. Speed (measured in terms of

response times) often comes at the expense of accuracy (measured in terms of errors) and vice versa, and this has been known to reflect the underlying information processing dynamics (Wickelgren, 1977) in the human brain. At its core, this tradeoff reduces to a simple consideration – "is there sufficient information to act on?" If the individual waits, then they acquire more information, in all likelihood increasing the accuracy of the response, while taking longer to respond. Responding immediately or relatively quickly might lead to poor decisions due to lack of information. Thus a computational tradeoff is born – where two variables (e.g. speed, accuracy) are complementary to each other.

In dual-system models, one system can be thought of as instantiating one aspect of the pertinent tradeoff, with the other representing its anti-correlated twin. A very interesting application of this idea to learning and behavioural control is the modelling framework of Daw, Niv and Dayan (2005). In this framework two systems work together as well as competitively for behavioural control. One system computes responses by performing a *tree-search* through possible response alternatives, whereas the other exploits *cached* responses. Caching can be thought of as using a previously learnt shortcut to reach a destination quickly, whereas tree-search would correspond to a more laborious estimation of the optimality of different routes. These two systems are not purely competitive, because the cached responses could also come from responses selected by the tree-search system, upon repeated application and reinforcement. This tradeoff between deliberative action and rapid action underlies many seemingly opposite aspects of human behaviour, such as decision making (Trimmer et al., 2008), ethics (utilitarian judgments versus deontological ones; Fox, 2013) and learning. It can also be cast in the light of a learning-action continuum (cf. also Boureau & Dayan, 2011). The tree-search system learns fast but takes longer to respond, whereas the cached system responds quickly and learns slowly. This is the computational tradeoff we explore in our general computational framework. We derive two models from it; a model of instruction-following and a model of self-control in decision making. Our models take into account known neural substrates of the phenomena of interest. We assume a general competitive Hebbian learning process to capture learning in each system. After this description of the general framework, we now provide a broad overview of the purpose of the different chapters of the thesis.

1.4 Computational model of instruction following

Recent investigations into the nature of instruction following have yielded interesting insights into its neural correlates (Brass et al., 2009; Cole et al., 2010; Hartstra et al., 2011; Ruge & Wolfensteller, 2010). A common theme has been the centrality of the lateral prefrontal cortex (LPFC) to the acquisition and implementation of instructions and rules.

On the theoretical side of things, Noelle and Cottrell (1996) proposed a computational model of instruction following which was arguably the first model to tackle this open question. The neuro-computational basis of instruction following has received more attention only recently (Biele, Rieskamp, Krugel & Heekeren, 2011; Doll et al., 2009).

Following Doll et al. (2009), we developed a computational model of instruction following that accounts for behavioural findings while employing a neuroanatomically informed architecture (Ramamoorthy & Verguts, 2012). This model consisted of two learning systems; a lateral prefrontal cortical (LPFC) system and a striatal system. The former acquired and implemented instructions using fast Hebbian learning, whilst the latter learnt from contingencies, also through Hebbian learning. The LPFC system was capable of learning fast but responded slowly, whereas the striatal system learnt relatively slowly but responded more rapidly. The simulation studies exploring the usefulness of this model are reported in Chapter 2.

1.5 Computational model of self-control

Given its importance to everyday life as well as long-term well-being, self-control has been studied extensively from multiple vantage points (Baumeister, 1998; Casey et al., 2011; de Ridder et al., 2012, Moffitt et al., 2011; Muraven & Baumeister, 2000; Tangney, Baumeister, & Boone, 2004). As far as the neurobiology of self-control is concerned, it is tightly associated with processes underlying value computation (Rangel and Hare, 2010). Value computation occurs in different regions

such as the anterior cingulate cortex (ACC) (Silvetti, Seurinck & Verguts, 2012), ventral striatum (Kable & Glimcher, 2007; Knutson, Taylor, Kaufman, Peterson & Glover, 2005), medial prefrontal cortex (Hare et al., 2009; 2011; Rangel & Hare, 2010) and the orbitofrontal cortex (Sescousse, Redoute & Dreher, 2010). The nature of the value term being computed may differ, with some regions, such as the ACC being associated with reward predictions (Silvetti et al., 2012) and others, such as the ventro-medial prefrontal cortex (vmPFC), being associated with overall goal-value computation (Hare et al., 2009). Here we focus on evaluation of different stimuli in the process of decision-making. In this context, the vmPFC has been hypothesised to perform the role of a comparator (Basten, Biele, Heekeren & Fiebach, 2010; Wunderlich, Dayan & Dolan, 2012) into which value inputs from other regions flow. Therefore, it can be reasoned that self-control processes must influence the computations in the vmPFC. Empirically, there appears to be a consensus on the fact that the LPFC plays an influential role in self-control (Figner et al. 2010; Hare et al., 2009; 2011). Hare et al (2009, 2011) demonstrate that self-control in dietary choice is characterized by increased activation in the LPFC as well as increased functional coupling between the LPFC and the vmPFC.

The application of self-control can be seen as deliberative processing corresponding to a fast-learning, slow-acting system, while the lack of control would correspond to the output of a system that responds to, immediacy. We constructed a general model of self-control which included a lateral prefrontal LPFC pathway and a ventral-striatal pathway (recruiting the nucleus accumbens specifically) converging upon the ventro-medial prefrontal cortex (vmPFC), which is believed to be a value computational terminus as noted above. Self-control was conceptualized as the preponderance of LPFC inputs in the computation of the overall value of a stimulus, over the ones from the one computing immediate reward value. We augmented the model to include representations of delay, to explore inter-temporal decision making. Temporal information was represented using the widely-used tapped delay line approach (Freeman & Nicholson, 1970; Medina & Mauk, 2000). This model performed decision making by integrating magnitude and delay across model iterations to compute the goal value in the vmPFC region. The corresponding simulation studies and their results are reported in Chapter 3.

1.6 Testing the theory

Given their common point of origin and the fact that both instruction-following and self-control are reliably ascribed to processes in the lateral-prefrontal cortex (LPFC) in empirical studies as well as our models, we predicted that these two phenomena will be related to each other. More specifically, we hypothesised that a high degree of instruction-following would predict a high degree of self-control. We tested this hypothesis using behavioural experiments designed to elicit instruction-following behaviour as well as self-control. The findings of this study are reported in chapter 4. Finally, Chapter 5 presents the larger picture emerging from our use of computational tradeoffs to study instruction-following and self-control, questions raised by the research presented here, and directions for future research.

CHAPTER 2

WORD AND DEED: A COMPUTATIONAL MODEL OF INSTRUCTION FOLLOWING

Brain Research (2012) [1]

Instructions are an inextricable, yet poorly understood aspect of modern human life. Here we propose that instruction implementation and following can be understood as fast Hebbian learning in prefrontal cortex, which trains slower pathways (e.g., cortical-basal ganglia pathways). We present a computational model of instruction following that is used to simulate key behavioural and neuroimaging data on instruction following. We discuss the relationship between our model and other models of instruction following, the predictions derived from it, and directions for future investigation.

1. Introduction

[1] This paper was co-authored by Tom Verguts.

There are many types of learning. Human beings can learn through trial-and-error by interacting with the environment (Thorndike, 1911). This, however, is costly, time-consuming, and dangerous. Consider for example learning the consequences of touching fire or eating a strange berry, by trying it out. Learning by observation can circumvent the costs associated with trial-and-error (Bandura, 1977). With the advent of language, yet a third option became available. Verbal information sharing promotes group cohesion while facilitating learning at reduced temporal cost. Understandably, learning from instructions (often verbal) became an integral part of the learning repertoire of the human brain.

From *to do* lists to software manuals, instructions influence the lives of modern humans on multiple levels. Yet, how they are understood and implemented by the brain remains a mystery (Monsell, 1996). On the empirical side, a few recent studies have explored the effects of instructions on performance. A few themes emerge when the literature is surveyed. First, instructions can be rapidly and accurately implemented without explicit training (Ruge and Wolfensteller, 2010). Second, even verbally instructed mappings that have never been applied, can interfere with well-applied mappings if the two sets of mappings share stimulus dimensions (Cohen-Kdoshai & Meiran, 2009; De Houwer, Beckers, Vandorpe & Custers, 2005; Waszak, Wenke & Brass, 2008). A few recent studies have also explored the neural correlates of instruction following (Brass, Wenke, Spengler & Waszak, 2009; Cole, Bagic, Kass, & Schneider, 2010; Hartstra, Kühn, Verguts, & Brass, 2011; Ruge and Wolfensteller, 2010). Third, dissociations may arise between instructions and their implementation: In particular, patients exhibiting goal neglect can verbally report the required instructions without being able to implement them (Duncan et al., 1996; Luria, 1966).

On the theoretical front, instructions are crucial yet unexplained components in theories of cognition. For example, models of cognitive control typically incorporate task representations which implement the demands imposed by the task at hand (Botvinick, Braver, Barch, Carter & Cohen, 2001; Cohen, Dunbar & McClelland, 1990; Verguts & Notebaert, 2008). However, investigations of instruction following itself are less common. One exception is the computational modeling work by Noelle and Cottrell (1996). More recently, Helie and Ashby (2009, 2010) proposed a

computational account of learning that begins with explicit rules and ends with procedural knowledge. However, they do not explore the acquisition of the rule itself. Doll, Jacobs, Sanfey, and Frank (2009) recently proposed a model of instruction control of reinforcement learning. In their model prefrontal cortex (PFC) projects directly to motor cortex and the basal ganglia (BG) to select responses consistent with the instructions.

Here, we follow up on the Doll et al. (2009) approach and consider instruction learning and implementation as instantiations of Hebbian learning. For that purpose, we combine two complementary models of learning and automatisation, namely SPEED and COVIS. In SPEED (Ashby, Ennis & Spiering, 2007), learning occurs initially in the basal ganglia and is eventually transferred to cortex with an attendant increase in automaticity. In the COVIS framework (Ashby et al., 1998, 2011), performance is governed by two systems – a rule-based one (dependent on prefrontal cortex) and a procedural one (dependent on BG). We propose that a more general model that combines the main features of both SPEED and COVIS is suited to explain various forms of learning, including instruction following. In this framework, when instructions are provided, the prefrontal cortex learns them quickly, but executes them slowly (Boureau & Dayan, 2010; Daw et al., 2002). Indeed, novel learning typically activates prefrontal cortex (e.g., Miller & Cohen, 2001; Toni et al., 2001). Upon repeated application, the BG (which learn more slowly but execute more quickly) pick up the appropriate stimulus-response mapping by Hebbian learning, where the appropriate response is provided by the prefrontal route. Finally, after extensive application another cortical pathway would take over (hyperdirect pathway; Ashby et al., 2007).

With this general framework, we describe and test a model that focuses on the acquisition and transfer of instructed mappings. In the next section we describe the model and discuss its biological plausibility. This is followed by the simulation studies. Theoretical considerations and empirical predictions are elaborated in the General Discussion.

1.1. The model

1.1.1 Architecture

The first route in the model is indirect (left part of Figure 1); here, new instructions (such as “if you see a hexagon, press the left arrow key” and “if you see a square, press the right arrow key”) can be rapidly learnt. The second one is the direct route (right part of Figure 1); it gradually picks up the regularities implemented by the indirect route.

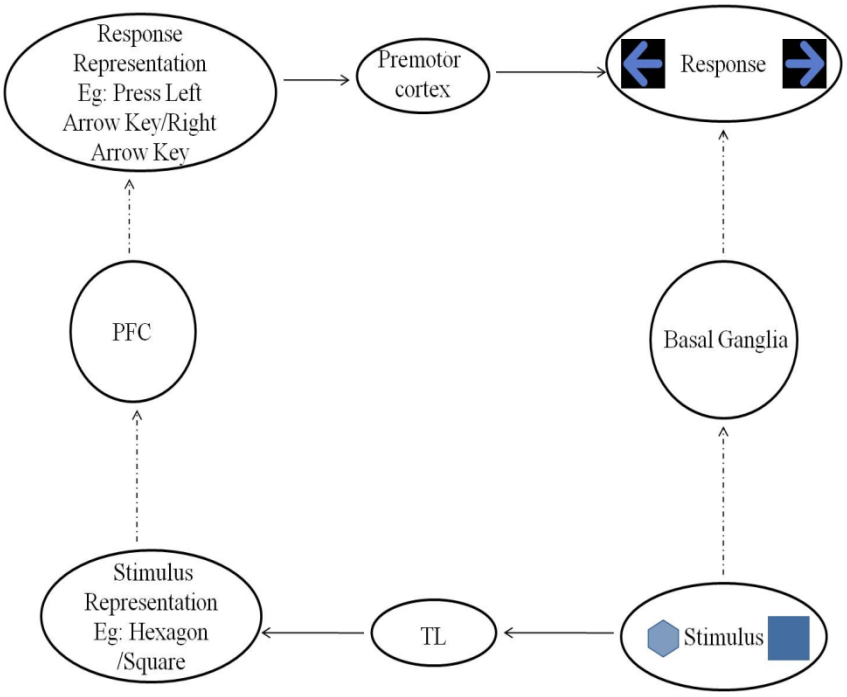

Figure 1. A schematic representation of the model. Weights from one layer to another are represented by an arrow connecting the two layers. Plastic weights are represented by dotted arrows. PFC = Prefrontal Cortex. TL = Temporal Lobe.

In the indirect route, the instruction is represented in terms of its components. One component contains stimulus representations (e.g., "hexagon") and the other response representations (e.g., "press right key"). These two components are typically (but not necessarily) verbal, encoded by two distinct layers (see Figure 1) and related to sensory and motor areas via long-term memory (temporal lobe (TL) and premotor cortex, respectively; see Figure 1). By premotor cortex we refer to the human analogue of the dorsal premotor cortex (PMd) in monkeys, which is a region associated with abstract motor planning (Nakayama, 2008).

In the case of verbal instructions, stimuli and responses are connected to their verbal equivalents (e.g., "hexagon", "left key"). It is reasonable to assume that a tight association between an object or attribute and its verbal analogue comes to be encoded during development (Fischer & Zwaan, 2008). Also, action verbs evoke activation of motor representations (Fischer & Zwaan, 2008; Hauk, Shtyrov & Pulvermüller, 2008). The two components of the indirect route are linked together by PFC; in particular, PFC subregion Inferior Frontal Junction (IFJ) appears to be a candidate for this role given that it is active in circumstances that require task-set switching or the loading of novel task sets (Derrfuss, Brass, von Cramon, Lohmann, & Amunts, 2009; Derrfuss, Brass, Neumann, & von Cramon, 2005), both of which require flexible verbal mapping. In the model, associating stimulus and response representations is achieved by fast Hebbian learning during the instruction phase. Neurally, the associative striatum is probably also part of the indirect path (Ashby et al., 2010), but we don't include it here for simplicity.

The direct route includes, in addition to the stimulus and response areas, the basal ganglia. The circuitry of the cortico-striato-pallido-thalamo-cortical pathway (Mink, 1996) is approximated by a one-layer excitatory path. In particular, fronto-striatal loops are not included, and the direct path can be considered a simplified version of subcortical control of action selection (e.g., Ashby, Turner, & Horvitz, 2010; Dominey, 2005; Frank, 2005, 2006). The direct route gradually acquires stimulus-response associations by Hebbian learning, where the correct stimulus-response pairs are provided by the indirect route. The hyperdirect cortical pathway mentioned in the introduction (e.g., Ashby et al., 2007) is not currently implemented.

It may be argued that the characterization of the indirect route as one having more intermediate steps than the direct one is contrary to the actual neural organization, where the BG has more intermediate synapses (e.g., Mink, 1996). However, the layout is consistent with the generally acknowledged finding that the PFC route is a slow processing route (e.g., Miller & Cohen, 2001). More generally, the number of synapses between two processing layers may be an imperfect measure of processing speed. Nevertheless, we report simulation studies that explore the influence of varying the respective path lengths in the two routes.

1.1.2 Dynamics

Information flows along the directions indicated by arrows in Figure 1. In each trial, activation in the stimulus layer is clamped, i.e., set at a specific value as opposed to allowing the layer to reach the activation level over time. Stimuli are represented by localist coding in a vector with the element corresponding to the stimulus set to 1 and all other elements set to zero. Activation of other model units is described by standard difference equations of the form (activation of input and output units denoted x and y, respectively):

$$y_j(t) = \tau y_j(t-1) + (1-\tau)\sum_i x_i w_{ij} \qquad (1)$$

where $y_j(t)$ is the activation of the j^{th} output unit at time t in the trial, $x_i w_{ij}$ is the net input from unit i to unit j, and τ is a cascade rate parameter (set at 0.9). A response is chosen if one of the response units reaches a threshold of 1. Reaction times (RTs) are calculated by counting the number of activation cycles (within a trial) needed to reach this threshold.

We now describe how learning occurs in the model. Initially, all weights are random values sampled from a uniform distribution between 0 and 0.01. After response, a competitive process selects the most active unit in the PFC and the most

active unit in BG. Typically, the winner's activation in a competitive model is a function of its original activation before the competition (e.g., Grossberg, 1973). As a simplified implementation of this process, we scaled the winner's activation (stimulus *j*) after competition by $\frac{1}{2}(1-e^{-N_{rep}(j)})$ with N_{rep} (*j*) the repetition number of stimulus *j* in the trial at hand. In the instruction phase, learning occurs only in the connections between stimulus representations, PFC and response representations. In the test phase, learning occurs only in the BG. In both layers, learning follows the rule given below from trial $n-1$ to n:

$$w_{ij}(n) = w_{ij}(n-1) + \lambda\left(x_i - dw_{ij}(n-1)\right)y_j \qquad (2)$$

where *d* is a weight decay parameter, set to 0.1. The term $dw_{ij}(n-1)$ is subtracted from the input x_i to constrain the learning process in the direction of the input (as is typical in a Hebbian / competitive learning algorithm, e.g., Fritzke, 1997).

The fact that only the PFC learned in the instruction phase, and the BG only in the test phase, implemented our assumption of fast learning in PFC. In their respective phases (instruction and test, for PFC and BG respectively), the learning rate was $\lambda = 9$ for each layer. In the simulations, we also explore the case when PFC learns in both instruction and test phases.

1.2 Phenomena simulated

1.2.1 From instructed to pragmatic representations (Ruge & Wolfensteller, 2010)

Instructions can be implemented with a high degree of accuracy on the very first trial (e.g., Cohen-Kdoshai et al., 2009; Cole et al., 2010). With increasing practice, the mapping loses novelty and becomes automatic as reflected, for example, in RT. Ruge and Wolfensteller (2010) studied the transition from instructed to implemented stimulus-response mappings using fMRI. They used a simple stimulus-response mapping task to identify the neural correlates of mappings that had been instructed and subsequently applied. Each stimulus was mapped onto one of two possible responses

(press left or right key). Four stimuli (two for each key) and their mappings were first instructed (instruction phase). This was followed by 32 practice trials in which each stimulus appeared 8 times in a randomized sequence (test phase). This procedure was repeated over 20 blocks with new stimuli in each block to obtain accurate fMRI data. Within each block, responses became faster with repetition. Error rates also decreased with repetition. At the neural level, activation levels across repetitions decreased in the left IFJ and increased in the BG (in particular, the caudate nucleus). The reported changes in other areas (e.g., decrease in left posterior intraparietal sulcus) are beyond the scope of the current study and hence not discussed.

1.2.2 Crosstalk of instructed and applied arbitrary S-R mappings (Waszak et al., 2008)

Waszak et al. (2008) investigated the effect of merely instructed and applied visuomotor mappings. They used stimuli varying on two dimensions (colour and shape) and subjects were presented with colour-task or shape-task trials intermixed. The tasks involved applying arbitrary mappings from colours and shapes to left and right responses (for instance, "if circle, then press left arrow key" or "if brown object, then press right arrow key"; see Figure 6). A third of the stimulus-response associations in each task were merely instructed and the other two-thirds were applied (i.e., trained). The irrelevant stimulus dimension allowed the stimulus in any given trial to be categorized as univalent, bivalent, or instructed. In the case of univalent stimuli, only the relevant stimulus dimension had a valid response mapping; for example, the stimulus would be a shape in a particular colour with only the shape being associated with a response. For bivalent stimuli, both relevant and irrelevant dimensions had valid response mappings. The instructed stimuli were similar to the bivalent ones, except that the irrelevant stimulus dimension and its mapping were merely instructed and had never been applied.

The experiment consisted of an instruction phase, followed by five practice blocks of 96 trials each. The fourth and fifth practice blocks were preceded by two test blocks (each lasting for 36 trials) in which the instructed stimuli were presented as

valid targets. This rendered these stimuli effectively bivalent for the final two practice blocks.

Waszak et al. (2008) hypothesized that the presentation of a stimulus with two dimensions having valid response mappings (with one of the two mappings being valid for a second task) would lead to an interference effect. *Interference effect* refers to the delay in responding to bivalent or instructed stimuli (congruent or incongruent) relative to univalent stimuli. In addition, the RT differences between incongruent and congruent stimuli for the bivalent and instructed types were also computed (*congruency effect*).

Waszak et al. (2008) found an interference effect for both bivalent and instructed stimuli, but it was larger for bivalent stimuli. Moreover, the interference effect was larger for the practice blocks after the test blocks than before the test blocks (see Figure 7a). They also found a congruency effect for bivalent stimuli across all practice blocks. In contrast, the instructed stimuli did not show a congruency effect in the first three practice blocks but an effect was observed in the final two practice blocks (see Figure 7c).

1.2.3 Goal neglect

Goal neglect refers to being able to describe an instruction while not being able to implement it. First reported in frontal lobe patients (Luria, 1966), Duncan, Emslie and Williams (1996) demonstrated that goal neglect can also be observed in normal subjects if instructions are sufficiently complex. In a recent study, Duncan, Parr, Woolgar, Thompson, Bright, Cox, et al. (2008) demonstrated that it is specifically the total number of elements to be remembered in the task which determines the extent of goal neglect and its correlation with general intelligence. We hypothesize that the modeling framework presented here may yield insights into goal neglect. Currently, we apply an adapted version of the Ruge design as a first step toward modeling goal neglect.

2. Experimental Procedure

2.1.1 Simulation study 1.1: Ruge and Wolfensteller (2010)

There were four units in the stimulus layer, corresponding to the four stimuli. Left and right responses were similarly encoded by two units in the response layer. The same coding scheme was applied to the (verbal) stimulus representations and (verbal) response representations (indirect route), with the same number of units in the respective layers. The PFC and BG layers had 200 units each.

The design of Ruge and Wolfensteller (2010) was replicated exactly, with 32 trials in the test phase. During the instruction phase, the stimuli and responses were presented as activation patterns in the stimulus representation and response representation layers (indirect route, left part of Figure 1) and the association was acquired by the PFC (Equation (2)). During the test phase, activation and learning obeyed Equations (1) and (2), respectively. PFC and BG activation vectors were collected before response competition for analysis. Mean RTs and mean error percentages were calculated across the 50 simulated subjects (as in Ruge and Wolfensteller, 2010).

2.1.2 Simulation study 1.2: Continuous learning in the PFC

Model parameters and settings were the same as in study 1.1, except that PFC was allowed to learn throughout the experiment. Behavioural and activation data were collected and analysed as above.

2.1.3 Simulation study 1.3: Model with equidistant paths from stimulus to response in both routes

The intermediate layers in the indirect pathway (temporal lobe and premotor cortex in the original model) and the verbal representations (verbal stimulus representation and response representation) were removed. As a consequence, the PFC and BG paths were rendered equidistant. Otherwise, model parameters were the same as in study 1.1. Behavioural and activation data were collected and analysed as above.

2.1.4 Simulation study 1.4: Model with equidistant paths from stimulus to response in both routes and PFC learning throughout the simulation

The architecture used in study 1.3 was used with one modification, namely the continuation of learning in the PFC throughout the simulation (as in study 1.2).

Model parameters were the same as in study 1.1. Behavioural and activation data were collected and analysed as above.

2.2 Simulation study 2: Waszak et al. (2008)

Model parameters were the same as in Simulation 1.1. Attention to the relevant stimulus dimension was encoded by a gating parameter (set at 0.6) by which the activation of the relevant stimulus dimension was multiplied (Verguts & Notebaert, 2008). The activation of the irrelevant stimulus dimension was scaled by 0.4. Colour and shape stimuli were encoded using two vectors of equal length (six elements each). There were two stimulus layers (6 nodes/layer), each providing input to the PFC and BG layers. The PFC and BG layers were divided into task-relevant sub-populations (S and C, for shapes and colours, respectively). Both PFC sub-populations and BG sub-populations had 1000 units each. Responses were the same across tasks and were encoded using two units.

The model was instructed in two consecutive phases (one for the color task, one for the shape task), before the start of the experiment, similar to Simulation 1. After the instructions, the stimuli (composed of two dimensions as described above) were presented. The model was given exactly the same task as the subjects in Waszak et al. (2008). It was exposed to 5 practice blocks of 96 trials each in which univalent, bivalent and instructed stimuli were randomized. The 4^{th} and 5^{th} practice blocks were preceded by two test blocks of 36 trials each (following the design of Waszak et al., 2008). RTs were calculated as before. Stimulus-specific RTs were averaged across the different subjects (simulations).

2.3 Simulation study 3: Simulating goal neglect

For a population of 100 models, the Ruge and Wolfensteller (2010) paradigm was adapted to include a verbal reporting task immediately following each of the 32 S-R trials. In the verbal reporting task, a stimulus probe was presented (the same stimulus presented for the S-R task) and the activation of the (verbal) response representation layer (on the left-hand side of Figure 1) was read out directly, corresponding to the production of a verbal response by the subject.

To simulate the rapid presentation of stimuli as in Duncan et al (2008) , the time to respond in the S-R task was limited (temporal deadline, tmax = 20, 40, or 60 cycles), and the time to respond to the verbal task held constant (temporal deadline 200 cycles). To implement the hypothesis that goal neglect is a function of the integrity of prefrontal learning, the learning rate of the PFC was set at 3 different levels (λ_{PFC} = 1, 5 and 9). The learning rate of the BG was kept at its original value of 9; however, similar results as those reported here were obtained when we varied this learning rate too. Goal neglect was quantified as the discrepancy between the mean percentage errors on the S-R and verbal task.

3. Results

3.1 From instructed to pragmatic representations

3.1.1 Simulation study 1.1

The major behavioral findings of Ruge and Wolfensteller (2010) were replicated. First, mean error percentage was zero (in the empirical study, it was very low and decreased across repetitions). Next, RTs decreased as a function of stimulus repetition (empirical and simulated data in Figure 2a and 2b, respectively).

In the empirical data, activation decreased in PFC across stimulus repetition (Figure 2c). The same is observed in the model (Figure 2d). Also the activation increase across repetitions in basal ganglia (caudate nucleus in Ruge & Wolfensteller, 2010; Figure 2e) was obtained in the model (Figure 2f). The reason why activation decreased in the PFC across trials is because the basal ganglia become faster with additional

learning (i.e., across trials), leaving less opportunity for the PFC to be highly active before response threshold is reached.

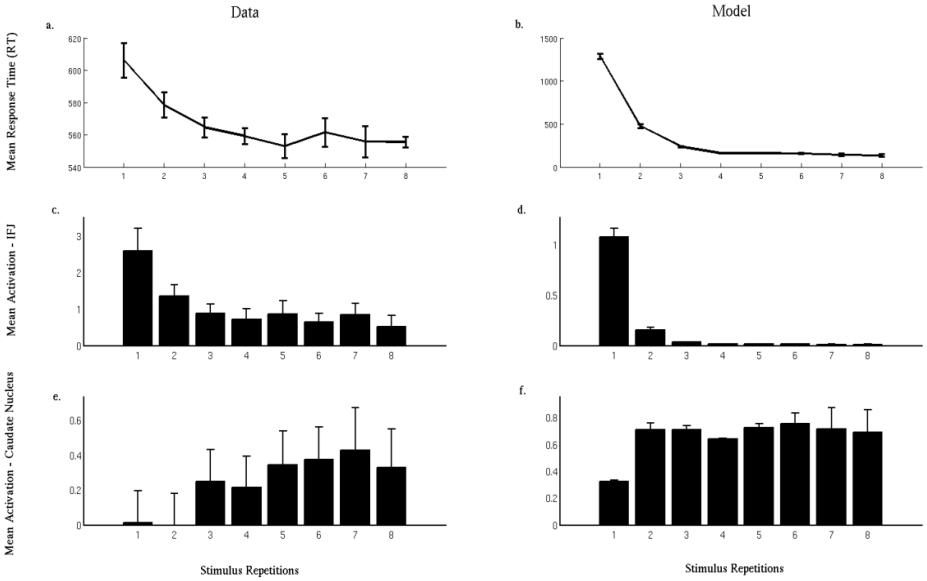

Figure 2: Simulation 1.1. 2a, RT curve as a function of stimulus repetition (Ruge & Wolfensteller, 2010). 2b, Model RT curve. Data points represent mean RT averaged across four stimuli and 16 subjects. Error bars indicate 95% confidence intervals. 2c, BOLD estimates for left-IFJ from Ruge and Wolfensteller (2009). 2d, activation levels in model PFC across repetitions. 2e, BOLD estimates for caudate nucleus from Ruge and Wolfensteller. 2f, activation levels in model basal ganglia across repetitions.

3.1.2 Simulation study 1.2: Continuous learning in the PFC

In the current simulation, learning continued throughout the task in PFC. Of the 50 simulations, 1 was excluded due to a large number of errors (more than 25%).The analyses were performed with the remaining 49 simulations. The major behavioral

findings of Ruge and Wolfensteller (2010) were replicated. First, mean error percentage was zero (in the empirical study, it was low and decreased across repetitions). RTs decreased as a function of stimulus repetition (empirical and simulated data in Figure 3a and 3b, respectively).

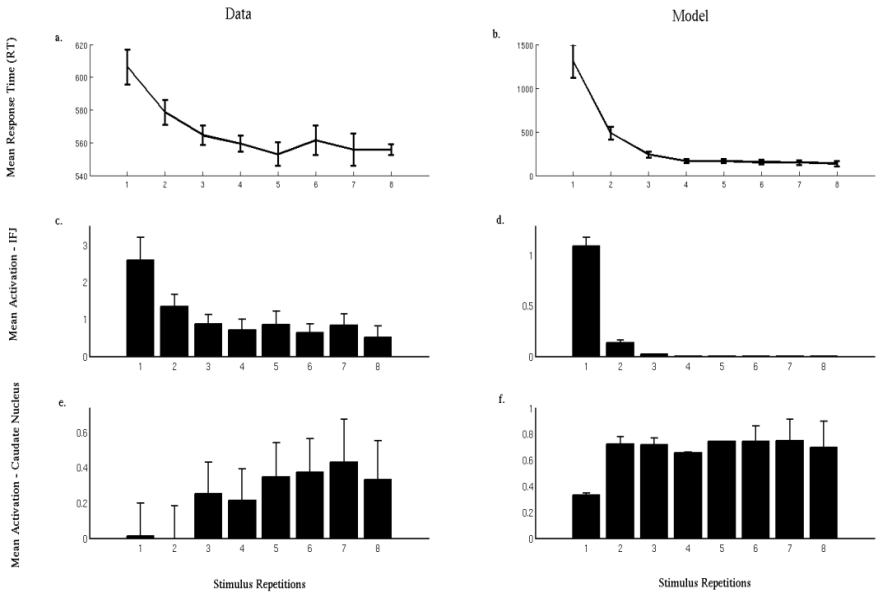

Figure 3. Simulation 1.2. 3a, RT curve as a function of stimulus repetition (Ruge & Wolfensteller, 2010). 3b, Model RT curve. Data points represent mean RT averaged across four stimuli and 16 subjects. Error bars indicate 95% confidence intervals. 3c, BOLD estimates for left-IFJ from Ruge and Wolfensteller (2009). 3d, activation levels in model PFC across repetitions. 3e, BOLD estimates for caudate nucleus from Ruge and Wolfensteller. 3f, activation levels in model basal ganglia across repetitions.

Empirically, activation decreased in PFC across stimulus repetition (Figure 3c). The same trend appears in the model (Figure 3d). Also the activation increase across

repetitions in BG (caudate nucleus in Ruge & Wolfensteller, 2010; Figure 3e) was obtained in the model (Figure 3f).

3.1.3 Simulation study 1.3: Model with equidistant paths from stimulus to response in both routes

In this study the synapses from stimulus to response layers through both routes were made equal in number so that the two routes were of equal length. The major behavioral findings of Ruge and Wolfensteller (2010) were replicated. First, mean error percentage was very low and remained below 0.01% (zero errors in the simulation). RTs decreased as a function of stimulus repetition (empirical and simulated data in Figure 4a and 4b, respectively).

As before, activation decreased in PFC across stimulus repetition (data, Figure 4c; model, Figure 4d). Again, as BG learn more, there is less opportunity for PFC to be as active as in the initial phase of responding, even though the paths are equidistant. Activation increases across repetitions in BG (caudate nucleus in Ruge & Wolfensteller, 2010, Figure 4e; model, Figure 4f).

Although the BG becomes increasingly active with learning, a complete switch from PFC to BG does not occur in this case (see Figure 4). For a complete switch to occur, it is required that PFC learns earlier than BG, but also that PFC acts slower than BG. This is not the case with equidistant paths, in which case the two paths act at approximately the same speed when both are well-trained.

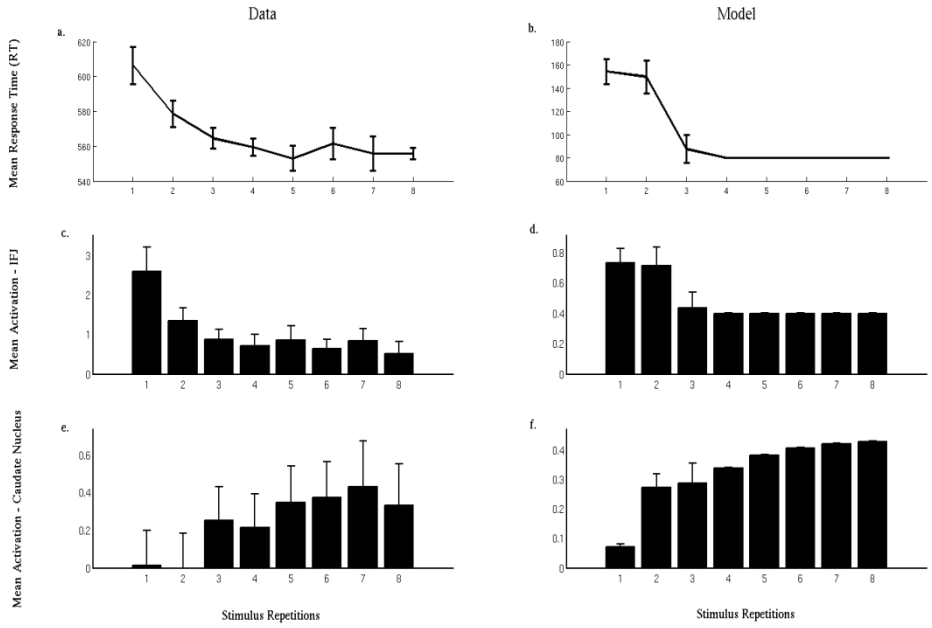

Figure 4: Simulation 1.3. 4a, RT curve as a function of stimulus repetition (Ruge & Wolfensteller, 2010). 4b, Model RT curve. Data points represent mean RT averaged across four stimuli and 16 subjects. Error bars indicate 95% confidence intervals. 4c, BOLD estimates for left-IFJ from Ruge and Wolfensteller (2009). 4d, activation levels in model PFC across repetitions. 4e, BOLD estimates for caudate nucleus from Ruge and Wolfensteller. 4f, activation levels in model basal ganglia across repetitions.

3.1.4 Simulation study 1.4: Model with equidistant paths from stimulus to response in both routes, with PFC learning throughout the simulation

Of the 50 simulations, 2 were excluded due to a large number of errors (25%). The analyses were performed with the remaining 48 simulations. Results were very similar as those obtained for study 1.3. Again, if the two paths act at approximately the same speed, there is no complete switch from PFC to BG.

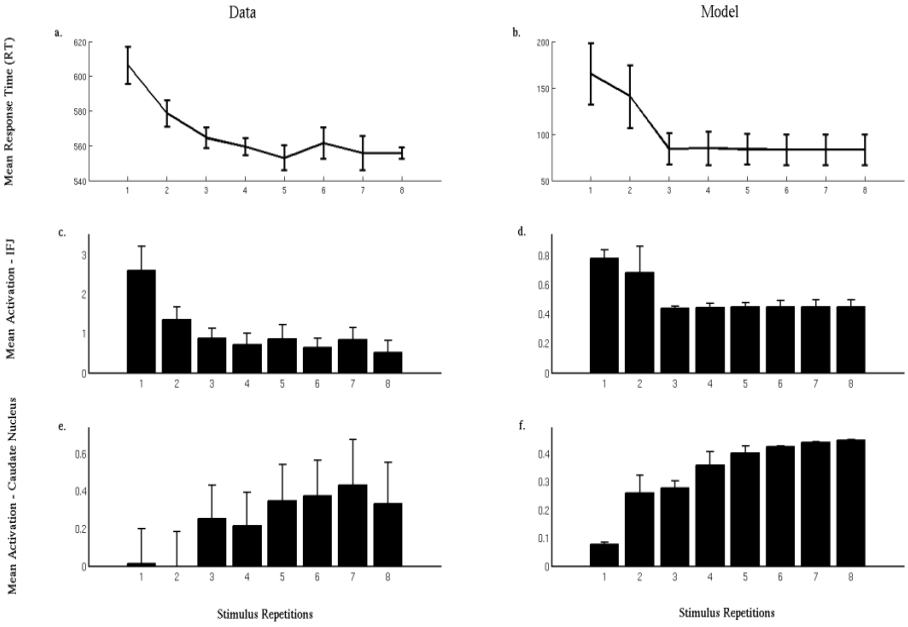

Figure 5: Simulation 1.4. 5a, RT curve as a function of stimulus repetition (Ruge & Wolfensteller, 2010). 5b, Model RT curve. Data points represent mean RT averaged

across four stimuli and 16 subjects. Error bars indicate 95% confidence intervals. 5c, BOLD estimates for left-IFJ from Ruge and Wolfensteller (2009). 5d, activation levels in model PFC across repetitions. 5e, BOLD estimates for caudate nucleus from Ruge and Wolfensteller. 5f, activation levels in model BG across repetitions.

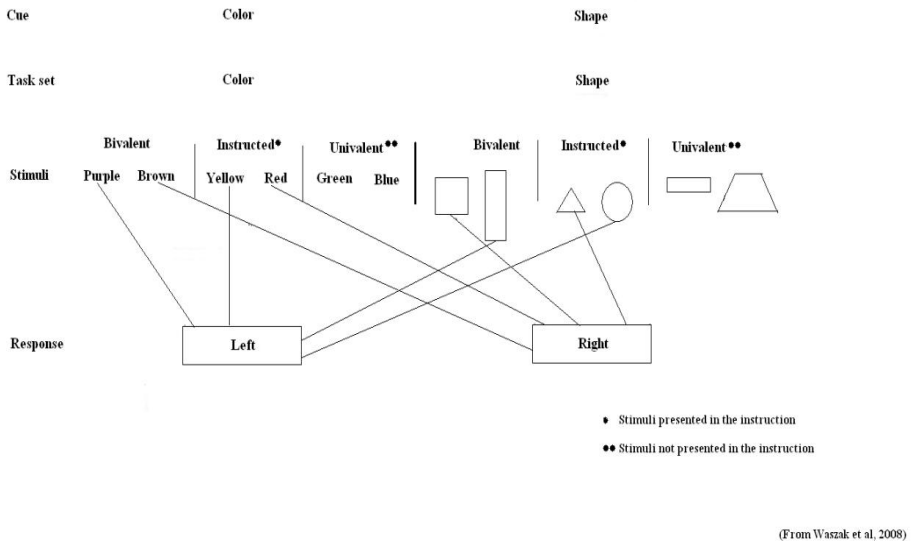

Figure 6 A representation of the task used in Waszak et al (2008), based on their original illustration.

3.2 Crosstalk of instructed and applied mappings

Of the 50 simulated subjects, 6 were excluded due to less than 90% accuracy. The simulation results were very similar to those of the original study (Waszak et al., 2008). First, there was an interference effect for both bivalent and instructed stimuli, which increased with practice (Figure 7b). Second, there was a congruency effect for bivalent stimuli which remained approximately constant after practice. In contrast, for the instructed stimuli, the congruency effect was initially very small for instructed stimuli (Figure 7d) but became larger after practice. One discrepancy is that the Waszak et al. study did not obtain a congruency effect for instructed stimuli whereas the model did. However, the model is consistent with other empirical studies that did find a congruence effect for instructed stimuli (Cohen-Kdoshai & Meiran, 2009; De Houwer et al., 2005).

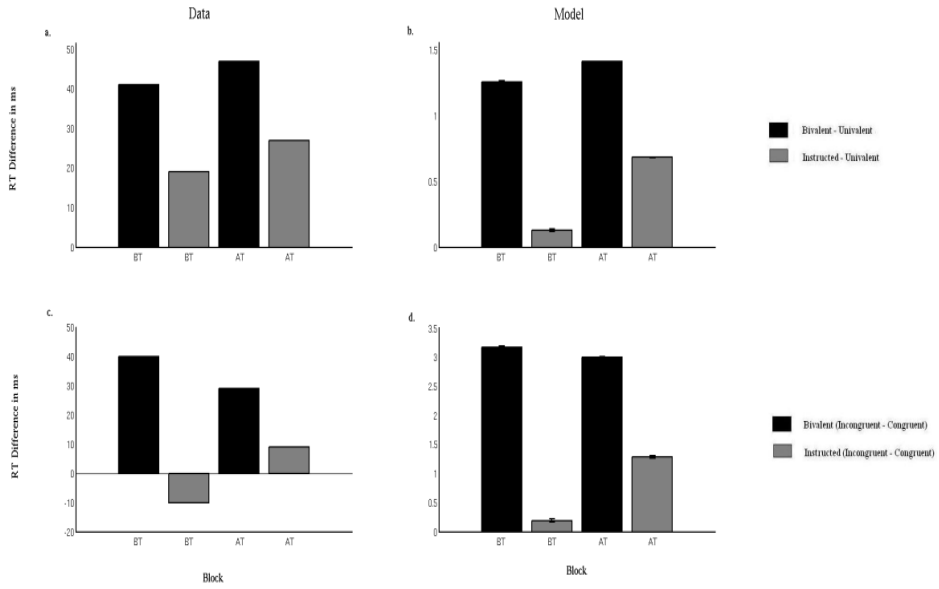

Figure 7: Simulation 2. Interference effect before test blocks (BT) and after test blocks (AT) in data (7a) and model (7b). Congruency effect (BT and AT) in data (7c) and model (7d)

3.3 Simulating Goal Neglect

Here, we applied the model to the important finding that instruction understanding and following can be dissociated (i.e., goal neglect, Luria, 1966). For this purpose, the Ruge et al. (2010) paradigm was extended to include a verbal reporting task immediately after each S-R trial. We chose this approach, rather than emulating the Duncan et al. design in great detail, because it allowed us to simulate the core phenomenon of goal neglect in the simplest way possible. In line with typical behavioral paradigms, different response deadlines were used for the two tasks (S-R

and verbal report), with the reporting task deadline much less strict. Percentage error in the S-R task relative to the verbal task was used as a measure of goal neglect (see Duncan et al., 2008) and evaluated across different PFC learning rates (λPFC) and S-R task response deadlines (tmax).

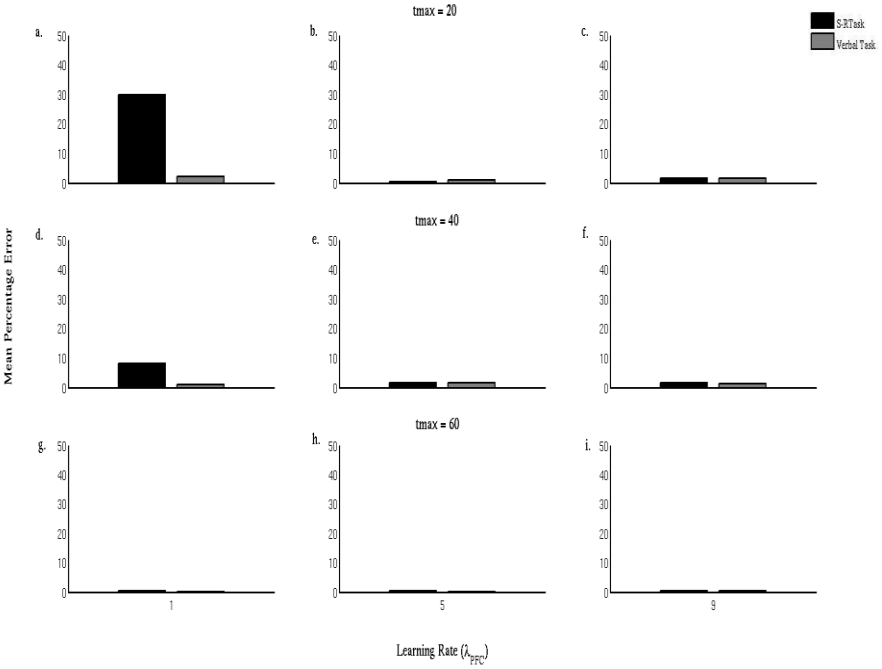

Figure 8: Simulation 3. Goal neglect. 3 × 3 plot of mean percentage error in the S-R and verbal tasks across prefrontal learning rates (λPFC, different columns) and S-R task response deadlines (tmax, different rows).

A very low PFC learning rate (λPFC = 1) in combination with a strict S-R task response deadline (tmax = 20) resulted in a greater percentage of errors on the S-R task (30%), with respect to the verbal task (2.3%) (similar to goal neglect; Figure 8a). This was ameliorated by increasing the S-R task response deadline tmax from 20 to 40 (resulting in error rates of 8.3% and 1% for the S-R and verbal tasks respectively) and

from 40 to 60 (resulting in errors rates of 0.5% and .3%, respectively) cycles (Figure 8d, 8g). In contrast, with moderately high PFC learning rates (λPFC =5, Figure 8b, 8e, 8h), or with high PFC learning rate (λPFC = 9, Figure 8c, 8f, 8i) there was overall low error rate (and the error rates were similar for the two tasks).

4. General Discussion

We presented a unified framework for instruction implementation, with Hebbian learning at its core. We derived a dual-route model from this framework and applied it to instruction following. The model was tested by simulating two recent experiments on instructions. The first simulation showed that sufficient practice can cause a switch from one route to another without need for a homunculus. However, a complete switch only occurred if the faster-learning path also acted more slowly. The second simulation study showed that merely instructed mappings can influence applied mappings (interference and congruency effects), but that they increase with practice. Finally, we proposed that goal neglect can be interpreted in this framework as emerging from a combination of an overall low learning rate (perhaps due to brain damage or dual tasking) and a strict response deadline for the S-R task. Consistent with this view, Czernochowski (2011) report evidence for impaired rule representations in older adults following short preparatory intervals but not long ones.

Learning in the BG was implemented as Hebbian learning. In general, though, Hebbian learning may be modulated by reward. In particular, if Hebbian learning is coincident with reward (phasic dopamine), then the Hebbian learning process is more efficient (e.g., Reynolds et al., 2001). However, given the very high accuracy rates in the experimental paradigms that we modeled, and given that explicit reward is never given, trial-to-trial variability in reward (phasic dopamine) is probably small. Hence, these influences are captured by the learning rate.

The model also yields a number of predictions for future empirical work. One prediction issuing from the model is that the connectivity between instruction-related areas should change across repeated implementation of the instruction. This prediction will be tested in future empirical work. Another prediction is that disruption to the

PFC will abolish influences of merely instructed stimuli in the Waszak et al. paradigm. This can be tested using rTMS or working memory loading. In addition, when these merely instructed stimuli become practiced, the disrupting effects should disappear. Finally, Simulation 1 suggests that the relative speeds of the PFC and BG routes determine the extent to which the processing switches from the PFC route to the BG one. Although this is difficult to test in humans, it may be testable using single-unit recordings.

In the remainder, we first discuss the relation between our approach and earlier models. Finally, we place the model in the broader framework of dual-route architectures and discuss the related but ill-understood topic of suggestions.

4.1 Models of instruction following

Noelle and Cottrell (1996) described a connectionist model of instruction following based on the simple recurrent network architecture (Elman, 1990; St-John & McClelland, 1990). Their approach to instruction following was *activation-based*, in the sense that a novel instruction was encoded as an activation pattern (Botvinick & Plaut, 2006). For instance, an instruction such as "if you see a hexagon, press the left key" would be encoded as a pattern of activation without changing the connection weights in the system. To be able to do this for any novel (untrained) instruction, without any learning (weight changes) during the presentation of these instructions, their model needed a long training phase with error backpropagation in which it acquired a set of instruction+stimulus-to-response mappings that could then be generalized to novel mappings of this type. Learning in our model is, by contrast, Hebbian, *weight-based* (cf. Botvinick & Plaut, 2006) and it occurred in the instruction phase (in the prefrontal connections) as well as in the practice phase (in the direct route).

In one sense, the long training phase of the Noelle et al. model corresponds to our hypothesized developmental acquisition of verbal and pragmatic associations. There are important differences, however. First, the Noelle et al. training regime used backpropagation, whereas ours can emerge from the simpler and more biologically

plausible Hebbian learning. Second, the associative structure emerging from the developmental process in our model can be more broadly recruited by other parts of the cognitive system, because it simply consists of associations between corresponding concepts (e.g., color red and word red).

As mentioned in the Introduction, Doll et al. (2009) also presented a model of instructional control based on the well-known model of reinforcement learning in BG advanced by Frank (2005, 2006), which is implemented in a simplified form as the direct route in our model. The Doll et al. model also employed fast learning in PFC to implement instructions. It was applied to a probabilistic selection task in which a third of the instructions were incorrect with regard to the reward probabilities associated with a particular pair of stimuli. Their model was applied to a case of conflict between instructions and contingencies. While it was not inconsistent with our own approach, we think that the current model adds several features of interest, in particular, a demonstration of how the Hebbian framework can deal with different "instruction following" phenomena described above; a theoretical account of when and why switching should occur (complementary systems are either fast-learning and slow-executing or vice versa); and an integration of the influential COVIS and SPEED models.

4.2 Dual-route architectures

Dual-route architectures are cognitive systems where one processing route is explicit, propositional, and effortful, and the other is implicit, associative (contingency-based), and effortless (e.g., Sloman , 1996). They have been claimed to underlie diverse neurocognitive phenomena from reasoning (Sloman, 1996) to self-regulation (Carver, Johnson & Joorman, 2008). Our model belongs to this general class with a propositional indirect route (of the form, "if …then…") and a contingency-based direct route. Despite clear and important differences between the two routes, we have proposed that a similar learning mechanism may underlie both. This similarity between explicit and implicit routes may have an evolutionary basis, given that selection tends to recycle previously successful structures and processes (Dehaene &

Cohen, 2007). Moreover, the modelling process demonstrated that sufficient practice can cause a switch from one route to another without any need for a homunculus.

Additionally, the model yielded the prediction that shifting from instructed to procedural knowledge consists of a gradual transition of response selection from PFC to BG. If the model were to be extended to the development of automaticity (e.g. Ashby et al., 2007), the dual-route architecture would be modified by the addition of a hyperdirect pathway which associates stimuli with responses without a mediating region. In our framework, learning in such a hyperdirect route would be even slower while it would be the fastest to execute the learnt mapping. Empirical (fMRI) tests of this prediction will be reported later on.

4.3 Instructions and verbally mediated phenomena

Suggestions are propositions used to influence an individual's beliefs and behaviour. Examples include alterations of autobiographical memories (Mazzoni et al, 2001), eyewitness accounts (Frenda, Nichols and Loftus, 2011), modification of Stroop effects (Raz and Campbell, 2011), and the placebo effect (Benedetti et al, 2005). While several studies have aimed at uncovering the neural correlates of phenomena mediated by suggestion (Craggs, Price, Perlstein, Verne, and Robinson, 2008; Petrovic, Kalso, Petersson, Andersson, Fransson and Ingvar, 2010), its nature remains poorly understood.

We propose that suggestions can be understood from the modeling framework proposed here. A suggestion would be rapidly learnt by the indirect route (specifically, the PFC) and applied immediately. This suggestion can be either cooperative or competitive with other processing routes (direct paths).The indirect route is expected to influence regions associated with attentional and evaluative processing, such as the anterior cingulate cortex (e.g., Ploghaus et al, 2003) or ventromedial PFC (Hare et al, 2011). Consistently, recent studies have identified increased PFC activation in suggestion-mediated analgesic effects (Derbyshire et al, 2009). Hence, the core principles proposed here can be used to generate a model to account for suggestion-

mediated phenomena. Currently, however, this remains an area for future investigation.

CHAPTER 3
OF VALUES AND THE WILL:

A COMPUTATIONAL MODEL OF SELF-CONTROL

Manuscript in preparation [1]

Self-control is a key aspect of adaptive behaviour and a predictor of long-term well-being in humans. Despite extensive empirical studies and more recent neurobiological investigations into the nature and substrates of self-control, computational accounts are scarce. Here we propose a computational model of self-control situated in a general computational framework based on a learning/acting tradeoff realized in fronto-striatal and striatal pathways. We conceptualize self-control as increased lateral-prefrontal influence over value computation. We report the results of two sets of simulations, one on dietary choice and the other on intertemporal choice. The first simulation study replicates the behavioural findings of Hare et al. (2009) on self-control in food-choice. The second simulation study explores the influence of self-control in intertemporal choice. It captured the phenomenon of preference-reversal (reflecting poor or no self-control) as well as choice-invariance over time (reflective of self-control). We discuss the model in the light of theoretical work on self-control as well as recent empirical findings and propose avenues for future research.

[1] This paper was co-authored by Tom Verguts.

1. Introduction

A hungry person steers clear of the neighborhood fast-food restaurant to go home and prepare a healthy meal. A child awaits her turn to be served dessert at the dinner table, and picks a single truffle when offered a few. A police officer subdues a hostile with force proportionate to the threat. Diverse though these scenarios may be, they have a common denominator; self-control. This prized human ability remains valued across cultures and has historically been celebrated in myths, legends and certain schools of philosophy (for example, Aurelius, circa 180 AD).

Self-control can be defined as the ability to overcome a habitual, automatic, impulsive or prepotent response or evaluation in order to achieve an instrumental goal (Muraven & Baumeister, 2000). In that it entails overcoming one response in favour of a more relevant one, self-control is akin to cognitive control (Botvinick et al., 2001). It is vital for adaptive behaviour and has profound implications for long-term success in various aspects of life (Casey et al., 2011; de Ridder et al., 2012; Moffitt et al., 2011; Tangney, Baumeister, & Boone, 2004). Loss of self-control is associated with maladaptive behaviours that lead to potentially disastrous consequences for the individual as well as society. For instance, addiction is considered to be a malfunction of the control mechanisms involved in adaptive behaviour (Baler and Volkow, 2006). Also, according to the general theory of crime, criminality is considered a consequence of a dysfunction of self-control (Gottfredson & Hirschi,1990; but see Wikstrom & Svensson, 2010).

Self-control has been characterized in various ways, including: a limited resource that gets depleted with use (ego-depletion; Baumeister, Bratslavsky, Muraven & Tice, 1998; Hagger et al. 2010; Hofmann, Vohs & Baumeister, 2012); internal commitment (Benhabib & Bisin, 2005); and a predominance of lateral-prefrontal cortical activity in decision making and valuation (McClure et al., 2004, 2007; Hare et al. 2009, 2011). However, it has not been studied as extensively from a computational perspective. In addition to parsimony and mechanistic integration of empirical

findings within (e.g. behavioural) and across (e.g. behavioural, neural) levels of investigation, computational modelling yields predictions that inform future research. To address the relative paucity of computational work on self-control in decision making, we extended our previous framework (Ramamoorthy & Verguts, 2012) to this phenomenon. In particular, we explored two instances of self-control in decision-making; dietary choice and intertemporal choice. In what follows, we discuss these two instances in detail and present the theoretical framework in which the model is situated.

Individuals are faced with dietary choices on a daily basis – whether to consume healthy food items or to consume foods of convenience which are more readily available or less expensive. Hare et al. (2009) investigated the behavioural and neural aspects of dietary choice in self-controllers and non-self-controllers. In their study, active dieters were recruited to perform a dietary choice task. The subjects rated a set of food items on the basis of taste and health independently. An item rated neutrally on both scales by a particular subject was chosen as the reference item for that subject. This was followed by a decision block in which subjects had to choose between the reference item and a test item on every trial. Food items were classified into four types; disliked-unhealthy, disliked-healthy, liked-unhealthy and liked-healthy. Subjects were classified into self-controllers and non-self-controllers based on performance. Figure 4a presents the key behavioural results of this study. All subjects, regardless of self-control status, tended to avoid disliked-unhealthy items but chose liked-healthy items. The food-types of interest were the disliked-healthy and liked-unhealthy types. Self-controllers chose the disliked-healthy items more often, and rejected the liked-unhealthy items, whereas the non-self-controllers rejected the disliked-healthy items and chose the liked-unhealthy items more often. Simulation 1 details the replication of these findings by our model.

The second instance concerns choices that affect rewards over time. Intertemporal decision making has been extensively studied in economics (Ainslie, 1992, 2001; Laibson, 1997; Diamond, 2003) and has recently been investigated using the tools of cognitive neuroscience (McClure et al 2004, 2007). The core of this form of decision-making is as follows; given a choice, should an agent select a sooner but

smaller reward, or defer selection in favour of a larger reward that can only be attained later? In reinforcement learning terminology, this is the problem of delayed gratification (Sutton & Barto, 1998). Economists expect a rational human agent to discount exponentially when faced with intertemporal choices, meaning that value at a given delay is only worth a fraction exp(-*K**delay) of the immediate value, in which *K* is a discounting parameter. This implies that the agent's preferences remain unchanged across delays. If the agent prefers a larger reward to be experienced later relative to a smaller reward available more immediately, this preference should remain unchanged if a fixed amount of time is added to the waiting times for each reward. Human behaviour, however, is largely inconsistent with this picture. Subjects presented with intertemporal choices do not typically display choice-invariance across delays to the rewards (Loewenstein & Prelec, 1992).

Consider the larger-later and smaller-sooner rewards in Figure 1. Time T3 and T4 correspond to the points at which the smaller-sooner and larger-later rewards are delivered, respectively. At time T1, the larger-later reward

is valued more than the smaller-sooner reward. As time advances and approaches T2, the values of the smaller-sooner and larger-later rewards become more similar. At T2, the agent is indifferent to the options. Between T2 and T3, the curves have crossed, implying a reversal of preferences relative to T1. Humans typically choose the smaller-sooner reward instead of waiting until T4 (Ainslie, 1992, 2001; O'Donoghue and Rabin, 1999). This has also been demonstrated in pigeons (Mazur & Biondi, 2009) and rats (Reynolds, de Wit & Richards, 2002).

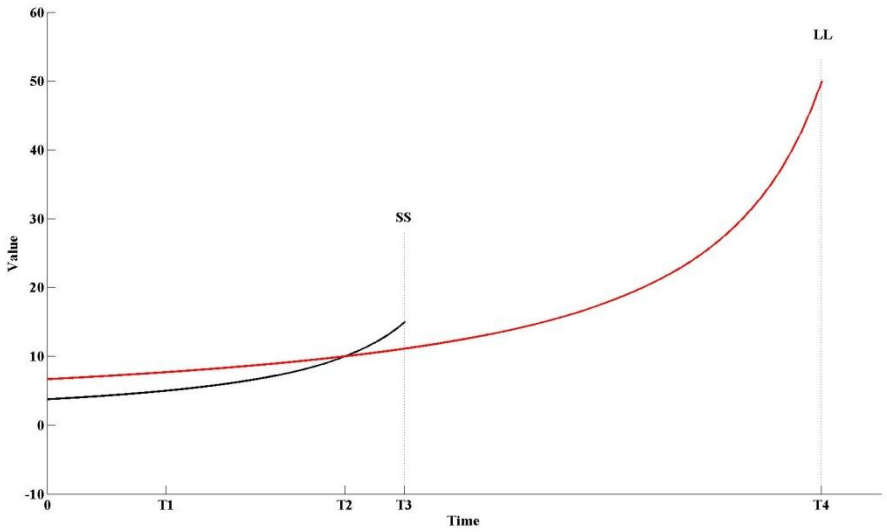

Figure 1: The evolution of subjective value over time in intertemporal choice. SS refers to the smaller-sooner reward and LL refers to the larger-later reward. T1 is the initial time point, T2 is the point of indifference and T3 and T4 represent the time-points at which the SS and LL are delivered, respectively.

Instead of exponential, value can also be discounted hyperbolically. In this case, value changes over delays by a factor of 1/(1+*K**delay), with *K* a discounting parameter. Hyperbolic models can predict the reversal of preferences shown in Figure 1, and accordingly fit behavioural data pertaining to intertemporal choices better than exponential models (Ainslie, 1992; Kirby & Marakovic, 2002; Kirby & Herrnstein, 1995; for a contrasting viewpoint, see Rubinstein, 2003). Models that also are able to predict this shift, but have a functional form different than hyperbolic, are termed quasi-hyperbolic (Laibson, 1997). From a behavioural/economic perspective, the reversal of preferences at a time when the smaller-sooner reward is closer, is interpreted as a lack of self-control. Self-control would correspond to remaining

consistent in valuation of rewards over time. Simulation 2 explores reversal as well as non-reversal, of preferences, in intertemporal choices, within the context of self-control.

Here we present an integrative view on self-control situated in a broader theoretical framework on the computational tradeoff between fast acting versus fast learning (Boureau & Dayan, 2007). In a general decision making context Daw et al. (2005) argued that a prefrontal tree-search system and a dorsolateral striatal cached system compete for behavioural control on the basis of the relative uncertainty associated with responses computed by the two systems. In their model, tree-search would correspond to fast learning with slow responding (as traversing the tree takes time). Cached learning would be slower, but cached responses faster in execution (no tree search required). Our model (Ramamoorthy & Verguts, 2012) implemented this principle (fast learning vs. fast acting) as applied to instruction following. Here, we present a computational account of self-control that is consistent with the general theoretical framework. We focus in particular on self-control in the context of competition between valuation pathways, such that one pathway learns values fast, while exerting influence over decisions slowly, while the other pathway acquires values slowly but responds faster. This scheme is used to account for self-control scenarios involving different dimensions of value (for instance, taste and health in Simulation 1) or different predictions of value over time (Simulation 2).

2. Method

2.1 Model Architecture

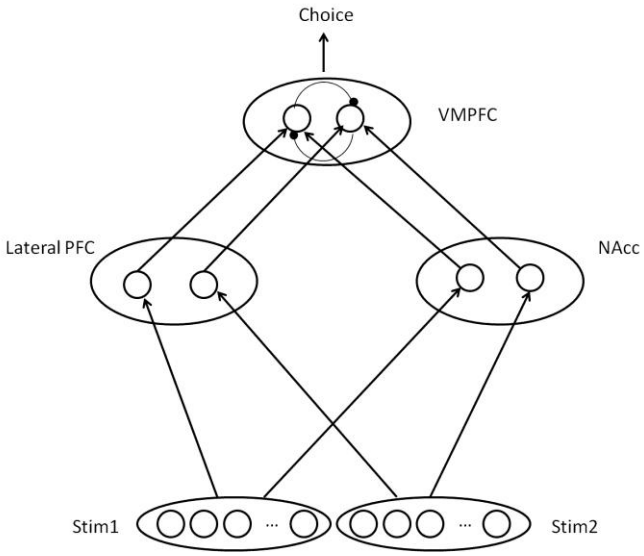

Figure 2: Schematic diagram of the model of self-control as applied to dietary choice. NAcc = Nucleus Accumbens. PFC = Prefrontal Cortex. VMPFC = Ventromedial Prefrontal Cortex

The model has two separate valuation pathways, the lateral-prefrontal (LPFC) pathway and the ventral-striatal pathway, each having the ventromedial prefrontal (VMPFC) cortex as a terminus (see Figure 2).

The VMPFC has been hypothesised to integrate different value-inputs from other regions (Basten et al., 2010; Hare et al., 2010; Philiastides, Biele & Heekeren, 2010; Smith et al., 2010;) and computes a value (Harris et al., 2011) that guides decision making (Grabenhorst & Rolls, 2011).

The ventral-striatal pathway consists of sensory input areas, the Nucleus Accumbens (NAcc), and the VMPFC (Hare et al., 2009) as a terminus. The NAcc receives input from the sensory cortices, and projects (via ventral pallidum and thalamus) to VMPFC (Pierce & Kumaresan, 2006). The conception of this pathway is consistent with several findings showing that the NAcc plays a role in value-computation and decision making (Knutson et al. 2005; Samanez-Larkin et al., 2010; Peters & Büchel, 2010).

The LPFC pathway starts with sensory input areas projecting to lateral-prefrontal areas. These areas are functionally connected to the VMPFC (Hare et al., 2009, 2011). The role of LPFC in self-control is supported by several empirical findings. Increased LPFC function is proposed as a correlate of self-control by Hare et al. (2009, 2011). Consistent with this, reduced LPFC sensitivity to reward reflects impaired self-control in addiction (Goldstein et al., 2009).

The model was implemented as a feedforward network with the activation propagating along the direction specified by the arrows in Figure 2. Activation in the stimulus layers (Stim1 and Stim2 for first and second stimulus, respectively) was encoded locally (unit *i* coding for stimulus *i*).

2.2 Simulation 1: Self-control in dietary choice

The taste and health ratings associated with a given stimulus were encoded in the connection strengths (weights) between the stimulus layers and the NAcc and LPFC layers respectively. The connection strengths varied in increments of 1, from 1 to 5 (to reflect the rating scale used). Taste ratings are construed to reflect innate values (learned at phylogenetic scale) or instead reflect experientially acquired values that take time to learn but are subsequently cached. In either case, responding is fast. Health ratings, on the other hand, are acquired fast, perhaps socially through instructions (e.g. parent pontificating to child on benefits of broccoli) but are applied more slowly.

Activations of NAcc and LPFC units were calculated using standard difference equations of the form:

$$V_{NAccj}(t) = \tau V_{NAccj}(t-1) + (1-\tau)\sum_i x_i w_{NAcc_{ij}} \quad (1)$$

$$V_{PFCj}(t) = \tau V_{PFCj}(t-1) + (1-\tau)\sum_i x_i w_{PFC_{ij}} \quad (2)$$

where V_{NAccj} and V_{PFCj} are the activations of the j^{th} unit in the corresponding area (NAcc , LPFC) at time t in the trial (j = 1 or 2, because there are two options on each trial), $x_i w_{ij}$ is the net input from input-layer unit i to unit j and corresponds to the magnitude of the pertinent dimension of value, and τ is a cascade rate parameter (set to 0.5).

Several studies (Hare et al., 2011; Hollman et al., 2012) report a regulatory influence of the LPFC on VMPFC. We implemented this influence through a gating variable S, which constrained the LPFC and NAcc inputs to the VMPFC as follows:

$$V_j(t) = \tau V_j(t-1) + (1-\tau)(SV_{PFCj} + (1-S)V_{NAccj} + V_k w_{inh}) \quad (3)$$

where $V_j(t)$ is the value computed by the j^{th} unit in the VMPFC layer at time t, representing the value of the j^{th} stimulus (option). V_{NAccj} and V_{PFCj} are value-inputs from the NAcc and LPFC layers respectively, τ is a cascade rate parameter (set to 0.5) and S is a gating variable that constrains the contributions of the two pathways to the overall input to the VMPFC. The VMPFC units compete through lateral inhibition. This is represented by the inhibitory input $V_k w_{\mathrm{inh}}$, where V_k is the activation of the competing VMPFC unit, and w_{inh} (< 0) is the strength of the inhibitory connection between the two VMPFC units.

2.2.1 Simulation 1.1: No ego-depletion

We replicated the design of Hare et al. (2009) which investigated self-control in food choice (Fig. 2). 50 stimulus items (tokens for food items) were evaluated relative to a reference item and a choice was made on each trial. Following Hare et al. (2009), we used a rating scheme to qualify the items on two dimensions taste and health. In the model, connections between stimulus layer and the NAcc varied from 1 to 5 units in strength, to correspond to the rating scheme used. Health as an abstract value was encoded in the weights between stimulus layer and PFC with the same range of values. The VMPFC integrated the two value components. The model was considered to have made a choice when the maximally responsive VMPFC unit reached threshold (set at 2), or if the maximum number of model cycles, set at 250, was reached. This selection informed the choice of the corresponding option.

The reference object was selected to be neutral in health and taste. Each stimulus carried value on two dimensions and this was randomized so that any given stimulus could have any combination of the two. Four kinds of stimuli were identified on this basis; liked-healthy, liked-unhealthy, disliked-healthy and disliked-unhealthy.

To generate individual differences between models, gating variables were randomly sampled from a uniform distribution between 0 and 1 for each of 50 simulated subjects. Simulated subjects were classified as self-controllers or non-self-controllers based on whether the gating parameter was bigger or smaller than 0.5.

All choices made by each simulated subject were classified according to the taste-health rating scheme described above. The mean proportion of choices per category were calculated across subjects and plotted to examine the influence of the gating parameter on choice.

2.2.2 Simulation 1.2: Ego-depletion case

Model architecture, parameters and simulation specifics were the same as in Simulation 1.1 except for the gating parameter *S*, which was subject to ego-depletion. Ego-depletion (Baumeister, et al., 1998; Baumeister et al., 2008) is a conception of

self-control as a limited resource and refers to the exhaustion of self-control as a function of repeated application. It was implemented here as a trial-dependent decay in the gating variable. In particular, the gate grew smaller over time and with each decision made by the model across trials:

$$S(\mathrm{n}) = \mu^{\mathrm{n}} S_0 \qquad (4)$$

where *S(n)* is the gating variable for trial *n*, μ is a decay constant set to 0.99 and S_0 is the initial value of the gating variable. An alternative labeling for the process described by equation (4) could be habituation or neural fatigue (Grill-Spector et al., 2006). We are theoretically not committed to the biological origin of the effect.

2.2.3 Simulation 1.3 Self-control as a function of gating variable – without ego-depletion

To study the effects of gating on choice, we produced a family of models with different gating parameters. Each had the architectural and parametric specifications described in study 1.1 but the gating variables were systematically varied from 0.05 to 0.9 (with intervals of 0.05). Each model was subjected to the Hare paradigm as described in Simulation 1.1, and the extent of self-control exerted by the model was quantified as the difference between the percentage of selections of disliked-healthy items and liked-unhealthy items. This quantity was estimated for each value of the gating parameter and plotted to examine the effect of gating strength on self-control.

2.2.4 Simulation 1.4: Self-control as a function of gating variable – with ego-depletion

A family of models conforming to the architectural and parametric specifications described in Simulation 1.1 was constructed, with gating variables varied from 0.05 to 0.9 (with intervals of 0.05). Ego-depletion was introduced in the form of a reduction in gating parameter values with each trial, consistent with equation (4) above. The procedure was the same as in Simulation 1.3 in all other respects.

2.3 Simulation 2: Intertemporal choice

In Simulation 2, we extended the model to discounting in intertemporal choice. We reasoned that for such choices, especially hypothetical ones, the two valuation systems would simulate the evolution of value across time for each option separately. This would correspond to prospection, or simulating future possibilities to guide decision-making. To capture this, the architecture was modified slightly to include time representations in the LPFC and NAcc layers (see Figure 3). The LPFC and NAcc layers were divided into two sets of units; one computed the magnitude of the stimulus and the other the delay associated with it.

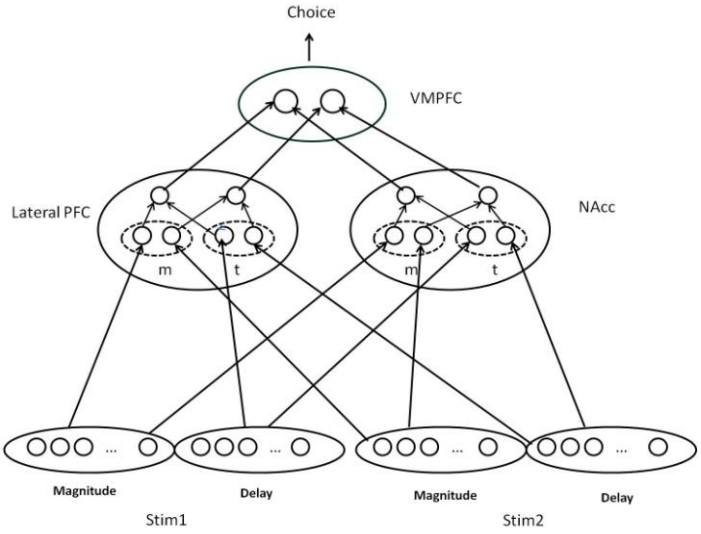

Figure 3. Schematic diagram of the model of self-control as applied to intertemporal choice. NAcc = Nucleus Accumbens. PFC = Prefrontal Cortex. VMPFC = Ventromedial Prefrontal Cortex. 'm' and 't' sub-layers correspond to magnitude and time representations, respectively**.**

Reward magnitudes (values) were encoded in the connection strengths (weights) between the stimulus layers and the NAcc and LPFC layers respectively (cf. Simulation 1). In the NAcc and LPFC layers, net activation was calculated as a product of the magnitude (net input) and the delay to reward. Magnitude was computed as in Simulation 1. This applied to units in NAcc and LPFC.

Following Buonomano and Merzenich (1995), we encoded temporal information as a change in response patterns across time. Delay was encoded as a travelling spike, with the amplitude of the spike decreasing across time. The basic time representation scheme assumed an array of neurons, with each neuron responding to a certain unit of time. This is akin to the tapped delay line formalism used in the study of temporal processing in the cerebellum (Freeman & Nicholson, 1970; Medina & Mauk, 2000) and basal ganglia (Brown, Bullock & Grossberg, 1999; Niv, O Duff & Dayan, 2005).

Time representations were integrated with magnitude information to compute the value components of each region as follows:

$$V_{NAccj}(t) = \frac{1}{t}\sum_{i} x_i w_{ij} \tag{5}$$

$$V_{PFCj}(t) = \frac{1}{t^r}\sum_{i} x_i w_{ij} \tag{6}$$

for option *j*. This prediction of the future value $V(t)$ was calculated across time points t until the delay associated with each option was reached (i.e., V(delay) was calculated). Given that prefrontal and ventral-striatal valuation networks encode abstract and immediate values respectively, it follows that this is reflected in their responses to delay. This was captured by making the time representations in the NAcc and LPFC layers decay across their respective delay lines. In the NAcc layer the decay was the reciprocal of the current timestep; in the LPFC layer, the decay assumed an exponent of the timestep of the current model cycle, with the exponent r set to 0.5.

In each timestep VMPFC units integrated these two value components as per equation (3) above. In this simulation, the VMPFC units representing the two alternatives receive temporally decaying inputs from the NAcc and LPFC regions, with the unit representing the smaller-sooner option completing its computation of value sooner than the unit representing the larger-later reward. This obviates the need for active competitive inhibition and therefore the inhibitory connections in the VMPFC layer were set to 0.

2.3.1 Simulation 2.1: Weak gating without ego-depletion

We presented the intertemporal choice paradigm to each of 50 simulated subjects. Each simulated subject was exposed to 50 trials. Within a given trial, the simulated subject was presented with varying delays to reward delivery, for a given pair of rewards and delays (the smaller-sooner and larger-later options). Each simulated subject had an architecture corresponding to Figure 3. We model delay as described earlier. The overall value was computed by the VMPFC layer, as a weighted (gated) sum of the PFC and NAcc components (cf. equation (3)). The gating parameter was set to 0.3.

In each trial the model was presented with stimuli representing two magnitudes and their respective delays. The smaller-sooner (SS) reward was fixed at a magnitude of 10 units and the larger-later (LL) reward was 100 units. The time of delivery for the smaller-sooner reward was always immediate (i.e., within a fixed number of model cycles, less than or equal to 20) and the larger-later reward had a delivery time

corresponding to a maximum of 110 model cycles. The model simulated the evolution of each stimulus value over its corresponding delay period. This was iterated over a number of delays. In each trial (choice between two stimuli), the stimulus values were computed over the corresponding delays to reward (each delay took the subject closer to the rewards). These values were collected and averaged across blocks and subjects to generate the value versus time plots.

2.3.2 Simulation 2.2: Strong gating without ego-depletion

Model architecture and parameters were the same as in Simulation 2.1, except for the gating parameter which was set to 0.9. This was done to explore the effect of increasing LPFC input to the VMPFC on intertemporal choice. We hypothesized that self-control would be captured by increased LPFC influence on VMPFC computations, given that LPFC is less sensitive to immediacy than the NAcc in our model.

2.3.3 Simulation 2.3: Weak gating with ego-depletion

In this study, we explored the effect of ego-depletion on model performance in intertemporal choice. Model architecture and parameters were the same as in Simulation 2.1, except for the gating parameter. As in Simulation 2.1, it was set to 0.3, but ego-depletion was introduced in the form of reduction in gating parameter values with each trial, consistent with equation (4) above.

2.3.4 Simulation 2.4: Strong gating with ego-depletion

In this study, we explored the effect of ego-depletion on model performance in intertemporal choice. Model architecture and parameters were the same as in Simulation 2.1, except for the gating parameter. As in Simulation 2.2, it was set to 0.9, but ego-depletion was introduced in the form of reduction in gating parameter values with each trial, consistent with equation (4) above.

3. Results

3.1.1 Simulation 1.1: Self- control in food choice (no ego-depletion)

The simulations replicated the findings of Hare et al (2009). Fig. 4b shows the proportion of choices made by the self-controller and non-self-controller groups of simulated subjects (identified on the basis of the gating parameter values). As in the original study, both self-controllers and non-self-controllers among the simulated subjects choose the liked-healthy option and tend not to select the disliked-unhealthy option. Non-self-controller models select the unhealthy-liked option whereas the self-controllers do not. In

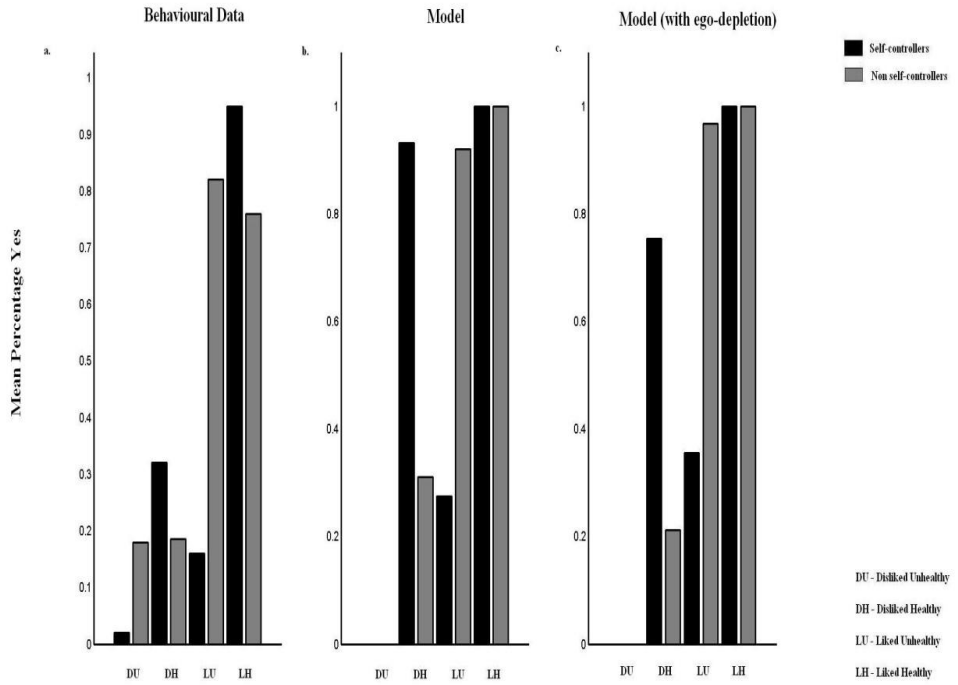

Figure 4: Simulation 1: Self-control in dietary choice

(4a) Behavioural results from Hare et al. (2009) (4b) Simulation results from model (4c) Simulation results from model with ego-depletion.

contrast, self-controllers choose the disliked-healthy option more often than the non-self-controller models.

3.1.2 Simulation 1.2: Self- control in food choice (with ego-depletion)

Fig. 4c shows the proportion of choices made by the self-controller and non-self-controller groups of simulated subjects (identified on the basis of the gating parameter values). The pattern of results remains the same as those reported above, with two minor differences. Self-controller models choose the disliked healthy option, but slightly less often than in the no ego-depletion case. They also chose the liked-unhealthy option more often.

3.1.3 Simulation 1.3: Self-control as a function of gating variable (without ego-depletion)

The extent of self-control, quantified as difference between percentage of disliked-healthy choices and liked-unhealthy choices increased as a function of gating. In the absence of ego-depletion, this increase reaches its highest point at a gate value of 0.7 and remains stable thereafter (Figure 5a).

3.1.4 Simulation 1.4: Self-control as a function of gating variable (with ego-depletion)

The extent of self-control increased as a function of the gating variable. Given ego depletion, this increase is slower (Figure 5b) compared to that seen in Simulation 1.3 (Figure 5a).

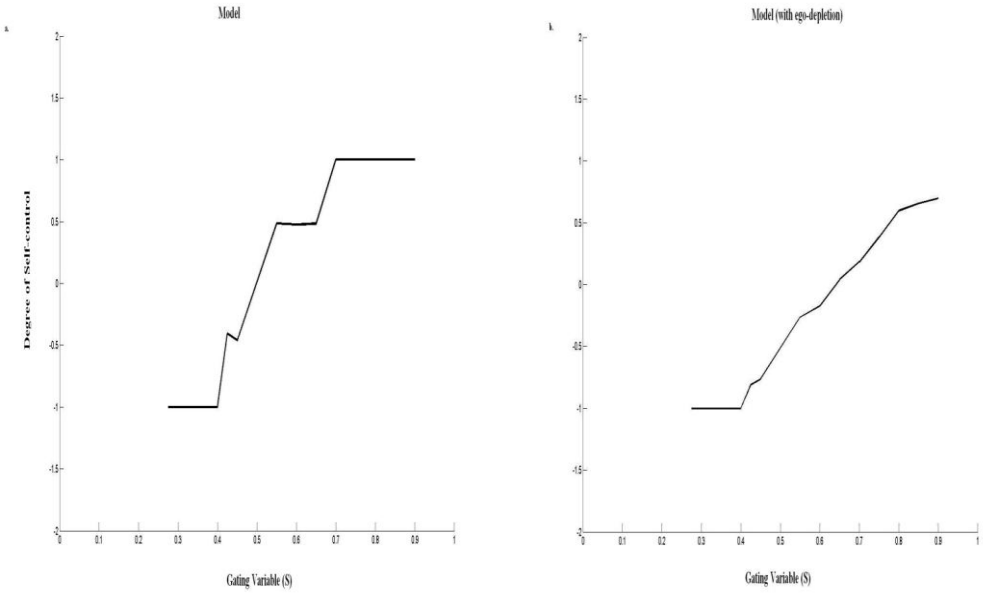

Figure 5: Extent of self-control as a function of the gating variable: (5a) without ego-depletion; (5b) with ego-depletion.

3.2.1 Simulation 2.1: Self-control in intertemporal choice (weak gating without ego-depletion)

The major behavioral finding of quasi-hyperbolic discounting in intertemporal choice was replicated, corresponding to an absence of self-control. In Fig. 6a, the model is shown to be tracking the evolution of value over time. At Time T1, the larger-later reward has a higher value compared to the smaller-sooner reward. As time approaches T2, there is a crossing of the curves. At time T3, the model shows reversal of preferences and chooses the smaller-sooner reward.

3.2.2 Simulation 2.2: Self-control in intertemporal choice (Strong gating without ego-depletion)

Fig. 6b shows the evolution of value over time in intertemporal choice. Here the larger-later reward has a higher value from the start of the trial (at time T1), and this increases with decreasing delay to reward. The smaller-sooner reward increases in value but always remains below the curve corresponding to the larger-later reward's subjective value over time, even at the time point T3 that marks its delivery. This is due to the preponderance of the LPFC valuation system over the NAcc system consistently across delays. As noted in section 1.1, this corresponds to exponential discounting.

.

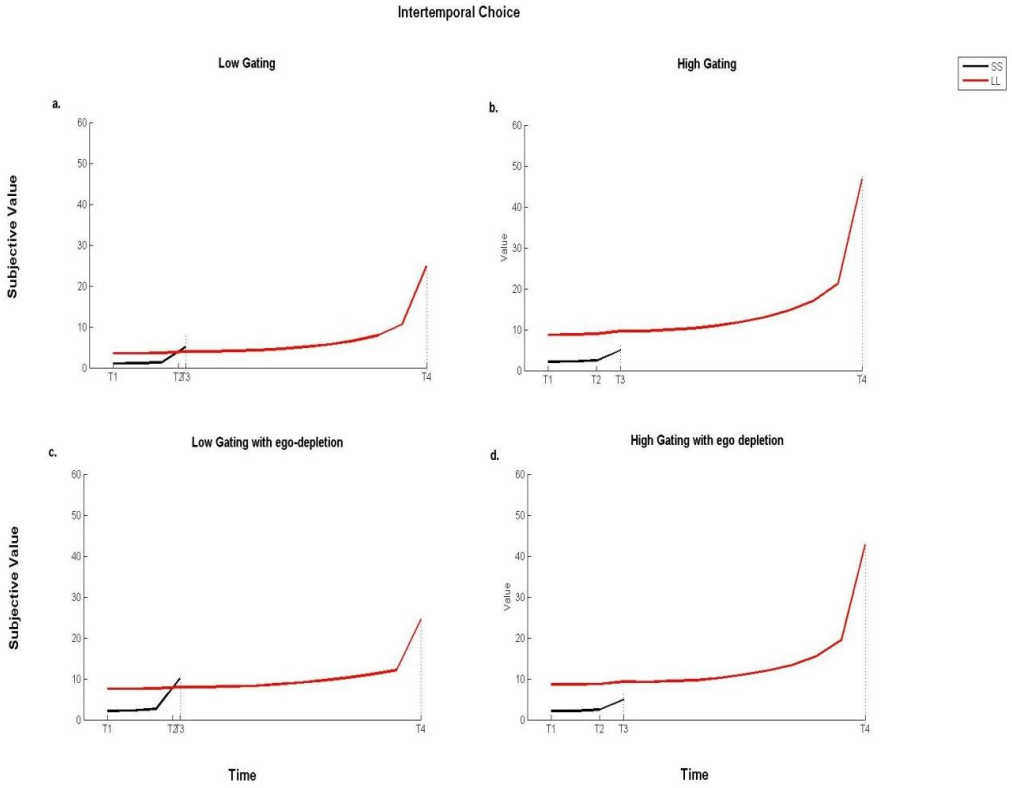

Figure 6: Simulation 2. Self-control in intertemporal choice as a function of gating variable. (6a) Low gating; (6b) High gating; (6c) Low gating with ego-depletion; (6d) High gating with ego-depletion. T1 is the initial time point, T2 is the point of indifference in curves (6a) and (6c) and T3 and T4 represent the time-points at which the SS and LL are delivered, respectively

3.2.3 Simulation 2.3: Self-control in intertemporal choice (weak gating with ego-depletion)

The major behavioral finding of quasi-hyperbolic discounting in intertemporal choice was replicated, corresponding to an absence of self-control. In Fig. 6c, the model is

shown to be tracking the evolution of value over time. At Time T1, the larger-later reward has a higher value than the smaller-sooner reward. As time approaches T2, the curves cross. At time T3, the model shows a preference reversal and chooses the smaller-sooner reward.

3.2.4 Simulation 2.4: Self-control in intertemporal choice (strong gating with ego-depletion)

Fig. 6d shows the evolution of value over time in intertemporal choice. Here the larger-later reward has a higher value from the start of the trial (at time T1), and this increases with decreasing delay to reward. The smaller-sooner reward increases in value but always remains below the curve corresponding to the larger-later reward's subjective value over time, even at the time point T4 that marks its delivery. The effect of ego-depletion reduces the difference between the two curves, although the strength of the gating preserves dynamic consistency in choice.

4. General discussion

We extended a general computational framework of learning-action tradeoffs to self-control in decision making. This complements the model of instruction following developed under the same framework (Ramamoorthy & Verguts, 2012). The self-control model replicated the food-choice data from Hare et al. (2009); a model with an additional time component replicated behaviour in intertemporal choice. Here we examine the theoretical framework in relation to other accounts of self-control. Further, we discuss inhibitory control and ego-depletion. We conclude with questions and directions for future research.

4.1 Theoretical considerations

Theoretical accounts of temporal discounting have led to several kinds of models; exponential discounting, hyperbolic discounting and quasi-hyperbolic discounting. Quasi-hyperbolic discounting models have recently been used in the interpretation of

neural correlates of temporal discounting (McClure et al, 2007). McClure et al. (2007) discuss the β-δ model where overall value is a weighted sum of the contributions of the two exponentially discounted value components computed by two valuation systems, the β-system and the δ-system, with slightly different exponential discounting factors.

Recently, Scherbaum, Dschemuchadse and Goschke (2012) have proposed a connectionist model of temporal discounting. This model employs an additive valuation process that integrates value and time information to compute relative values of different options. Representations of the discounted options compete through lateral inhibition, and the winning option determines the choice. Our model shares these aspects with the Scherbaum et al. model, but differs in a few crucial respects such as prefrontal influence (gating) over value computation in the VMPFC and the representations of delay using tapped-delay-lines. Most importantly, our model focuses on self-control and how it is implemented in biologically specified interactions between brain areas, thus providing a computational base for future tests and manipulations that may modulate self-control.

4.2 Inhibition and self-control

The ability to inhibit oneself is considered vital to adaptive behaviour. Inhibitory control has been extensively studied at the behavioural and neural levels.

At face value, the importance of inhibition to self-control appears obvious. Explorations of the neural bases of self-control tend to identify regions associated with inhibitory control, such as the right inferior frontal gyrus (rIFG) (Aron, 2008) and some conceptual accounts of self-control have proposed inhibition as a putative mechanism (Jasinska, Ramamoorthy & Crew, 2011). A recently proposed theoretical framework suggests that inhibitory control can be of two types (Munakata et al. 2011). Direct inhibition is associated with rIFG, and leads to a global shutdown of processing in the target region (i.e. to stop a response). It may be achieved by rIFG representations that project excitatory connections to GABAergic interneurons which then inhibit neurons in the target region. The other means of achieving inhibition is by

biasing relevant areas in a competitive, winner-takes-all mechanism. The losing representation and the behaviour it would have engendered are thus indirectly inhibited. Our model is an instance of the latter type. Self-control in our model emerges through the interplay between LPFC representations, gating of LPFC and NAcc inputs to the VMPFC and competition between values within the VMPFC, such that the winning value-representation laterally inhibits the competing value-representations.

4.3 Ego-depletion

As mentioned earlier, ego-depletion is a key construct in the self-control literature (Baumeister, 1998). It refers to the progressive weakening of the ability to exert self-control, when the agent is called upon to exert such control repeatedly. In our model self-control is a consequence of strong LPFC gating of value computation in the VMPFC. We incorporated ego-depletion into this account by allowing the gating variable to decay with time. As we view deliberate processing as akin to tree-search in Daw et al. (2005), repeated use of this mechanism would be exhausting, given the limited time for such searches to be performed. The learning/acting tradeoff could then account for ego-depletion, as a consequence of the computational cost of tree-search as opposed to caching (Daw et al., 2005).

Interestingly, ego-depletion is counteracted by motivation as well as beliefs, particularly beliefs concerning willpower (Job, Dweck & Walton, 2010). This effect, however is contingent upon the extent of depletion (Vohs, Baumeister & Schmeichel, 2011). Severe depletion is not reversed by beliefs and motivation. Within our framework, the effect of motivation would correspond to a boost to the prefrontal learning system.

4.4 Self-control as rule-following

In our earlier computational work (Ramamoorthy & Verguts, 2012), we speculated on the influence of the prefrontal-striatal instruction-following machinery on various

phenomena (e.g., suggestions). Given the conservation of theoretical principles across the models, interesting predictions emerge when the individual phenomena are considered in the light of the general theoretical framework. Instruction following influences response selection against actual contingencies (Doll et al., 2009), which means abstract representations counteract the effects of experience. In the context of self-control, value inputs of an immediate, appetitive or hedonic nature (such as taste), are overcome by abstract value inputs (such has health). This leads us to hypothesise that the extent of instruction-following observed in an individual would predict the extent of self-control they can exert under a given circumstance. We note that both instruction following and exertion of control can be mal-adaptive if they do not co-occur with flexible updating of policies. Strong rule-following in the absence of flexible updating of the rules or incorporation of contextual information could have potentially damaging consequences. In this case, the self-control exerted ceases to be regulatory and is a merely unilateral influence.

4.5 Directions for future research

Self-affirmation is effective in sustaining self-control in the face of ego-depletion (Schmeichel & Vohs, 2009). Self-affirmation refers to the activation of a core value, i.e., an abstract value associated with oneself. Affirmation of core values, rules of conduct and other such abstract beliefs, is widely recognized as a coping mechanism in diverse scenarios including combat (Asken, Christensen & Grossman, 2010). In the light of our framework, this would correspond to a lateral-prefrontal representation being strengthened. It is conceivable that this leads to heightened prefrontal processing overall and therefore has an effect on value computation, resulting in self-control in the face of temptation or fatigue. Consistent with this notion is the recent empirical finding that cue-based consideration of health-values leads to increased selection of healthy items in a food-choice task (Hare et al., 2011). Avenues for future research include integrating the instruction-following model with the self-control model to explore the mechanistic underpinnings of this phenomenon, behavioural investigation of the relation between instruction-following and self-control and simulating self-

control malfunction to examine implications for addiction, decision-making and other real-world phenomena.

CHAPTER 4

TWO SIDES OF THE SAME COIN? A BEHAVIOURAL STUDY ON INSTRUCTION-FOLLOWING AND SELF-CONTROL

Instruction following and self-control are potent aspects of human behaviour that influence our lives every day. Following recent empirical investigations into instruction following and self-control, we proposed computational models, derived from a general computational framework, to further theoretical understanding of these phenomena. A key feature of our models was the significance of the lateral prefrontal cortex (LPFC) to instruction following as well as self-control. An immediate prediction that emerges from the framework and its offshoots, is that instruction following and self-control must be related. More specifically, a strong tendency to follow instructions is expected to correlate significantly and positively with a high degree of self-control. To test this prediction, we conducted a study in which three tasks were used to compute an individual's tendency for instruction following, self-control and their working-memory capacity. The tasks used were, a dietary choice task (to measure self-control), a probabilistic selection task with an instructed component (to measure instruction following) and a working memory task (to control for the effect of working memory). Here we describe the study in detail, and report the results obtained. We found no significant correlations between the variables of interest. We discuss the results and what they might mean in the light of the general framework, and in turn, what they imply for the framework itself.

1. Introduction

Learning takes many forms and is ubiquitous in the biological world. Unique to human beings, however, is the use of verbal/symbolic information in learning. Given the finitude of human existence, this provides a tremendous advantage to humans in adapting to and shaping the environment they inhabit. Imagine having to learn every aspect of everyday life and human knowledge by trying to discover everything anew through trial-and-error. Such a task would be arduous, ambitious and would likely render the learner less adapted to surviving and thriving in a human environment. Of course, exploration lies at the root of all revolutionary insights into the nature of the world, but a majority of those who engage in such exploration do not strike gold. Chance, as Pasteur observed, favours the prepared mind and culturally/socially mediated learning contributes immensely to such preparation. A particularly influential instance of such learning is learning from instructions, or instruction following. Instruction following, once described as an unsolved mystery (Monsell, 1996), has been explored in recent times, both empirically (Ruge & Wolfensteller, 2010;Waszak et al., 2008) and computationally (Doll et al. 2009; Noelle & Cottrell, 1996; Ramamoorthy & Verguts, 2012).

Self-control, too, is all-pervasive in organisms and is essential to survival. Self-control can be defined as the ability to overcome prepotent, impulsive or habitual responses or evaluations to produce behaviours that are more compatible with longer-term benefits or goals. A good example would be the avoidance of unhealthy snacks which are consumed exclusively for their taste. While such avoidance would deprive the individual of momentary pleasures, it would promote long-term health. Indeed self-control as a human behaviour is very important to individuals as well as societies. Given its importance it has been studied extensively, from behavioural and neural perspectives (Baumeister, 1998, Hare et al. 2009, Hare et al. 2011, McClure et al. 2004, McClure et al. 2007).

Given the paucity of theoretical work on these phenomena, we developed a framework in order to integrate them conceptually and thus to understand them better. The framework was based on the concept of computational tradeoffs. Such tradeoffs are

well known in cognitive neuroscience, the most well-known being that between generalization and (avoidance of) interference, hypothesised to be handled by cortex and hippocampus, respectively (McClelland et al., 1995; O'Reilly & Norman, 2002). To understand instruction following, we looked at another tradeoff, one between fast learning and fast acting. This computational tradeoff pits learning speed against the speed of acting or responding. Rapid learning is typically associated with slow responding and rapid responding is associated with slow learning. We proposed a computational framework which instantiated this tradeoff in the form of complementary systems /processing pathways, capable of learning and producing responses to stimuli. The lateral prefrontal cortical (LPFC) system learnt fast but responded slowly, and the second system, situated in the basal ganglia (BG) (mostly the striatum) learnt slowly while responding rapidly. We extended this framework to instruction following and self-control respectively. In the instruction following model (Ramamoorthy & Verguts, 2012), instructions were learnt rapidly by the LPFC and implemented slowly, and in turn, learnt slowly by the BG and applied rapidly once learnt. This model replicated empirical findings such as the quickening of responses, the change in relative contributions of frontal and striatal regions (Ruge & Wolfensteller, 2010), interference between instructed and applied mappings (Waszak et al., 2008) and also suggested a possible mechanism behind goal neglect (Duncan, 1996).

Following the application of the framework in the study of instruction following, we extended it to self-control. Recent research suggests that self-control is realised through the influence of the LPFC on value computation in the ventro-medial prefrontal cortex (vmPFC) (Hare et al., 2009; 2011). We further hypothesised that the LPFC could influence (or gate) the overall value computation in the vmPFC by accentuating its contribution and attenuating the contribution of the other valuation system. The model successfully replicated the pattern of choice-behaviour reported for self-controllers and non-self-controllers (delineated computationally via a gating parameter) by Hare et al. (2009). Next, we used the model to capture self-control in intertemporal choice. The crux of intertemporal choice is the problem of rewards crossed with delays; when a larger reward can be reached only after a relatively long delay period, whilst a smaller reward is accessible more immediately. The former is

typically referred to as the larger-later reward and the latter, the smaller-sooner reward. If faced with a choice between a larger-later reward and a smaller-sooner one, a supposedly rational agent's preference would remain unchanged over time . Economists who assumed this (referred to as *exponential discounting*, where the difference between the subjective value of the two rewards remains constant as they depreciate with delay), were stymied by human behaviour. Humans (among other animals) appear to prefer the larger reward at first, but as time elapses and the smaller-sooner reward becomes more appealing due to its proximity, they switch preferences (*preference reversal*) and opt for the smaller-sooner reward (Ainslie, 1992). This is typically interpreted as a lack of self-control.

From our framework, we reasoned that the LPFC system (fast-learning, slow acting) could be thought of as simulating future reward value and consequently would be less vulnerable to delays in obtaining a desired good. By contrast, the BG system (slow-learning, fast acting), and the ventral-striatal reward-value processing regions in particular would simulate future value with a high sensitivity to delay. The model successfully captured both preference-reversal (Ainslie, 1992) and no-reversal (interpreted as self-control).

Recent research has suggested that self-control can be sustained through the use of self-affirmations (Schmeichel & Vohs, 2009). Self-affirmations are statements that promote positive beliefs or evaluations of the self or serve to link specific goals or desired states with the self (e.g. "I am calm"). We hypothesised that self-affirmations are instructions to oneself and therefore, self-control can be treated as an instance of instruction-following. This would imply that individuals with a marked tendency to follow instructions are also likely to be better at controlling themselves (i.e., self-control). From a cognitive neuroscience perspective, the importance of lateral prefrontal cortex (LPFC) to both instruction-following (Hartstra et al., 2012; Ruge et al. 2010) and self-control (Figner et al, 2010; Hare et al.,2009; 2011) appears to lend credence to this idea. As described previously, the LPFC pathway in our models implements both instruction following and self-control. An important prediction issuing forth from the framework is that instruction-following and self-control may be functionally related. More specifically, we hypothesize that a strong tendency to

follow instructions would correlate with a high degree of self-control. In other words, those who are good at instruction following should also be good at controlling themselves in the face of temptation. As a first step towards examining this prediction, we designed a behavioural study consisting of three components; a probabilistic selection task with an instructional component to measure the tendency to follow instructions, a dietary choice task to measure self-control, and a working memory task to control for working memory effects.

2. Methods

The study consisted of three behavioural experiments completed sequentially. A dietary choice task adapted from Hare et al. (2009), a probabilistic selection task adapted from Doll et al. (2009) and a working memory task sourced from the WAIS IV (2008). Ethical approval for the study was obtained from the local Ethics committee (faculty of Psychology and Educational Sciences).

Subjects were screened for neuro/psychopharmacological drug use through a standard screening checklist. Only those without such a history or ongoing prescription were invited to participate in the study. All subjects had to abstain from consuming food for a period of three hours prior to the experimental session. This restriction was conveyed through written instruction during the screening process. Compliance to this instruction was tested in the questionnaire administered after the dietary choice task. Consumption of water was allowed.

The two tasks of interest, dietary choice and probabilistic selection were counterbalanced across subjects with all subjects completing the working memory task last.

Sixty-four subjects participated in the study after providing informed consent. Twelve subjects were excluded due to one of the following reasons: non-completion of task, interruptions during experiment session, non-adherence to experiment protocol (e.g. not abstaining from food 3 hrs prior to participation despite instructions provided in advance). The remaining 52 subjects (37 female, mean age 22.56 standard deviation (sd) 3.28) completed all three tasks successfully.

24 subjects completed the food choice task first and the remaining 28 completed the probabilistic selection task first. This was encoded as an order parameter and used as a between-subjects factor in the analysis of the probabilistic selection task.

2.1 Dietary choice task

Following Hare et al. (2009), we presented images of food items sequentially to the subjects in three blocks. We used 60 images downloaded from the internet for this purpose. The first two blocks were rating blocks. In the "Taste" rating block, subjects rated the tastiness of each item on a Likert scale ranging from 1 to 5 (where 1 = "not very tasty", 2 = "not tasty", 3 ="neutral", 4 = "tasty" and 5 = "very tasty"). In the "Health" rating block, subjects rated the healthiness of each item on a Likert scale ranging from 1 to 5 (where 1 = "very unhealthy", 2 = "unhealthy", 3 = "neutral",4 = "healthy" and 5 = "very healthy"). The order of these two rating blocks was counterbalanced across subjects.

Once the ratings had been collected, a reference item unique to the subject was chosen from the items that had been given a neutral rating (i.e., rating 3) on both scales. In the absence of such an item (with taste and health ratings of "3" each), we selected an item rated neutrally on Taste and slightly positively on Health (i.e., a rating for "4") as the reference (see Hare et al., 2009, supplementary materials). The reference image was then presented on screen until the subject chose to move to the third and final block.

The third block was a decision block. In this block, all images excluding the reference (therefore a set of 59 items) were presented sequentially and in random order. On each trial, the subjects had to make a simple binary choice; would they like to eat the item on the screen or the reference (shown before the decision block) at that moment? Subjects were told to choose naturally and that one of the choices they made would be implemented at the end of the experiment (cf. Hare et al., 2009). Choices were recorded and used in the analysis. The design is represented schematically in Figure 1.

After the task, subjects answered the following questions with yes/no responses:

1. Are you vegetarian/vegan?
2. Are you allergic to any of the food items presented in the task?
3. Did you consume any food during the 3 hours before the experiment?
4. Are you currently on a diet?

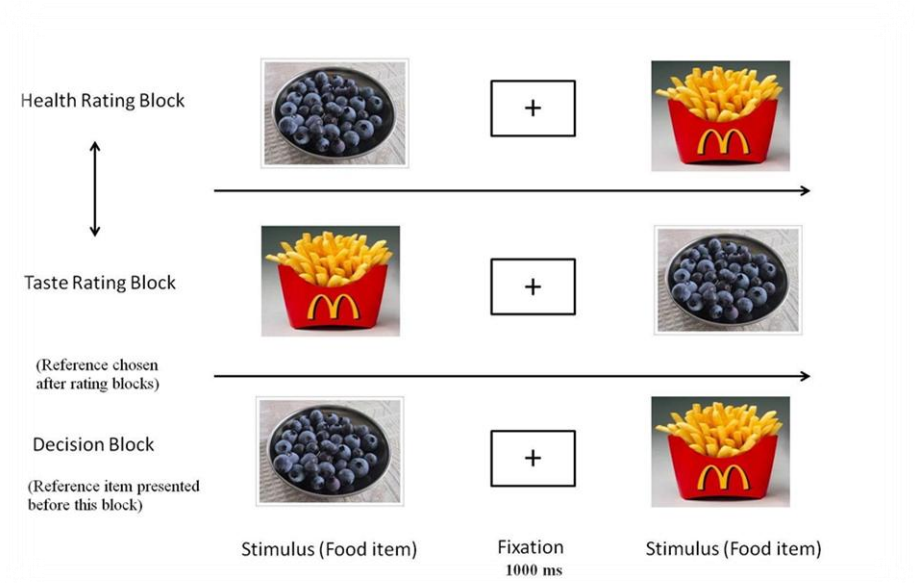

Figure 1: The design of the dietary-choice task. In the rating blocks , subjects rated each item presented on the screen on a 5-point Likert scale assessing one of two dimensions (taste, health) specific to the block, (i.e., all items were rated on the taste dimension in the taste-rating block). Following the ratings, a reference item was chosen (from the subset of items rated neutrally on both health and taste in previous blocks) and presented before the decision block. In the decision block, subjects had to choose between the reference and the item currently being shown on screen. In all blocks, the stimuli were presented until a response was made, to facilitate natural responses on the part of the subject.

Using the choice data from the decision block, we computed a self-control measure using the following equation:

$$S = \sum_{i=1}^{n} (y_i - x_i) v_{Choice_i} \quad (1)$$

where S is the self-control measure, y_i is the health rating of the i^{th} item (from a set of n trials in the decision block) and x_i is the taste rating of the i^{th} item and $v_{Choicei}$ indicates whether the ith item was chosen over the reference or not. This computation resulted in a self-control score that was a continuous variable.

Here and in the other tasks, we examined split-test correlations to assess reliability of the data. Across subjects, we computed the variable of interest twice, from the odd-numbered and even-numbered trials to obtain trial-parity-specific scores. We then examined the correlations between odd and even scores between subjects.

2.2 Probabilistic selection task

Following Frank et al. (2004), and Doll et al. (2009), we designed a probabilistic selection task to test the effect of instructions on reinforcement learning.

The stimuli were 6 Hiragana characters, labeled A to F (see Figure 2). During training, only pairs AB, CD and EF were presented. Each item was associated with a probability of being rewarded i.e., reward contingency). The contingencies associated with the stimuli were the same for all subjects. The Stimulus A was determined to be rewarding on 80% of the trials. Its counterpart B was rewarded only on 20% of the AB trials. C and E were set at 70% with D and F being rewarded only on 30% of the respective trials. This marks a departure from the design used by Doll et al. (2009) where they used a different reward distribution for CD (70/30) and EF (60/40) pairs.

We used the same reward probabilities for CD and EF to use CD as a control condition for pair EF. Due to an error in the program used for the probabilistic selection task, 24 of the 52 subjects were exposed to a slightly different set of contingencies for the uninstructed pairs in the training blocks. They were also exposed to one additional CD pair in the test block in place of a BD pair, but these combinations were not relevant to the calculation of the instruction following measure. In particular, the contingencies were (76.19/23.81) for AB and (68.42/32.58) for CD. The effect of this change is checked in the analysis (see below, factor contingency type). The remaining 28 subjects experienced contingencies (AB (80/20), CD (70/30), EF (70/30)) and test block pairings strictly as per the design described above.

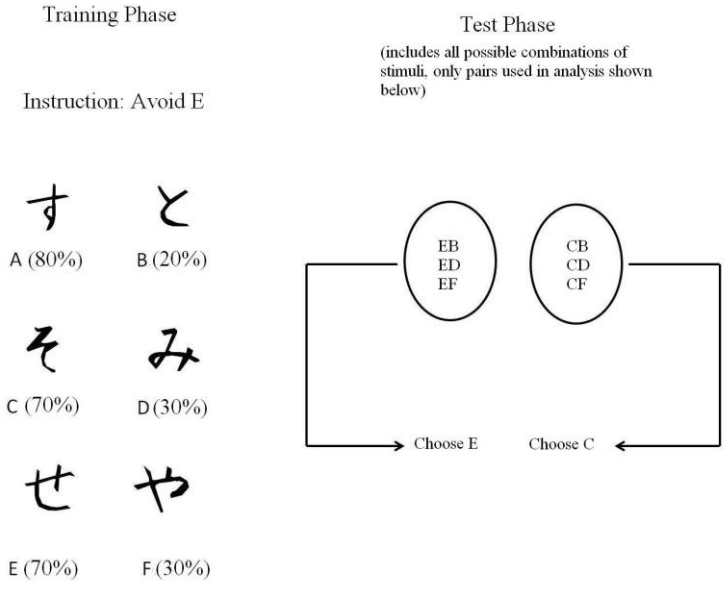

Figure 2: The design of the probabilistic selection task. In the training phase, subjects were exposed to pairs of stimuli (Hiragana characters labeled A-F for convenience here). The stimuli were rewarded probabilistically (reward contingencies shown in percentages next to each stimulus). The pairs were AB, CD and EF. During the training phase, only these pairs were presented. Hiragana characters were chosen to

minimize verbal encoding as in Doll et al. (2009). Prior to training, subjects were told that "E" would be an inferior stimulus and that they should try to avoid it. Following instruction, each trial presented a randomly selected pair (from AB, CD and EF) and the subjects had to choose one of the stimuli from it. Feedback indicating whether the stimulus was rewarded ("Correct") or not ("Incorrect") was provided after each trial and the distribution of feedback reflected the contingencies associated with the stimuli. Subjects trained until they achieved a certain level of performance (see probabilistic selection task in the Methods section for details). Following this, they had to undergo a test phase where the stimuli were presented in old as well as novel pairs, exhausting the possible binary combinations available with 6 stimuli. Responses to stimuli with 70% reward contingency, i.e., C and E were of particular interest, as one of the two (E) had been devalued by prior instruction. The figure depicts pairs of interest for this instructed stimulus and indicates optimal behaviour for each type of pairing (deviation from this would imply they were influenced by the instruction to avoid E).

The task consisted of a training phase and a test phase. In the training phase, each stimulus pair was presented a fixed number of times within a block. In this design, there were 60 trials in a training block, with each pair occurring 20 times. Stimulus presentation was randomized within blocks. Each trial was preceded by a fixation cross and followed by feedback as to whether the symbol chosen was correct or not. Following Doll et al. (2009) we imposed a performance criterion to ensure learning of the uninstructed contingencies (with EF being the instructed pair; see below for explanation). This criterion was the simultaneous (as in within one training block) attainment of (at least) 65% A choices on AB and (at least) 60% C choices on CD training pairs. To minimize the risk of some subjects learning only one of the pairs and passing the criterion by chance selection of correct symbols on the other pair, we imposed a minimum of two training blocks per subject, with the assessment being performed only from the second block onwards. If a subject failed to achieve the requisite level of performance, they had to undergo further training until they did.

Performance was always measured within a given training block and not across blocks.

Upon successful passing of the criterion, subjects completed a single-block test-phase in which all the symbols appeared in all possible combinations (e.g., AB, AC, AD, etc.). The test phase comprised of 90 trials without feedback.

Prior to the training phase, the experimenter provided detailed verbal instructions to the subject as follows:

"This is a probabilistic selection task. It has two parts – a training phase and a test phase. In this task, on each trial you will be presented with two symbols on the screen. You have to choose one of them. Following this you will be informed whether your choice was correct or not. As this is a probabilistic task, no symbol will be correct 100% of the time. There is no absolute correct answer. Some symbols are likely to be correct more often than others. Your task is to guess the better symbol in any given pair."

After this the subjects were asked to read the instructions presented on the computer screen very carefully. The same set of general instructions were presented in Dutch to reiterate the basic concept of the probabilistic selection task. This was followed by a screen that displayed the following misleading instruction (in Dutch):

"The symbol shown below is least likely to be correct. You should avoid selecting it" and the Hiragana character for the "E" symbol was provided below. This was followed by the following instruction:

"To choose the symbol on the left, press 1 and to choose the symbol on the right, press 0".

Subjects read through the instructions at their own pace and commenced training afterwards. Upon completion of the training phase the subjects received an instruction that they would proceed to the test phase in which they will be exposed to all possible pairs of the stimuli they were trained on (training phase pairs interleaved with novel pairs). They were informed that the test phase would not include feedback following each trial and that they would have to go with their intuition. This was followed by the

test phase. Each pair appeared 6 times within the test block, and the occurrence of any given pair was randomized across the block. No feedback was provided in this block.

Following successful completion of the task, subjects were asked to provide feedback on the following questions:

1. Please provide general feedback on the experiment
2. Did you notice anything in particular about the experiment?
3. Did you follow all instructions given to you?
4. Did you notice anything in particular about the instructions?

An instruction-following measure was computed by calculating the difference (C-E) between mean C choices and E choices in conditions where they were paired with statistically inferior stimuli (i.e., B, D, and F) across the test block of the probabilistic selection task. The rationale behind this was that the instruction to avoid E would result in a positive difference between C choices and E choices on these trials, where both C and E are clearly the better stimuli.

2.3 Working memory task

The working memory task (taken from WAIS IV, 2008) consisted of subjects being asked to report sequences of letters and numbers presented to them on a trial. There were two blocks in this task. In the first block, the subject saw a sequence of letters and numbers appearing one element at a time in the middle of the screen. They then had to reproduce (i.e., type in) the elements, with the numbers being typed in first and the letters next. The numbers had to be in increasing order and the letters, in alphabetical order. The sequences increased in complexity over trials (2 items in the first trial and 8 in the final trial). There was a short training phase consisting of 5 trials prior to the commencement of the experiment. This was followed by 21 trials. The second block consisted of 16 trials in which subjects had to remember and reproduce the exact sequence in which numbers had appeared on screen during stimulus presentation, without reordering them. The stimuli were presented as in the first block. A working memory score was calculated by dividing the number of correct responses, up until the first erroneous one, by the total number of trials in the block.

For the purpose of our study, we used the score from the first block to compute a measure of working memory. The second block was not used in the analysis due to its focus on short-term memory span (cf. Daneman & Carpenter, 1980).

3. Results

3.1 Dietary choice task

Taste and health ratings for each item were averaged across subjects (N= 52) to obtain a colormap (Figure 3) indicating the rating distribution of the items. Most items were found to be distributed around a taste value of 3 and a health value of 4. Figure 4 presents the distribution of self-control scores across the sample (equation 1). Most subjects received negative self-control scores (mean = -22.11, sd = 24.07). 19.2% of the subjects self-identified as vegetarians. 11.5% reported being on a diet. 2 of the 52 subjects reported having consumed food within the 3 hours prior to the study. The reliability of the task was, $r(50) = 0.605$, ($p < 0.001$).

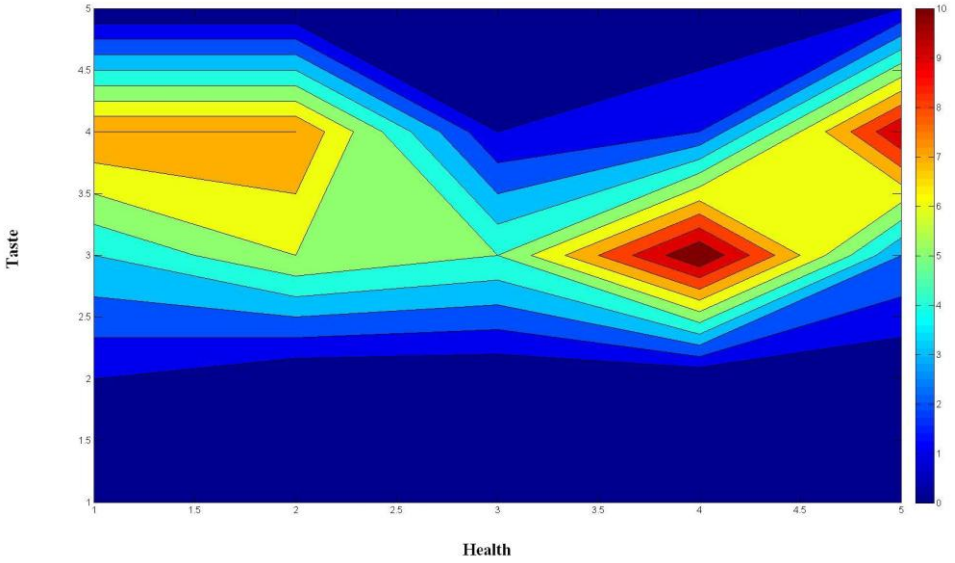

Figure 3: Colormap generated from health and taste ratings assigned to different food items by subjects in the dietary choice task. This shows the distribution of rating combinations (taste, health; each rated on a 5-point Likert scale) for the 60 items rated

.

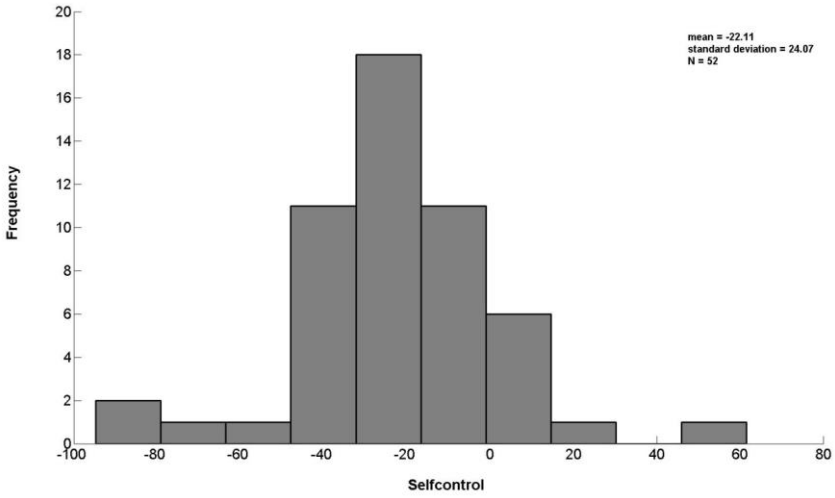

Figure 4: Self-control scores from dietary choice task

3.2 Probabilistic selection task

Our dependent variable of interest in the probabilistic selection task was percent choice of the 70%-stimulus, with independent variables stimulus (C or E, control or instructed) and block (training block 1, training block 2, or test). Given the error in contingencies in one version of the task we used contingency type as a between-subjects factor and tested for interactions with variables of interest .

Subjects completed the training phase before proceeding to the test phase. Subjects completed a minimum of 2 training blocks prior to test. The number of training blocks needed by subjects to reach performance criterion ranged from 2 to 5 (mean= 2.289, sd = 0.637). Due to the fact that only 21.15 % of the 52 subjects required more than 2 training blocks, we used only the first two blocks from each subject's data for

analysis. The reliability of the test (measured as indicated in methods section) was equal to ($r(50) = 0.834$, $p < 0.01$).

To measure the effect of instructions on EF pair relative to the uninstructed pairs, we obtained choice and accuracy information for each pair in the training blocks. Figure 5 provides a comparison of choice behaviour on A, C and E across training and test blocks.

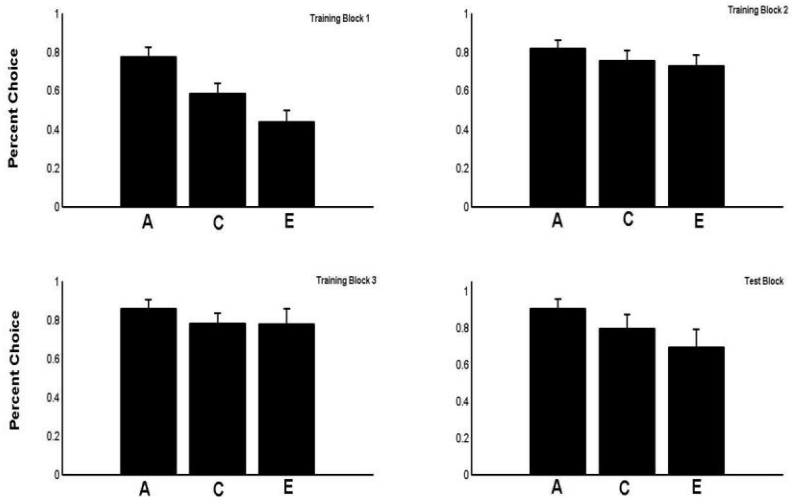

Figure 5: Percent choice on stimuli A, C, and E across blocks in probabilistic selection task.

In the training phase, subjects chose A (in the AB pair) often throughout training (77.30% of the time (sd = 19.31%) in the first training block; and 82.11% of the time (sd = 15.67%) in the second training block). Choice of C in the CD pair was above chance in the first training block (58.36% of the time, sd = 19.91%) and was close to the actual contingency in the second training block (75.19% of the time, sd: 19.60%). Subjects' choice of E in the EF pair (which had the same reward probability as C in CD, 70%) was suboptimal in the first training block (44.03% of the time, sd: 22.05%) but improved in the second block (72.69 % of the time, sd: 22.19%). These results show that by the end of the training phase subjects had learnt to select the better stimulus in each training pair. They are summarised in Figure 5. An ANOVA was performed with within-subjects variables stimulus, C or E; and block (training block 1, training block 2, test) and with order (Dietary choice task first or Probabilistic selection task first), contingency type, and the answer to the feedback query (whether the subjects followed all instructions (yes/no/no useful answer)), as between-subjects factors. This revealed a significant effect of block (Training block 1/ Training block 2 / test) (Greenhouse-Geisser corrected, $F(1.775, 76.317) = 14.306$, $p < 0.01$), meaning that the optimal stimulus was chosen more frequently as the task progressed. In other words, subjects learned from the experimental contingencies. There was no significant effect of stimulus ($F(1, 43) = 0.338$, $p = 0.564$), but the interaction between block and stimulus was significant ($F(1.726, 74.232) = 3.477$, $p = 0.043$). Hence, despite being nominally in the predicted direction (see percentages above and Figure 5), the instruction manipulation did not have a robust effect on subjects' performance. The between-subject factors did not have any significant interactions with stimulus, (contingency type: $F(1, 43) = 0.898$, $p = 0.349$; order: $F(1, 43) = 0.949$, $p = 0.335$; feedback query: $F(2, 43) = 1.725$, $p = 0.190$) and block (contingency type: $F(1.775, 76.317) = 0.867$, $p = 0.413$; order: $F(1.775, 76.317) = 2.391$, $p = 0.104$; feedback query: $F(1.775, 76.317) = 0.868$, $p = 0.477$). Among the between-subjects factors, contingency type ($F(1, 43) = 5.130$, $p = 0.029$) and feedback query ($F(2, 43) = 5.339$, $p = 0.008$) had significant main effects, unlike order, which had no significant effect ($F(1, 43) = 0.5$, $p = 0.483$).

In the post-task questionnaire, the question "did you follow all instructions?" received a positive answer ("yes") from 76.9 % of the subjects, with 19.2% replying "no" and

the remaining 3.8 % providing unclear feedback. The question "did you notice anything in particular about the instructions?" received replies in the negative from 55.8% of the subjects. Of the remaining 44.2%, 30.8% reported that the instruction to avoid E was wrong with regard to the contingencies, and the remaining 13.4% provided feedback that could not be quantified within the binary encoding used. The feedback question "did you notice anything in particular about the instructions?" showed a weak correlation with instruction-following, tending towards significance ($r(50) = -0.238$, $p = 0.089$). The feedback question "did you follow all instructions" was not correlated with instruction-following as measured in the main task ($r(50) = 0.211$, $p = 0.133$). This latter query was however negatively and significantly correlated with the choice of E in the test phase ($r(50) = -0.309$, $p = 0.026$), meaning that adherence to all the instructions resulted in a reduction in choice of the E stimulus in the test phase, which is consistent with the fact that the stimulus-specific instruction was to avoid selecting E.

3.3 Working memory task

Working memory scores were calculated by dividing the number of correct responses up to the point of the first error, by the total number of trials in the block. We computed this score using data from the first block of the working memory task (see above). Working memory scores ranged from 0.048 to 0.952 (mean = 0.509, sd = 0.235). The reliability of the task was, $r(50) = 0.979$ ($p < 0.001$).

3.4 Correlations across tasks

Contrary to our prediction, self-control was not significantly correlated with the instruction following measure ($r(50) = 0.165$, $p = 0.242$). Likewise, the correlations between working memory and self-control ($r(50) = -0.162$, $p = 0.251$), and working memory and instruction following ($r(50) = -0.175$, $p = 0.216$) were not significant. The statistical power of these correlations given the sample size (N= 52) was lower than 35% (32% for instruction following versus self-control, 31.22% for working

memory versus self-control and 34.64% for working memory versus instruction following, respectively).

4. General discussion

In this study, we aimed to test whether an increased tendency towards instruction-following predicted a high degree of self-control. We used tasks designed to elicit instruction-following tendency and self-control within the context of dietary choice to this end. A working memory task was used a control task. The study yielded no clear evidence in favour of the prediction derived from the models. Here we detail possible interpretations of the results and discuss avenues for future research.

4.1 Dietary choice task

The sample contained a majority of non-self-controllers (in the context of dietary choice). This differs from the distribution of subjects in the Hare et al. (2009) study. There were few dieters in the sample (11.5%) and it is conceivable that this contributed in part to the observed distribution of self-control scores.

4.2 Probabilistic selection task

The probabilistic selection task did not yield a marked effect of instructions on contingency-learning. This could be due to a number of factors. Firstly, the instructed (EF) and control (CD) pairs shared a contingency that was easy to disambiguate (70/30), as opposed to the one used by Doll et al. (2009), (where EF pair had a reward contingency of 60/40). Our reason to equalize these percentages was to make sure that a proper control stimulus (C) would be available for the manipulated (E) stimulus. Yet, as a result of this change our experiment does not serve as a strict replication of the earlier work by Doll et al. (2009). Secondly, a focus on the instructed symbol (E) coupled with clear-cut contingencies could have led to abandonment of the instructed policy (avoid E). Thirdly, given that we imposed a minimum of two training blocks (i.e., subjects could not proceed to the test phase even if they had learnt the

uninstructed contingencies within one block), it is conceivable that this could have been sufficient for subjects to learn that the "avoid E" instruction was inconsistent with the actual reward contingencies associated with the EF pair.

4.3 Correlations across tasks

Our prediction that high self-control scores would correlate positively with instruction-following behaviour was not supported by the evidence. Reliability of the individual tasks was probably not the reason, as reliabilities were generally quite high. In addition to the possible factors that led to the observed task-specific results, the absence of an interaction between the variables of interest could be due to multiple factors.

First, given the computational framework described previously, these results could also be interpreted to mean that the subjects were skilled at contingency-learning and responded to immediate rewards (tasty food items). Second, as it stands, the sample comprises of subjects with low instruction-following tendency and low self-control. It is likely that there were floor effects in the variables of interest. Third, subjects performed all three tasks (probabilistic selection, dietary choice and working memory) in one session, and did not consume food for 3 hours prior to it. This requirement is pertinent only to the dietary choice task. It is conceivable that this might have presented the subjects with increased cognitive demand, thereby introducing an additional, uncontrolled, factor influencing the results.

Null results are akin to an ambiguous silence punctuating a spirited conversation. The results yielded by the experiment do not categorically invalidate our prediction but they also do not favour it. Given the success of the modelling framework in replicating recent datasets dealing with both instruction-following and self-control, we opine that the experiment described in this study needs to be repeated with one or more of the following modifications. First, the probabilistic selection task might benefit from a modification of the contingencies of the two stimuli of interest; C and E. Instead of the easy to disambiguate reward contingency, (70/30) a finer one (60/40) can be used for both control (CD) and instructed (EF) stimulus pairs. We predict that this would

render the effect of instruction more pronounced. Second, testing the same group of subjects on each task separately at different time periods or on different days could be a way to circumvent any unintended influence of cognitive demand, on overall performance. Third, the sample should be larger to increase power. Indeed, the correlation was nominally in the right direction, so it is a possibility that low power (< 35%) contributed to the null result, given the sample size. And finally, a more balanced sample (including dieters in greater number) could yield a more balanced distribution of self-control scores and increase the sensitivity of our tests.

CHAPTER 5

GENERAL DISCUSSION

Human behaviour, be it the utterly arbitrary or the most arcane, is the product of interactions that span multiple levels of description, from the biophysics of the action potential to coordinated information processing in brain systems. In this thesis, we investigated the computational underpinnings of two very important facets of human behaviour; learning and cognitive control, and explored the predictions derived from the computational approach to these phenomena, empirically. Here, we discuss the two aspects of this program of research (i.e., theoretical and empirical) in detail and systematically outline pertinent ideas for future research.

5.1 The Computational theory: an overview

David Marr (1982) proposed a three-level framework for understanding information processing systems; the computational ("what is the computational problem being solved?"), the algorithmic ("what are the ways in which the problem can be solved") and the physical (implementation) ("how does a physical system support the process described by the preceding levels?") (Poggio, 1981). Originally applied to the problem of vision, this framework has a well-deserved generality and can be applied to the study of those facets of behaviour that we seek to understand as well. Our approach covered the computational, algorithmic and high-level mechanism levels of analysis. We used a broadly defined theoretical framework to address open questions in the study of learning and cognitive control.

Following Daw et al. (2005), our framework consisted of complementary learning/control systems that instantiated a learning-action tradeoff. One of the systems learnt rapidly while being slow to respond, whereas the other required more time to learn while responding with rapidity once said learning had occurred.

5.1.1 Models and Mechanisms

We derived two distinct (but related) models from the framework described above; a model of instruction following and a model of self-control. Here we discuss the

individual models and their relation to other models aimed at capturing the same phenomena.

5.1.1.1 Model of Instruction following

The model described in chapter 2 was a straightforward instance of the framework in that it had dual-systems architecture carrying out learning and responding across the computational tradeoff of interest (rapid-learning/slow-acting versus slow-learning/fast acting). The LPFC system (fast-learning/slow-acting) acquired instructions (encoded as instructed S-R mappings in the model, to capture a simple, but sufficiently rich form of instruction following) through fast-Hebbian learning and responded slowly. While this is a simplification, it can also be seen as an instance of model-based learning (as described in Daw et al. 2005), where the system acquires a model of the world (in this case, a simple rule for responding to specific stimuli or categories of stimuli) and uses it to constrain its responses. Typically, (world) models (and their learning algorithms) are more complex, but in their simplest form they can be reduced to straightforward associations between required stimuli and required responses. See Gläscher et al. (2010) for an example how such more complex models may be learnt. The second, slow-learning system, on the other hand, is model-free, in that it learns by integrating contingencies over time but acquires "cached" responses that can be implemented rapidly. In the context of instruction following, the "cached" learning system (hypothesised to be predominantly striatal), acquires its mappings by picking up reinforced responses delivered initially by the LPFC system. This is an instance of cooperative learning between the different systems along a tradeoff. Competitive interactions could also occur, for example, when the rule learnt by the instruction-following system is no longer applicable. The flow of information can also go the other way round: The "cached" learning system integrates novel contingencies and could then contribute to the extraction of a new rule (Pasupathy & Miller, 2005) which would then lead to updating of the previous "(world)model".

It is interesting to note that this competition/conflict between the two learning systems can also be cast in the light of reward prediction-error driven learning in the brain (Gläscher, Daw, Dayan & O'Doherty, 2010; Silvetti & Verguts, 2012). The "model"

held by the LPFC system (or the instruction following system) is associated with a response to an external circumstance. This amounts to a prediction of a specific outcome given a specific response. If the rule is no longer valid, then the model-based response would lead to an outcome that would fall short of the expected reward and this would result in a prediction error. Given that the brain appears to perform optimization, which corresponds to the computational level in Marr's sense, it would try to minimize this error by updating its representation of the world and the rules used to respond in the world (for a discussion of this idea see Friston, 2010). When this fails to occur, mal-adaptive behaviours and/or cognitive biases can be observed.

5.1.1.2 Model of self-control

The model discussed in chapter 3 is derived from the same framework as the model of instruction following, but its correspondence to the framework is less direct. This is because, as it stands, the model of self-control takes value information of different kinds and integrates them at a specific site (the vmPFC) to compute a decision. The learning/action tradeoff is implicit in its architecture and function. The two learning systems are very much prevalent in this model. The LPFC valuation system which influences overall value computation and is thus seen as the seat of self-control, is yet another instance of a fast-learning, slow-acting, model-based system. Instructions can be conceived of as very simple models of the world (and complexes of such rules generate richer, more comprehensive models). Likewise, abstract values such as "health", "savings" etc., can be seen as being part of the model-based system. One has a notion of becoming healthier over time due to application of specific rules ("eat more vegetables", "avoid chocolate fudge ice-cream for breakfast") but the results are not immediately apparent. Therefore, such rules belong to a model of the world, of things to be done to achieve specific goals (better health, retirement money). The model-free element here is the ventral-striatal valuation system which learns value based on immediate reward and reward prediction errors (Gläscher et al; for a different perspective on the role of ventral-striatum in reinforcement learning, see McDannald, Lucantonio, Burke, Niv & Schoenbaum, 2011). The gustatory pleasure associated with a tempting snack is not experienced in some future state, as in if one persistently

consumes chocolate cake, one would then experience a very satisfying taste. It is an immediate one. Therefore, when it comes to "wanting" a certain reward, be it dietary or monetary, the ventral-striatal system performs similarly to the "cached" system in the previous model, by responding quickly to things that have been learnt to be rewarding immediately. The LPFC system, on the other hand, can be thought of as estimating value by simulating the future and selecting the response in accordance with rules or goals ("eat healthy", "spend wisely") and it is essentially akin to the instruction-based system seen previously. A self-control goal can be learnt or acquired rapidly (e.g. New Year's resolutions) but its implementation is not immediate or easy (Gollwitzer & Schaal, 1998; Koestner, Lekes, Powers & Chicoine, 2002) . If this account reflects something that is going on in the brain, then the following scenario must be demonstrable upon investigation. Just as the instruction-following system imparts instructed contingencies to the striatal system, the self-controlling, LPFC valuation system should, in principle, be able to "teach" the ventral-striatal valuation system to assign greater reward value to choices that are more in-line with the long-term goals. For example a self-controller's striatal valuation system would come to view healthier options (broccoli) as being tasty if not tastier, than more immediately rewarding but unhealthy food items (fried doughnuts). Such a change in evaluation should be discernible on the behavioural and neural levels. Indeed, *caching* of previously rule-based responses improves the likelihood of attaining personal goals requiring the exercise of self-control (Gollwitzer & Schaal, 1998).

5.1.1.3 Other models of instruction following and self-control

Instruction following As detailed in chapter 2 instruction following has received much attention in recent years, albeit from the empirical level of investigation. Theoretical accounts of instruction following were rare with no major modelling efforts directed at acquisition and implementation of instructions until the pioneering work of Noelle and Cottrell (1995).

Noelle and Cottrell's model employed a recurrent network architecture (Elman 1990) to capture instruction following. Their network acquired an *instructional language* over a long training phase, using the error-backpropagation learning algorithm

(Rumelheart, Hinton & Williams, 1986). Instructions were represented as activation patterns (see Botvinick & Plaut, 2006) in the network as opposed to synaptic changes. This model is a fine example of connectionist modelling of cognition, but is not grounded in neurobiological considerations and is therefore not a learning account situated in the human brain *per se*.

Ashby et al. (1998) proposed the influential COVIS model of category learning, in which category learning is realised in distinct verbal (rule-based) and implicit (procedural) learning systems (for a review, see Ashby, 2011). As noted in chapter 2, this modelling approach is compatible with our general framework and COVIS can be seen as a predecessor of our model of instruction following. However, COVIS does not focus on the acquisition of an instructed rule, and is therefore not a model of instruction following in the same sense as the one described here.

More recently, Doll et al. (2009) proposed a model of instructional control of reinforcement learning. This model shares an important feature with our model, in that instructions are learnt by fast Hebbian learning in the prefrontal cortex. A point of departure is the purported influence of the instruction following system on the striatal learning system. In our model (Ramamoorthy & Verguts, 2012), this influence is cooperative, with one system teaching the other. In Doll et al.'s account, the instructions manage to bias the contingency learning system in the basal ganglia, to the detriment of the cognitive agent. They also report a variant of their model similar to ours (with prefrontal cortex influencing the motor system directly), but they dismiss this model in favor of the other one.

More recently, Huang, Hazy, Herd and O'Reilly (2013) have proposed a model of instruction following consisting of complementary learning systems that allow rapid instruction-learning and slow automatisation, in a manner that is conceptually identical to our model of instruction following. The differences between the two models are architectural. In their model, instruction following is achieved by fast-learning in a fronto-hippocampal system, and the parietal learning system carries out slow learning. It is interesting to note that this idea of complementary learning pathways has been proposed as a means of capturing instruction following by multiple

investigators independently, suggesting a theoretical convergence that might have a meaningful correspondence to empirical reality.

Self-control Our model of self-control is, to the best of our knowledge, novel. However, phenomena such as temporal discounting and preference reversal have been explored using models derived from behavioural economics (Ainslie, 1992), and such models have been applied to fMRI data (McClure et al., 2007). McClure et al. (2007) apply the double-exponential discounting model (the so-called β-δ model) to intertemporal decision making in the brain and suggest that the brain recruits two distinct valuation systems for intertemporal choice; the reward-sensitive/delay-aversive β-system and the delay-tolerant δ-system. This model is very interesting and indeed, provides neurobiological inspiration for our model, but it does not provide a neurocomputational mechanism to account for intertemporal decision making. More recently, Scherbaum, Dschemuchadse and Goschke (2012), have proposed a connectionist account of intertemporal choice and discounting. Their model employs an additive valuation process that computes the relative values of competing options by integrating both reward magnitude and temporal information. The two discounted options compete and the winner inhibits the other option, thereby leading to a choice. Our model shares some similarities with this approach, but has a more specific neurobiological focus, and suggests that delay and magnitude are combined with varying levels of tolerance to delay in competing valuations systems (see chapter 3 for details).

Our model appears to be the first to bring phenomena such as self-control in dietary choice as well as self-control in intertemporal choice within the purview of a general computational framework.

5.2 The computational theory: Concluding remarks

The general computational framework described above derives from the dual-controller framework proposed by Daw et al. (2005) and is in some respects, a simpler version of the same. Constructs such as instruction following or self-control refer to complex phenomena: We attempted to provide a simplified account capable of

addressing these. Our models make contact with previous findings and offer testable predictions, but are by no means exhaustive in their capacity for describing instruction following and/or self-control. They are the sum of their strengths as well as shortcomings. Shortcomings, more than strengths, inform and influence the evolution of a theory, and the models spawned by it. Specific boundaries/limitations add to the scientific quality of theories and models, as they render them testable and therefore, refutable (Feynman, 1965, 1974; Popper, 1963). Here we summarize the strengths and weaknesses of the models, and in turn, the overarching framework and remark on the future of this line of research

5.2.1 Model of instruction following: strengths and shortcomings

Upon examining the strengths of the model of instruction following, the following facts become apparent. First, our model of instruction following (Ramamoorthy & Verguts, 2012), has been successful in replicating basic behavioural findings, such as progressive quickening of responses (Ruge & Wolfensteller, 2010), interference between instructed and applied mappings (Waszak et al., 2008) as well as some aspects of the neural dynamics of instruction following (Ruge & Wolfensteller, 2010), such as declining prefrontal contribution to responses over increased application of an instruction, and increased striatal contribution to responses over time. Second, it employs a Hebbian learning rule, which, as noted in chapter 1, is an elegant and biologically plausible learning mechanism. It recasts instruction following as associative learning. Third, the model is not burdened by more parameters than needed, and in fact, given its architecture, produces interesting behaviours with just one free parameter (learning rate). Other parameters such as response threshold, cascade rate etc. (see Chapter 2 for details) are standard parameters used in modelling neural networks, and are not unique to the model. Fourth it yields testable predictions; the dynamics and transfer of learning across the different processing routes in the model are empirically testable, as is the model's account of goal neglect (Duncan, 1996).

The instruction following model has its share of shortcomings. First, the proposed architecture does not support processes such as context representation, rule retrieval from long-term memory and is as such a relatively partial account of

instruction following. Second, it presumes the existence of previously learnt associations between components of an instruction and their physical counterparts, such as objects and actions. While it can be argued that this is a reasonable assumption to make, it does add to the problem of incompleteness as mentioned previously. Third, the model produces an interesting shift from the LPFC route to the striatal one, without being chaperoned, but this shift depends on the way the architecture is configured; if the LPFC route and the striatal route have the same number of synapses to the motor cortex (overseeing the response), then their activations come to resemble one another, as opposed to a "passing of the baton". This dependence on architectural specifics, coupled with the difficulty in estimating a realistic path length for the LPFC route (see Chapter 2) render the model vulnerable to the criticism that it is, in the worst case, a model of convenience. Fourth, the prediction offered by the model, that instructions can eventually be automatised by procedural learning followed by overtraining, is one shared by the COVIS and SPEED models of Ashby et al. (1998; 2007) and shares their difficulty in finding empirical support for the suspected transfer from rule-based to procedural learning (Helie, Roeder & Ashby, 2010).

5.2.2 Model of self-control: strengths and shortcomings

The model of self-control (Ramamoorthy & Verguts, under review) has the following strengths; first, it has been successful at capturing two very broad and important instances of self-control, dietary choice regulation (Hare et al., 2009) and avoidance of preference-reversal in intertemporal choice (Ainslie 1992). Second, it is consistent with several recent findings that point to an integrative account of value computation in the medial frontal cortex (Kable & Glimcher, 2007, 2009; Peters & Buchel, 2009; Rangel & Hare, 2010) and provides a simple mechanistic account of the same. Third, it *predicts* the behavioural distinctions between self-controllers and non-self-controllers in dietary choice *a priori*, as self-control in model terms, is a function of LPFC gating in the VMPFC (see chapter 3 for details) and this gating parameter also influences the distinction between steep and shallow discounting in the intertemporal choice version of the model, such that high gating of LPFC input to the VMPFC leads to self-control in both dietary choice as well as intertemporal choice. This adds to the

economy of the description provided by the model. Fourth, it yields testable predictions, namely, the possibility that repeated application of a self-control rule (e.g. "eat healthy") would then lead its being *cached* and that this would come to be reflected in the response of the striatal valuation pathway, such that the two systems have a cooperative dynamic. In the example of dietary choice, this would correspond to an oft-chosen healthy food item acquiring a higher subjective taste-rating over time, therefore becoming easy to choose over time. It also provides a testable mechanism for temporal discounting that does not involve the explicit representation of a discount function. Recent research appears to support the representation of delay used in our model (Jimura, Chushak & Braver, 2013; for a related yet subtly different perspective on the issue of discounting in the brain, see Marco-Pallares et al., 2010).

As was the case with the previous model, the self-control model has a few shortcomings, in addition to the strengths touched upon above. First, it comes with in-built values specific to the task-at-hand and the acquisition of those values is assumed rather than encapsulated in the model description. This weakens its connection to the model of instruction following, which is, primarily, a model of learning. Second, the model of intertemporal choice and the model of dietary choice are closely related, but distinguished by a single architectural difference; the intertemporal choice model has temporal representations, whereas the latter does not. While the two models do share an architectural core, and use the same mechanism for capturing self-control in decision making, this architectural detail renders the overall approach to self-control less parsimonious than it would have been with a single architecture capable of supporting both instances of self-control. Third, it does not address the widely studied and pertinent phenomenon of inhibition in cognitive control (for a review, see Munakata et al., 2011).

5.2.3 General computational framework: evaluation and perspectives

Just as the individual computational models have their strengths and weaknesses, so does their progenitor, the general computational framework. Briefly, this framework rests on two influential concepts, namely, complementary processing pathways and computational tradeoffs. These are general ideas and generality is often achieved at the

expense of granularity of description, i.e, a high-level idea such as a computational tradeoff offers no purchase on the specifics of how such a tradeoff might work in a very particular physical system that processes information. On one hand, this allows it to be applied to diverse phenomena, and as elucidated previously, such is the case with our framework too.

5.2.3.1 Complementary processing systems

Complementary processing systems have been proposed time and again, in psychology (Kahneman, 2011; Metcalfe & Mischel, 1999; Schneider & Shiffrin, 1977; Sloman, 1996), neuroscience (Hare et al., 2009; Milner & Goodale, 1993, 2006, 2008; McClure et al., 2004, 2007;). and computational modelling of cognition (Ashby et al. 1998; Daw et al., 2005). Despite its generality and influence, or perhaps because of them, this idea has also received its share of criticism in recent times (Keren & Schuul, 2009). Recent studies on reinforcement learning (Daw, O'Doherty, Dayan, Seymour, & Dolan, 2006), value computation(Hare et.al, 2009; McClure et al., 2004, 2007; for a contrasting viewpoint see Kable & Glimcher, 2007, 2009) and visual processing (see Milner & Goodale ,2008) suggest that complementary information processing routes do function in the human brain.

The two systems in our models represent processing pathways that operate at opposite ends of a computational tradeoff. The existence of relatively independent processing systems could also allow the brain to accomplish more in terms of information processing by using one system to perform low-cost computations (automatic responding, for instance) and the other to engage in high-cost computations (tree-search as described in Daw et al., (2005)). Finally, having complementary systems instantiate a tradeoff ensures that reduction in performance in one system does not immediately disadvantage the brain. A unitary system would suffer from performance degradation than one which has multiple layers of processing.

5.2.3.2 Homunculus-free computation

A key strength of our computational framework is that it is not home to a homunculus. In the model of instruction following, all information is learnt locally through Hebbian learning, and the transfer of learning from one system to another does not require the system to "know" in some higher sense of the term, when and where to effect such a transfer. Responses are produced by integrating the output of the available processing systems. Likewise, in the model of self-control, two valuation systems contribute to a goal-value computation in the vmPFC (see chapter 3 for details) without the need for a "self" overseeing self-control. The gating mechanism presented as a potential causal influence in self-control, is a parameter that is distributed across individuals and is not an implicit homunculus which tells the model to control or not.

5.2.3.3 Biological Realism and Marr's third level

O'Reilly (1998), articulated six principles for biologically realistic computational models of cortical cognition in an eponymous article; biological realism, distributed representations, inhibitory competition, bidirectional activation, error-driven task learning and Hebbian model learning (here the term is used in the sense of a "model of the world"). From this perspective, our framework realizes most of these principles; all learning in our models is Hebbian, response selection is achieved through inhibitory competition between response units and the architectures are informed by recent neurobiological findings. Below, we discuss the principle of biological realism and its implications for our framework.

Biological realism is an important test of a computational model of brain function. Seen from the perspective of Marr's tri-level analysis of information processing systems (Marr, 1982), biological realism corresponds to the *third* level, or the *implementation* level, i.e., e answer to "how are the algorithms supporting the computation of interest realized in a physical system?"

As mentioned at the outset, our models capture a general high-level computation (goal-directed behaviours, instructions, self-control etc., aimed at optimizing overall adaptive behaviour and reward) while retaining a strong focus on the algorithmic level. The neural architectures we propose are indeed grounded in neurobiological findings, but they do not correspond to Marr's third level. They are architectures that support specific algorithms (e.g., Hebbian learning, additive value computation) and hint at a biological foundation for the same. However, our models do not venture deeply into the biological implementation of the algorithms they represent. This derives in part, from an *a priori* choice to limit modelling efforts to the neurocognitive level as well as the lack of clear biological correlates for commonly assumed model parameters (such as *learning rate*). This leaves us with open questions on the precise biology of the brain mechanisms we propose, such as the gating parameter used to encapsulate LPFC control over valuation. In this regard, our models display a telling weakness, in that they do not reflect the reality of a changing, adaptive human agent. Learning rates and gating parameters are assumed to be static or utterly deterministic in our models. Reality is, of course, likely to be far richer than this simplification. It is not inconceivable that an integrative account of information processing in the brain across the three levels can be built, from membrane biophysics to behaviour and the extensions of our computational framework provide sufficient motivation to delve into the third level of analysis.

5.3 Predictions and reality: empirical contact

Ideas when subjected to rigorous empirical testing shape the landscape of theory, in science. The general computational framework described here and its instances (i.e., the models derived from it) belong to the theoretical level of scientific discourse. Simulation studies yield useful tests of models but they are no substitute for empirical testing. In the empirical work described below, we attempted to test a specific prediction that related the model of instruction following to the one dealing with self-control; a strong tendency to follow instructions should correlate positively and significantly with a high degree of self-control. This prediction appears to resonate with recent findings that the use of cognitive control is facilitated by the use of

implementation intentions (which are "if-then" statements that function as self-instructed stimulus-response contingencies or cue-response associations), (Cohen, Bayer, Jaudas & Gollwitzer, 2008) as well as the use of self-affirmations to counteract *ego-depletion* (Schmeichel & Vohs, 2009), (for a detailed account of *ego-depletion*, see Muraven & Baumeister, 2000).

5.3.1 The behavioural study and its implications

To test the prediction described above, we designed a behavioural study (see chapter 4 for details) that incorporated three different tasks; a dietary choice task to measure self-control (adapted from Hare et al., 2009), a probabilistic selection task with an instructional component (adapted from Doll et al., 2009) and a working memory task (from the WAIS IV, 2008). We calculated measures of instruction following tendency, self-control and working memory capacity from the tasks, and examined them across 52 subjects. Upon detailed analysis, the individual tasks and the measures computed from them were found to be reliable and the sample sufficiently powerful for each component of the overall study. The correlation between the variables of interest (instruction following, self-control) was found to be nominal and non-significant.

Prima facie, the results of this study do not support our principal hypothesis. However, they do not yield an unambiguous refutation either. This ambiguity derives from the existence of a nominal (non-significant) correlation between the two variables of interest (instruction following tendency and self-control score), the lack of sufficient power for the correlation of interest (<32%) and the fact that null results are typically difficult to interpret (Aberson, 2002). Additionally, nearly 50% of the subjects experienced a slightly different contingency in the probabilistic selection task with this factor (contingency type) having a significant effect (as reported in chapter 4). This implies that the sample suffered from a lack of homogeneity with respect to experimental conditions. Subjects performed all 3 tasks in one session, and as we have discussed in chapter 4, this could have contributed to noise in the data, reducing the quality of performance in individual tasks. Also, the sample perhaps did not contain enough instruction followers or self-controllers to adequately test for a relation between the two phenomena.

While it may well be the case that controlling for these factors and replicating the study with a larger sample could potentially yield a different outcome, we note here that the results of our study suggest that we reduce belief in the hypothesis and remain open to invalidation of the theoretical motivations for the same. Given recent findings that lend support to the individual models such as the dynamics of learning-transfer in instruction following (Wolfensteller & Ruge, 2012) and temporal representations in value discounting (Jimura et al., 2013) and the interaction between them, for instance, the influence of affirmations on self-control (Schmeichel & Vohs, 2009), we suggest that the framework requires additional testing to achieve true falsification and as such it has not been disconfirmed by the null result obtained from our behavioural study.

5.4 Taking stock, moving forward: directions for future research

Thus far, we have articulated a set of models to capture two indispensable aspects of human adaptive behaviour; instruction following (Ramamoorthy & Verguts, 2012) and self-control (Ramamoorthy & Verguts, under review). We tested an important prediction through a behavioural study and encountered a null result which remains open to interpretation. Here we address the following pertinent query.

5.4.1 Where do we go from here?

5.4.1.1 Computational modelling

The models described here and any future models derived from the aforementioned framework would likely benefit from more detailed biological specification to increase their testability. A model that lays down a specific way of realizing a systems-level mechanism is more refutable than one that is mechanistically vague or agnostic. We hope to achieve greater biological realism and therefore a realization of all three levels of analysis in Marr's framework in future modelling endeavors.

Of immediate and particular interest, in this context, are the findings concerning the role of neuromodulators in learning (Daw & Doya, 2006), cognitive control (Cools, 2008) and temporal discounting (Schweighofer, Tanaka & Doya, 2007). The current

models can be augmented to include the effects of neuromodulators on self-control; for example, recent findings suggest that serotonin depletion leads to impulsive choice and steep-discounting (Crockett, Clark, Lieberman, Tabibnia & Robbins, 2010; Schweighofer et al., 2007). It would be useful to explore the computational basis of this phenomenon.

5.4.1.2 Empirical testing

First, in the case of the instruction following model, an extensive model-based fMRI study to examine the temporal dynamics of initial instructional-learning of a stimulus-response association and eventual automatisation of the same would likely prove useful (for a review on model-based fMRI, see Gläscher and O'Doherty, 2010). Specifically, the model performs instruction following and automatically transfers the instructed mapping from one learning-system (LPFC) to another (striatal). One way of testing this proposed transfer empirically, would be to conduct an fMRI study where subjects perform an instruction-following task (for example, the task from Ruge & Wolfensteller (2010), where simple stimulus-response associations are instructed and tested) in the scanner over an extended period of time (i.e., multiple training sessions). Behavioural results for each session, from each subject would be used to generate the corresponding activations and temporal dynamics in the computational model, with the output from the model being used as a regressor in the analysis of the fMRI data. Given that transfer in the model is a consequence of the computational tradeoff, it depends on the speed of learning as well as responding in the two pathways. We predict that, given inter-individual differences in processing speed in the LPFC and striatum, the model's activation patterns for these regions over time should correlate with maximally active voxels in the brain that correspond to the LPFC and striatum of the individual subjects.

Second, a similar approach can be used to study temporal changes in valuation. Self-control is, in our framework, an instance of cognitive control. Given our computational framework, it can be argued that self-control can be learnt and automatised for specific stimuli or circumstances, just as instructions applied over and over again become *cached* responses to stimuli. The work of Gollwitzer and Schaal

(1998) lends support to this idea. It would be interesting to study whether repeated application of self-control leads to a transfer of the valuation from the LPFC to the striatal system, such that choices which needed deliberate control become valued as being immediately rewarding over time. An example of this would be the change in the subjective evaluation of a food item, which is initially considered merely healthy and not particularly tasty, but comes to be perceived as having higher taste value than before, due to repeated application of self-control to select it. A model-based fMRI study, as described above, can be used to delineate the occurrence of such a change in evaluation by tracking the activation in regions of interest across over time.

5.4.2 Conclusion

To summarize, our relatively simple models appear to capture otherwise complex phenomena such as instruction following and self-control and provide insights into how they *might* occur in the brain but we do not purport to *know* this in any conclusive manner. Our research raises a fresh set of questions to be explored and is best viewed as work-in-progress. We tentatively conclude (for science is an interminable process), by expressing belief in the utility of our theoretical framework and looking forward to exploring open questions in learning and cognitive control. In this ongoing endeavor, we concur with Richard Feynman; "*I think, it is much more interesting to live not knowing, than to have answers which might be wrong*." (Feynman, 1999); for in the acknowledgment of fallibility and uncertainty lies the hope for increased understanding.

NEDERLANDSTALIGE SAMENVATTING

Inleiding

Deze thesis bestudeert leren en cognitieve controle vanuit computationele en empirische perspectieven. Leren en cognitieve controle beïnvloeden menselijk gedrag grondig en op vele manieren. We kozen ervoor om te focusen op het volgen van instructies en zelfcontrole, omdat deze aspecten van cognitieve controle nog slecht begrepen zijn. In lijn met Daw et al. (2005) vertrokken we van een computationeel kader gekarakteriseerd door een tradeoff tussen leren en handelen, geïmplementeerd in complementaire paden in het brein. Meer bepaald, het ene pad leert snel maar handelt traag, terwijl het andere pad traag leert maar snel handelt. We pasten dit kader toe op de fenomenen waarvan sprake, namelijk het volgen van instructies en zelfcontrole. We bouwden computationele modellen gebaseerd op deze tradeoff en gebruiken deze modellen om relevante fenomenen te simuleren uit de literatuur omtrent het volgen van instructies en zelfcontrole.

Model voor het volgen van instructies

Elk experiment in de psychologie begint met instructies relevant voor de taak die de proefpersoon uit te voeren heeft. Vele modellen voor cognitieve controle gaan ervan uit dat de relevante taakinformatie op een of andere manier aanwezig is in het brein en toelaat om de taak uit te voeren. Echter, hoe dit precies gebeurt is nog slecht begrepen. Recent heeft het volgen van instructies verhoogde aandacht gekregen van zowel theoretici als experimentele onderzoekers. Onderzoek naar de gedragsaspecten ervan toonde aan dat instructies snel en met zeer hoge accuraatheid geïmplementeerd kunnen worden (zie bijvoorbeeld Ruge & Wolfensteller, 2010). Instructies kunnen ook interfereren met eerder geleerde responsen (zie bijvoorbeeld Waszak et al., 2008).

Men vermoedt dat het volgen van instructies een prefrontaal-striataal netwerk aanspreekt (zie bijvoorbeeld Hartstra et al., 2011; Ruge & Wolfensteller, 2010).

We construeerden een computationeel model voor het volgen van instructies. Het model heeft twee paden; een lateraal-prefrontaal pad en een striataal pad. Het eerste pad implementeert instructies via snel Hebbiaans leren. Het tweede pad zorgde voor geleidelijke automatizatie van de instructies. Dit model kon typische fenomenen verklaren zoals versnelde reactietijden naarmate training toeneemt, en zeer hoge accuraatheden vanaf de eerste proefbeurt. Het model gaf ook een verklaring voor de interferentie-effecten die gerapporteerd werden door Waszak et al. (2008), waar een geïnstrueerde stimulus-respons mapping die nooit echt geïmplementeerd werd, toch reactietijden kon vertragen van (andere) toegepaste stimulus-respons mappings. Tot slot bekeken we met dit model ook het fenomeen van zgn. "goal neglect" (Duncan, 1996) waar proefpersonen of patiënten een doel kunnen uitspreken (in dit geval, een simpele stimulus-response mapping) zonder het effectief te kunnen implementeren. Predicties van het model kunnen in gedrags- of fMRI vervolgonderzoek getest worden.

Model voor zelfcontrole

Zelfcontrole is een belangrijk maar minder goed bestudeerd aspect van cognitieve controle. Hoewel het op empirisch niveau wel bestudeerd werd in de psychologie en de cognitieve neurowetenschappen (bijvoorbeeld, Baumeister, 1998) is er zeer weinig geweten over de computationele basis ervan. Om dit aspect aan te pakken, pasten we het algemene kader toe op zelfcontrole, in de domeinen van dieet-gerelateerde beslissingsprocessen en intertemporele keuzes.

Met betrekking tot dieet-gerelateerde beslissingen probeerden we de bevindingen van Hare et al. (2009) te repliceren. Deze auteurs toonden aan dat mensen die zelfcontrole uitoefenen zowel gezondheids- als smaakinformatie in hun beslissingen incorporeren, en ze toonden aan welke hersengebieden bij deze soorten informatie betrokken waren. Het model had complementaire paden. Elk pad leverde een waarde-input aan een centrale waarde-vergelijker, die de informatie van beide paden integreerde.

Gezondheid werd "berekend" door de lateraal-prefrontale cortex, en smaak door de nucleus accumbens. De waarde-vergelijking werd uitgevoerd door de ventraal-mediale prefrontale cortex. Het model simuleerde o.a. de gegevens van Hare et al. (2009).

Hierna bouwden we een model voor zelfcontrole bij intertemporele keuzes. Dit verwijst naar een keuze tussen opties die op verschillende tijdsmomenten beschikbaar zijn. Het klassieke voorbeeld is tussen een kleine beloning die onmiddellijk beschikbaar is en een grote beloning die pas later komt. Het model had dezelfde algemene architectuur als voorheen beschreven. Zelfcontrole, of de afwezigheid ervan, werd gemoduleerd door een "gating" parameter in de lateraal-prefrontale cortex. Een sterke "gating" leidde tot keuzes die consistent waren over de tijd.

Empirische test van het model

Een predictie uit het modelleerkader was dat er een correlatie is tussen de mate waarin iemand instructies volgt en zelfcontrole uitoefent. Om dit te testen, voerden we een studie uit met de volgende taken; een dieetkeuze taak (gebaseerd op Hare et al., 2009); een probabilistische selectietaak (gebaseerd op Doll et al., 2009); en een werkgeheugen taak om te controleren voor werkgeheugen aspecten.

In de dieetkeuze taak moesten proefpersonen telkens zeggen of ze een bepaald voedsel verkozen ten opzichte van een referentie item. Zo berekenden we een mate voor zelfcontrole. In de probabilistische selectietaak moesten proefpersonen een keuze maken tussen één van twee stimuli (Hiragana karakters) die telkens op het scherm verschenen. De stimuli waren statistisch geordend (A beter dan B; C beter dan D; E beter dan F). We gaven echter de instructie aan proefpersonen om een statistisch superieure stimulus te vermijden (vermijd E in het paar EF). Op basis hiervan berekenden we per proefpersoon een mate van instructievolgen. Er bleek echter geen correlatie te zijn tussen de twee variabelen. We argumenteren dat de studie hernomen moet worden met een nieuw, strikter design om de predictie te testen.

Conclusie

Computationele modellering is een krachtig instrument voor het begrijpen van vele natuurlijke processen, inclusief cognitie in de hersenen. Het nut bestaat erin dat het fenomenen over vele niveaus van beschrijving heen toelaat te integreren. We pasten deze algemene benadering toe op het volgen van instructies en zelfcontrole. Na een succesvolle implementatie van deze modellen, testten we een predictie, die echter onduidelijke resultaten opleverde. Gegeven de algemene toepasbaarheid van het algemene kader, blijven we echter optimistisch dat latere experimenten meer duidelijkheid kunnen verschaffen. Hoe het ook uitdraait, op deze manier hopen we dat we in de toekomst een beter zicht zullen kunnen krijgen op deze processen.

REFERENCES

Abbott, L. F., & Nelson, S. B. (2000). Synaptic plasticity: taming the beast . *Nature neuroscience , 3*, 1178-1183.

Aberson, C. (2002). Interpreing Null Results: Improving Presentation and Conclusions with Confidence Intervals. *Journal of Articles in Support of the Null Hypothesis , 1* (3), 36-42.

Ainslie, G. (2001). *Breakdown of will.* Cambridge: Cambrdige University Press.

Ainslie, G. (1992). *Picoeconomics: The strategic interaction of successive motivational states within the person.* Cambridge: Cabridge University Press.

Alexander, W. H., & Brown, J. W. (2010). Computational models of performance monitoring and cognitive control. *Topics in cognitive science , 2* (4), 658-677.

Amat, J. E., Paul, E., Watkins, L. R., & Maier, S. F. (2008). Activation of the ventral medial prefrontal cortex during an uncontrollable stressor reproduces both the immediate and long-term protective effects of behavioral control. *Neuroscience , 154* (4), 1178-1186.

Antonov, I., Antonova, I., Kandel, E. R., & Hawkins, R. D. (2003). Activity-Depedent Presynaptic Facilitation and Hebbian LTP Are Both Required and Interact during Classical Conditioning in Aplysia. *Neuron , 37* (1), 135-147.

Aron, A. R. (2008). Progress in Executive-Function Research From Tasks to Functions to Regions to Networks. *Current Directions in Psychological Science , 17* (2), 124-129.

Ashby, F. G., & Crossley, M. J. (2010). Interactions between declarative and procedural-learning categorization systems . *Neurobiology of learning and memory , 94* (1), 1-12.

Ashby, F. G., & Maddox, W. T. (2005). Human category learning. *Annual Review of Psychology , 56* (1), 149-178.

Ashby, F. G., Alfonso-Reese, L. A., Turken, A. U., & Waldron, E. M. (1998). A neuropsychological theory of multiple systems in category learning. *Psychological Review , 105*, 442-481.

Ashby, G. F., Ennis, J. M., & Spiering, B. J. (2007). A Neurobiological Theory of Automaticity in Perceptual Categgorization. *Psychological Review , 114* (3), 632-656.

Ashby, G. F., Paul, E. J., & Maddox, T. W. (2011). COVIS. In E. M. Pothos, & A. J. Wills (Eds.), *Formal approaches in categorization.* New York: Cambridge University Press.

Ashby, G. F., Turner, B. O., & Horvitz, J. C. (2010). Cortical and basal ganglia contributions to habit learning and automaticity. *Trends in Cognitive Sciences , 14* (5), 208-215.

Asken, M. J., Grossman, D., & Christensen, L. W. (2010). *Warrior mindset: Mental toughness skills for a Nation's peacekeepers.* Washington , DC: Library of Congress.

Baler, R. D., & Volkow, N. D. (2006). Drug addiction: the neurobiology of disrupted self-control. *Trends in Molecular Medicine , 12* (12), 559-566.

Bandura, A. (1989). Human agency in social cognitive theory. *American psychologist , 44* (9), 1175-1184.

Bandura, A. (1977). *Social Learning Theory.* NJ: Prentice Hall.

Basten, U., Biele, G., Heekeren, H. R., & Fiebach, C. J. (2010). How the brain integrates costs and benefits during decision making. *Proceedings of the National Academy of Sciences , 107* (50), 21767-21772.

Baumeister, R. F., Bratslavsky, E., Muraven, M., & Tice, D. M. (1998). Ego Depletion: Is the Active Self a Limited Resource? *Journal of Personality and Social Psychology , 74* (5), 1252-1265.

Benedetti, F., Mayberg, H. S., Wager, T. D., Stohler, C. S., & Zubieta, J.-K. (2005). Neurobiological Mechanisms of the Placebo Effect. *The Journal of Neuroscience , 25* (45), 10390-10402.

Benhabib, J., & Bisin, A. (2005). Modeling internal commitment mechanisms and self-control: A neuroeconomics approach to consumption-saving decisions. *Games and Economic Behaviour , 52*, 460-492.

Biele, G., Rieskamp, J., Krugel, L. K., & Heekeren, H. R. (2011). The neural basis of following advice. *PLoS biology , 9* (6), e1001089.

Bienenstock, E. L., Cooper, L. N., & Munro, P. W. (1982). Theory for the development of neuron selctivity: orientation specificity and binocular interaction in visual cortex. *The Journal of Neuroscience , 2* (1), 32-38.

Botvinick, M. M., & Plaut, D. C. (2006). Short-Term Memory for Serial Order: A Recurrent Neural Network Model. *Psychological Review , 2*, 201-233.

Botvinick, M. M., Braver, T. S., Barch, D. M., Carter, C. S., & Cohen, J. D. (2001). Conflict Monitoring and Cognitive Control. *Psychological Review , 108*, 624-652.

Boureau, Y.-L., & Dayan, P. (2010). Opponency Revisited:Competition and Cooperation Between Dopamine and Serotonin. *Neuropsychopharmacology Reviews* , 1-24.

Brass, M., Wenke, D., Spengler, S., & Waszak, F. (2009). Neural Correlates of Overcoming Interference from Instructed and Implemented Stimulus-Response Associations. *The Journal of Neuroscience , 29(6)*, 1766-1772.

Brown, J., Bullock, D., & Grossberg, S. (1999). How the Basal Ganglia Use Parallel Excitatory and Inhibitory Learning Pathways to Selectively Respond to Unexpected Rewarding Cues. *The Journal of Neuroscience , 19* (23), 10502-10511.

Buonomano, D. V., & Merzenich, M. M. (1995). Temporal information transformed into a spatial code by a neural network with realistic properties. *Science* , 1028-1028.

Carver, C. S., Johnson, S. L., & Joorman, J. (2008). Serotonergic Function, Two-Mode Models of Self Regulation, and Vulnerability to Depression: What Depression Has in Common with Impulsive Aggression. *Psychological Bulletin , 134* (6), 912-943.

Casey, B. J., Somerville, L. H., Gotlib, I. H., Ayduk, O., Franklin, N. T., Askren, M. K., et al. (2011). Behavioural and neural correlates of delay of gratification 40 years later. *Proceedings of the National Academy of Sciences , 108* (36), 14998-15003.

Cohen, A.-L., Bayer, U. C., Jaudas, A., & Gollwitzer, P. M. (2008). Self-regulatory strategy and executive control: implementation intentions modulate task switching and Simon task performance. *Psychological Research , 72*, 12-26.

Cohen, J. D., Dunbar, K., & McClelland, J. L. (1990). On the Control of Automatic Processes: A Parallel Distributed Processing Account of the Stroop Effect. *Psychological Review , 97* (3), 332-261.

Cohen-Kdoshay, O., & Nachshon, M. (2009). The Representation of Instructions Operates Like a Prepared Reflex. *Experimental Psychology , 56(2)*, 128-133.

Cole, M. W., Bagic, A., Kass, R., & Schneider, W. (2010). Prefrontal Dynamics Underlying Rapid Instructed Task Learning Reverse with Practice. *The Journal of Neuroscience , 30(42)*, 14245-14254.

Cole, M. W., Laurent, P., & Stocco, A. (2013). Rapid instructed task learning: A new window into the human brain's unique capacity for flexible cognitive control. *Cognitive, Affective & Behavioral Neuroscience , 13* (1), 1-22.

Cools, R. (2008). Role of dopamine in the motivational and cognitive control of behavior . *The Neuroscientist , 14* (4), 381-395.

Cooper, L. N., & Bear, M. F. (2012). The BCM theory of synapse modification at 30: interaction of theory with experiment . *Nature Reviews Neuroscience , 13* (11), 798-810.

Craggs, J. G., Price, D. D., Perlstein, W. M., Verne, G. N., & Robinson, M. E. (2008). The dynamic mechanisms of placebo induced analgesia:Evidence of sustained and transient regional involvement. *Pain , 139*, 660-669.

Crockett, M. J., Clark, L., Lieberman, M. D., Tabibnia, G., & Robbins, T. W. (2010). Impulsive choice and altruistic punishment are correlated and increase in tandem with serotonin depletion. *Emotion , 10* (6), 855-862.

Cross, E. S., Kraemer, D. J., Hamilton, A. F., Kelley, W. M., & Grafton, S. T. (2009). Sensitivity of the action observation network to physical and observational learning . *Cerebral Cortex , 19* (2), 315-326.

Czernochowski, D. (2011). ERP evidence for scarce rule representations in older adults following short, but not long preparatory intervals. *Frontiers in Psychology , 2*.

Daw, N. D., & Doya, K. (2006). The computational neurobiology of learning and reward. *Current opinion in neurobiology , 16* (2), 199-204.

Daw, N. D., & Touretzky, D. S. (2002). Long-Term Reward Prediction in TD Models of the Dopamine System. *Neural Computation , 14*, 2567-2583.

Daw, N. D., Niv, Y., & Dayan, P. (2005). Unertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nature Neuroscience , 8* (12), 1704-1711.

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature , 441* (7095), 876-879.

Dayan, P., Niv, Y., Seymour, B., & Daw, N. D. (2006). The misbehaviour of value and the discipline of the will. *Neural Networks , 19* (8), 1153-1160.

De Houwer, J., Beckers, T., Vandorpe, S., & Custers, R. (2005). Further evidence for the role of mode-independent short-term associations in spatial Simon effects. *Perception & Psychophysics , 67(4)*, 659-666.

de Ridder, D. T., Lensvelt-Mulders, G., Finkenauer, C., Stok, F. M., & Baumesiter, R. F. (2012). Taking Stock of Self-Control : A Meta-Analysis of How Trait Self-Control Relates to a Wide Range of Behaviours. *Personality and Social Psychology Review , 16* (1), 76-99.

Dehaene, S., & Cohen, L. (2007). Cultural Recylcing of Cortical Maps. *Neuron , 56*, 384-398.

Derbyshire, S. W., Whalley, M. G., & Oakley, D. A. (2009). Fibromyalgia pain and its modulation by hypnotic and non-hypnotic suggestion: An fMRI analysis. *European Journal of Pain , 13*, 542-550.

Derrfuss, J., Brass, M., Neumann, J., & von Cramon, D. (2005). Involvement of the Inferior Frontal Junction in Cognitive Control: Meta-Analyses of Switching and Stroop Studies. *Human Brain Mapping , 25*, 22-34.

Derrfuss, J., Brass, M., von Cramon, D., Lohmann, G., & Amunts, K. (2009). Neural activation at the junction of the inferior frontal sulcus and the inferior precentral sulcus: interindividual variability, reliability, and association with sulcal morphology. *Human Brain Mapping , 30* (1), 299-311.

Diamond, P., & Koszegi, B. (2003). Quasi-hyperbolic discounting and retirement. *Journal of Public Economics , 87* (9), 1839-1872.

Doll, B. B., Jacobs, W. J., Sanfey, A. G., & Frank, M. J. (2009). Instructional control of reinforcement learning: A behavioural and neurocomputational investigation. *Brain Research , 1299*, 74-94.

Dominey, P. F. (2005). From Sensorimotor Sequence to Grammatical Construction: Evidence from Simulation and Neurophysiology. *Adaptive Behaviour , 13* (4), 347-361.

Duncan, J., Emslie, H., & Williams, P. (1996). Intelligence and the Frontal Lobe: The Organization of Goal-Directed Behaviour. *Cognitive Psychology , 30*, 257-303.

Duncan, J., Parr, A., Woolgar, A., Thompson, R., Bright, P., Cox, S., et al. (2008). Goal Neglect and Spearman's g: Competing Parts of a Complex Task. *Journal of Experimental Psychology , 137* (1), 131-148.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science , 14* (2), 179-211.

Feynman, R. P. (1974). Cargo cult science. *Engineering and Science , 37* (7), pp. 10-13.

Feynman, R. P. (1965). *The character of physical law*. Cambridge, MA: MIT Press.

Feynman, R. P. (1999). *The Pleasure of Finding Things Out* (Vol. 178). Cambridge: Perseus Books.

Figner, B., Knoch, D., Johnson, E. J., Krosch, A. R., Lisanby, S. H., Fehr, E., et al. (2010). Lateral prefrontal cortex and self-control in intertemporal choice. *Nature neuroscience , 13* (5), 538-539.

Fischer, M. H., & Zwaan, R. A. (2008). Embodied language: A review of the tole of the motor system in language comprehension. *The Quarterly Journal of Experimental Psychology , 61* (6), 825-850.

Fox, C. (2013). Formalising robot ethical reasoning as decision heuristics . *First UK Workshop on Robot Ethics (UKRE).*

Frank, M. J. (2005). Dynamic Dopamine Modulation in the Basal Ganglia: A Neurocomputational Account of Cognitive Deficits in Medicated and Nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience , 17* (1), 51-72.

Frank, M. J. (2006). Hold your horses: A dynamic computational role for the subthalamic nucleus in decision making. *Neural Networks , 19*, 1120-1136.

Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism . *Science , 306* (5703), 1940-1943.

Freeman, J. A., & Nicholson, C. N. (1970). Space-Time Transformation in the Frog Cerebellum through an Intrinsic Tapped Delay-line. *Nature* , 640-642.

Frenda, S. J., Nichols, R. M., & Loftus, E. F. (2011). Current Issues and Advances in Misinformation Research. *Current Directions in Psychological Science , 20* (1), 20-23.

Fritzke, B. (1997). *Some competitive learning methods.* Ruhr-Universitat Bochum.

Glaescher, J. P., & O'Doherty, J. P. (2010). Model-based approaches to neuroimaging: combining reinforcement learning theory with fMRI data. *Cognitive Science , 1*, 501-510.

Glaescher, J., Daw, N., Dayan, P., & O'Doherty, J. P. (2010). States versus Rewards: Dissociable Neural Prediction Error Signals Underlying Model-based and Model-free Reinforcement Learning. *Neuron , 66*, 585-595.

Goldstein, R. Z., Craig, A., Bechara, A., Garavan, H., Childress, A. R., Paulus, M. P., et al. (2009). The neurocircuitry of impaired insight in drug addiction. *Trends in cognitive sciences , 13* (9), 372-380.

Gollwitzer, P. M., & Schaal, B. (1998). Metacognition in Action: The Importance of Implementation Intentions . *Personality and Social Psychology Review , 2* (2), 124-136.

Gottfredson, M., & Hirschi, T. (1990). *A general theory of crime.* Stanford University Press.

Grabenhorst, F., & Rolls, E. T. (2011). Value, pleasure and choice in the ventral prefrontal cortex. *Trends in ognitive sciences , 15* (2), 56-67.

Grill-Spector, K., Henson, R., & Martin, A. (2006). Repetion and the brain: neural models of stimulus-specific effects. *Trends in cognitive sciences , 10* (1), 14-23.

Grossberg, S. (1973). Contour enhancement, short-term memory, and constancies in reverberating neural networks. *Studies in Applied Mathematics , 52*, 213-257.

Hagger, M. S., Wood, C., Stiff, C., & Chatzisarantis, N. L. (2010). Ego Depletion and the Strength Model of Self-Control: A Meta-Analysis. *Psychological Bulletin , 136* (4), 495-525.

Hare, T. A., Camerer, C. F., & Rangel, A. (2009). Self-Control in Decision-Making Involves Modulation of the vmPFC Valuation System. *Science , 324*, 646-648.

Hare, T. A., Camerer, C. F., Knoepfle, D. T., O'Doherty, J. P., & Rangel, A. (2010). Value computations in ventral medial prefrontal cortex during charitable decision making incorporate input from regions involved in social cognition. *The Journal of Neuroscience , 30* (2), 583-590.

Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing Attention on the Health Aspects of Foods Changes Value Signals in vmPFC and Improves Dietary Choice. *The Journal of Neuroscience , 31* (30), 11077-11087.

Hare, T. A., Malmaud, J., & Rangel, A. (2011). Focusing Attention on the Health Aspects of Foods Changes Value Signals in vmPFC and Improves Dietary Choice. *The Journal of Neuroscience , 31* (30), 11077-11087.

Harris, A., Adolphs, R., Camerer, C., & Rangel, A. (2011). Dynamic construction of stimulus values in the ventromedial prefrontal cortex. *PloS one , 6* (6), e21074.

Hartstra, E., Kuhn, S., Verguts, T., & Brass, M. (2011). The implementation of verbal instructions: an fMRI study. *Human Brain Mapping , 32*, 1811-1824.

Hauk, O., Shtyrov, Y., & Pulvermuller, F. (2008). The time course of action and action-word comprehension in the human brain as revealed by neurophysiology. *Journal of Physiology, Paris , 102* ((1-3)), 50-58.

Hebb, D. O. (1949). *The organization of behaviour: A neuropsychological approach.* John Wiley & Sons.

Helie, S., & Ashby, G. F. (2009). A Neurocomputational Model of Automaticity and Maintenance of Abstract Rules. *Proccedings of International Joint Conference on Neural Networks*, (pp. 1192-1198). Atlanta, Georgia, USA.

Helie, S., Roeder, L. J., & F.Gregory, A. (2010). Evidence for Cortical Automaticity in Rule-Based Categorization. *The Journal of Neuroscience , 30* (42), 14225-14234.

Hinton, G. E. (1989). Connectionist learning procedures . *Artificial intelligence , 40* (1), 185-234.

Hofmann, W., Baumeister, R. F., Forster, G., & Vohs, K. D. (2012). Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology , 102* (6), 1318.

Hollmann, M., Hellrung, L., Pleger, B., Schlogl, H., Kabisch, S., Stumvoll, M., et al. (2012). Neural correlates of the volitional regulation of the desire for food. *International Journal of Obsesity , 36*, 648-655.

Huang, T.-R., Hazy, T. E., Herd, S. A., & O'Reilly, R. C. (2013). Assembling Old Tricks for New Tasks: A Neural Model of Instructional Learning and Control. *Journal of Cognitive Neuroscience , 25* (6), 843-851.

Jasinska, A. J., Ramamoorthy, A., & Crew, C. M. (2011). Toward a neurobiological model of cue-induced self-control in decision making: relevance to addiction and obesity. *The Journal of Neuroscience , 31* (45), 16139-16141.

Jimura, K., Chushak, M. S., & Braver, T. S. (2013). Impulsivity and Self-Control during Intertemporal Decision Making Linked to the Neural Dynamics of Reward Value Representation. *The Journal of Neuroscience , 33* (1), 344-357.

Job, V., Dweck, C. S., & Walton, G. M. (2010). Ego Depletion- Is It All in Your Head? *Psychological science , 21* (11), 1686-1693.

Kable, J. W., & Glimcher, P. W. (2007). The neural correlates of subjective value during intertemporal choice . *Nature neuroscience , 10* (12), 1625-1633.

Kable, J. W., & Glimcher, P. W. (2009). The neurobiology of decision: consensus and controversy. *Neuron , 63* (6), 733-745.

Kahneman, D. (2011). *Thinking, fast and slow.* Farrar, Straus and Giroux.

Kandel, E. R., Abrams, T., Bernier, L., Carew, T. J., Hawkins, R. D., & Schwartz, J. H. (1983). Classical conditioning and sensitization share aspects of the same molecular cascade in Aplysia. *Cold Spring Harbor symposia on quantitative biology . 48*, pp. 821-830. Cold Spring Harbor Laboratory Press.

Kelso, S. R., Ganong, A. H., & Brown, T. H. (1986). Hebbian synapses in hippocampus . *Proceedings of the National Academy of Sciences , 83* (14), 5326-5330.

Kempter, R., Gerstner, W., & Van Hammen, J. L. (1999). Hebbian learning and spiking neurons. *Physical Review E , 59*, 4498-4514.

Keren, G., & Schul, Y. (2009). Two Is Not Always Better Than One: A Critical Evaluation of Two-System Theories. *Perspectives on Psychological Science , 4* (6), 533-550.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed Neural Representation of Expected Value. *The Journal of Neuroscience , 25* (19), 4806-4812.

Knutson, B., Taylor, J., Kaufman, M., Peterson, R., & Glover, G. (2005). Distributed neural representation of expected value. *The Journal of Neuroscience , 25* (19), 4806-4812.

Koestner, R., Lekes, N., Powers, T. A., & Chicoine, E. (2002). Attaining personal goals: Self-concordance plus implementation intention equals success. *Journal of personality and social psychology , 83* (1), 231-244.

Laibson, D. (1997). Golden eggs and hyperbolic discounting . *The Quarterly Journal of Economics , 112* (2), 443-478.

Loewenstein, G., & Prelec, D. (1992). Anomalies in intertemporal choice: Evidence and an interpretation. *The Quarterly Journal of Economics , 107* (2), 573-597.

Luria, A. R. (1966). *Higher cortical functions in man.* London: Tavistock.

Maier, S. F., & Watkins, L. R. (2010). Role of the medial prefrontal cortex in coping and resilience. *Brain research , 1355*, 52-60.

Marco-Pallares, J., Mohammadi, B., Samii, A., & Munte, T. F. (2010). Brain activations reflect individual discount rates in intertemporal choice. *Brain Research , 1320*, 123-129.

Marr, D. (1982). *Vision: A computational approach.*

Mazur, J. E., & Biondi, D. R. (2009). Delay-amount tradeoffs in choices by pigeons and rats:hyperbolic versus exponential discounting. *Journal of the experimental analysis of behaviour , 91* (2), 197-211.

Mazzoni, G. A., Loftus, E. F., & Kirsch, I. (2001). Changing Beliefs About Implausible Autobiographical Events: A Little Plausbility Goes a Long Way. *Journal of Experimental Psychology, Applied , 7* (1), 51-59.

McClelland, J. L., McNaughton, B. L., & O'Reilly, R. C. (1995). Why there are complementary learning systems in the hippocampus and neocortex: insights from the successes and failures of connectionist models of learning and memory. *Psychological Review , 102* (3), 419-457.

McClure, S. M., Ericson, K. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2007). Time Discounting for Primary Rewards. *The Journal of Neuroscience , 27* (21), 5796-5804.

McClure, S. M., Laibson, D. I., Loewenstein, G., & Cohen, J. D. (2004). Separate Neural Systems Value Immediate and Delayed Monetary Rewards. *Science , 306*, 503-507.

McDannald, M. A., Lucantonio, F., Burke, K. A., 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* (7), 2700-2705.

Medina, J. F., & Mauk, M. D. (2000). Computer simulation of cerebellar information processing. *nature neuroscience , 3*, 1205-1211.

Metcalfe, J., & Mischel, W. (1999). A Hot/Cool-System Analysis of Delay of Gratification: Dynamics of Willpower. *Psychological Review , 106* (1), 3-19.

Milner, A. D., & Goodale, M. A. (1993). Visual pathways to perception and action. *Progress in Brain Research , 95*, 317-337.

Milner, D. A., & Goodale, M. A. (2008). Two visual systems re-viewed. *Neuropsychologia , 46* (3), 774-785.

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

Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences , 108* (7), 2693-2698.

Monsell, S. (1996). Control of Mental Processes. In V.Bruce (Ed.), *Unsolved Mysteries of the Mind* (pp. 93-148). Hove, England: Erlbaum.

Munakata, Y., Herd, S. A., Chatham, C. H., Depue, B. E., Banich, M. T., & O'Reilly, R. C. (2011). A unified framework for inhibitory control. *Trends in cognitive sciences , 15* (10), 453-459.

Muraven, M., & Baumeister, R. F. (2000). Self-regulation and depletion of limited resources: Does self-control resemble a muscle? *Psychological Bulletin , 126* (2), 247.

Nakayama, Y., Yamagata, T., Tanji, J., & Hoshi, E. (2008). Transformation of a Virtual Action Plan into a Motor Plan in the Premotor Cortex. *The Journal of Neuroscience , 28* (41), 10287-10297.

Niv, Y., O'Duff, M., & Dayan, P. (2005). Dopamine, uncertainty and TD learning. *Behavioral and Brain Functions , 1* (6), 1-9.

Noelle, D. C., & Cottrell, G. W. (1995). A Connectionist Model of Instruction Following. In J. D. Moore, & J. F. Lehman (Ed.), *Proceedings of the Seventeenth Annual Conference of the Cognitive Science Society* (pp. 369-374). Hillsdale, NJ: Lawrence Erlbaum Associates.

O'Donoghue, T., & Rabin, M. (1999). Doing it now or later. *American Economic Review* , 103-124.

O'Reilly, R. C. (2001). Generalization in interactive networks: The benefits of inhibtory competition and Hebbian learning . *Neural computation , 13* (6), 1199-1241.

O'Reilly, R. C. (1998). Six principles for biologically based computational models of cortical cognition . *Trends in cognitive sciences , 2* (11), 455-462.

O'Reilly, R. C., & Norman, K. A. (2002). Hippocampal and neocortical contributions to memory: Advances in the complementary learning systems framework. *Trends in cognitive sciences , 6* (12), 505-510.

Pasupathy, A., & Miller, E. K. (2005). Different time courses of learning-related activity in prefrontal cortex and striatum. *Nature , 433* (7028), 873-876.

Pavlov, I. P. (1927). *Conditioned reflexes: an investigation of the physiological activity of the cerebral cortex.* (G. V. Anrep, Trans.) London: Oxford University Press.

Peters, J., & Buchel, C. (2010). Neural representations of subjective reward value . *Behavioural brain research , 213* (2), 135-141.

Petrovic, P., Kalso, E., Petersson, K. M., Andersson, J., Fransson, P., & Ingvar, M. (2010). A prefrontal non-opioid mechanism in placebo analgesia. *Pain , 150* (1), 59-65.

Philiastides, M. G., Biele, G., & Heekeren, H. R. (2010). A mechanistic account of value computation in the human brain. *Proceedings of the National Academy of Sciences , 107* (20), 9430-9435.

Pierce, C. R., & Kumaresan, V. (2006). The mesolimbic dopamine system: the final common pathway for the reinforcing effect of drugs of abuse? *Neuroscience & biobehavioral reviews , 30* (2), 215-238.

Ploghaus, A., Becerra, L., Borras, C., & Borsook, D. (2003). Neural circuitry underlying pain modulation: expectation, hynposis, placebo. *TRENDS in Cognitive Sciences , 7* (5), 197-200.

Poggio, T. (1981). Marr's computational approach to vision. *Trends in neurosciences , 4*, 258-262.

Popper, K. R. (1963). *Conjectures and refutations: the growth of scientific knowledge.*

Ramamoorthy, A., & Verguts, T. (2012). Word and deed: A computational model of instruction following. *Brain Research , 1439*, 54-65.

Rangel, A., & Hare, T. (2010). Neural computations associated with goal-directed choice. *Current opinion in neurobiology , 20* (2), 262-270.

Rao, R. P., & Ballard, D. H. (1999). Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects. *Nature neuroscience , 2* (1), 79-87.

Rao, R. P., & Sejnowski, T. J. (2001). Spike-timing-dependent Hebbian plasticity as temporal difference learning . *Neural computation , 13* (10), 2221-2237.

Raz, A., & Campbell, N. K. (2011). Can suggestion obviate reading?Supplementing primary Stroop evidence with exploratory negative priming analyses. *Consciousness and Cognition , 20* (2), 312-320.

Reyna, V. F. (2012). A new intuitionism: Meaning,memory and development in Fuzzy-Trace Theory. *Judgment and Decision Making , 7* (3), 332-359.

Reynolds, B., De Wit, H., & Richards, J. B. (2002). Delay of gratification and delay discounting in rats. *59* (3), 157-168.

Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature , 413*, 67-70.

Ruge, H., & Wolfensteller, U. (2010). Rapid Forumlation of Pragmatic Rule Representations in the Human Brain durign Instruction-Based Learning. *Cerebral Cortex , 20*, 1656-1667.

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning reprsentations by back-propagating errors. *Nature , 323* (6088), 533-536.

Samanez-Larkin, G. R., Kuhnen, C. M., Yoo, D. J., & Knutson, B. (2010). Variability in nucleus accumbens activity mediates age-related suboptimal financial risk taking . *The Journal of Neuroscience , 30* (4), 1426-1434.

Scherbaum, S., Dshemuchadse, M., & Goschke, T. (2012). Building a bridge into the future: dynamic connectionist modeling as an integrative tool for research on intertemporal choice. *Frontiers in Psychology , 3*.

Schmeichel, B. J., & Vohs, K. (2009). Self-affirmation and self-control: affirming core values counteracts ego depletion . *Journal of Personality and Social Psychology , 96* (4), 770.

Schneider, W., & Shiffrin, R. M. (1977). Controlled and automatic human information processing: I. Detection , search and attention. *Psychological review , 84* (1), 1-66.

Schweighofer, N., Tanaka, S. C., & Doya, K. (2007). Serotonin and the evaluation of future rewards . *Annals of the New York Academy of Sciences , 1104* (1), 289-300.

Sescousse, G., Redoute, J., & Dreher, J.-C. (2010). The architecture of reward value coding in the human orbitofrontal cortex . *The Journal of Neuroscience , 30* (39), 13095-13104.

Silvetti, M., Seurinck, R., & Verguts, T. (2012). Value and prediction error estimation account for volatility effects in ACC: a model-based fMRI study. *Cortex* .

Sloman, S. A. (1996). The Empirical Case for Two Systems of Reasoning. *Psychological Bulletin , 119* (1), 3-22.

Smith, D. V., Hayden, B. Y., Truong, T.-K., Song, A. W., Platt, M. L., & Huettel, S. A. (2010). Distinct value signals in anterior and posterior ventromedial prefrontal cortex. *The Journal of Neuroscience , 30* (7), 2490-2495.

Song, S., Miller, K. D., & Abbott, L. F. (2000). Competitive Hebbian learning through spike-timing dependent synaptic plasticity. *Nature neuroscience , 3* (9), 919-926.

St.John, M. F., & McClelland, J. L. (1990). Learning and applying contextual constraints in sentence comprehension . *Artificial Intelligence , 46* (1-2), 217-257.

Sutton, R. S., & Barto, A. G. (1998). *Reinforcement learning: An introduction* (Vol. 1). Cambridge, MA: MIT press.

Tangney, J. P., Baumeister, R. F., & Boone, A. L. (2004). High Self-Control Predicts Good Adjustment, Less Pathology, Better Grades, and Interpersonal Success. *Journal of Personality , 72* (2), 271-323.

Thorndike, E. L. (1911). *Animal Intelligence.* New York: MacMillan.

Toni, I., Rushworth, M. F., & Passingham, R. E. (2011). Neural correlates ofvisuomotor associations. *Experimental Brain Research , 141*, 359-369.

Trimmer, P. C., Houston, A. I., Marshall, J. A., Bogacz, R., Paul, E. S., Mendl, M. T., et al. (2008). Mammalian choices: combining fast-but-inaccurate and slow-but-accurate decision-making systems . *Proceedings of the Royal Society B: Biological Sciences , 275* (1649), 2353-2361.

Verguts, T., & Notebaert, W. (2008). Hebbian Learning of Cognitive Control: Dealing with Specific and Nonspecific Adaptation. *Psychological Review , 115* (2), 518-525.

Vohs, K. D., Baumeister, R. F., & Schmeichel, B. J. (2012). Motivation, personal beliefs, and limited resources all contribute to self-control. *Journal of Experimental Social Psychology , 48*, 943-947.

Waszak, F., Wenke, D., & Brass, M. (2008). Cross-talk of instructed and applied arbitrary visuomotor mappings. *Acta Psychologica , 127*, 30-35.

Wickelgren, W. A. (1977). Speed-accuracy tradeoff and information processing dynamics . *Acta psychologica , 41* (1), 67-85.

Wikstrom, P., & Svensson, R. (2010). When does self-control matter? The interaction between morality and self-control in crime causation. *European Journal of Criminology , 7* (5), 395-410.

Wunderlich, K., Dayan, P., & Dolan, R. J. (2012). Mapping value based planning and extensively trained choice in the human brain . *Nature neuroscience , 15* (5), 786-791.