Visual Attention and Dopamine in Value-Based Learning



Brónagh McCoy

Visual Attention and Dopamine in Value-Based Learning

VRIJE UNIVERSITEIT

Visual Attention and Dopamine in Value-Based Learning

ACADEMISCH PROEFSCHRIFT

ter verkrijging van de graad Doctor aan
de Vrije Universiteit Amsterdam,
op gezag van de rector magnificus
prof.dr. V. Subramaniam,
in het openbaar te verdedigen
ten overstaan van de promotiecommissie
van de Faculteit der Gedrags- en Bewegingswetenschappen
op donderdag 18 juni om 13.45 uur in een online bijeenkomst van de Universiteit,
De Boelelaan 1105

door

Brónagh McCoy

geboren te Drogheda, Ireland

promotor:	prof. dr. J.L. Theeuwes
copromotor:	dr. T. Knapen

Members of the committee:

Prof. dr. R. Cools	Radboud Universiteit Nijmegen
Prof. dr. H. Slagter	Vrije Universiteit Amsterdam
Prof. dr. S. van der Stigchel	Universiteit Utrecht
Dr. S.J. Fallon	University of Bristol
Dr. R. Lawson	University of Cambridge

Printing: ProefschriftMaken

VU VRIJE UNIVERSITEIT AMSTERDAM

iBBA Institute Brain and Behavior Amsterdam

Contents

Chapter 1: Introduction **9**

- Prioritization 11
- What drives selection? The role of goals, saliency, and selection history 14
- Reward 17
- Pavlovian and instrumental conditioning 19
- Attention and reward 20
- Overt and covert attention 22
- Oculomotor capture and reward 23
- Decision-making and learning 25
- Models of decision-making 26
- Models of learning 29
- Neural correlates of learning 32
- Special case: how dopamine deficits affect learning in Parkinson's disease 35
- Overview of imaging methods 37

Chapter 2: Effects of reward on oculomotor control **43**

Chapter 3: Overt and covert attention to location-based reward **83**

Chapter 4: Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease **113**

Chapter 5: Distractor inhibition and reinforcement learning in Parkinson's disease: behavioural and brain commonalities **169**

Chapter 6: Summary **199**

References **207**

List of Publications **237**

Acknowledgements **239**

Chapter 1

Introduction

Every day our attention is drawn towards an abundance of events occurring in the environment (Theeuwes, Kramer, Hahn, & Irwin, 1998; Theeuwes & Burger, 1998). What captures our attention in such an eventful world? One aspect that has garnered greater recognition in recent years is the experience of rewarding and pleasurable outcomes associated with these events (Anderson, 2019). Over time, the satisfaction you feel when eating a chocolate bar, seeing a close friend, or opening up your favourite software program, leads to greater attentional capture by these and associated items. Imagine you are a coffee-drinker. Due to its inherent bitterness, the rewarding effects of coffee generally begin with a sense of increased alertness and concentration occurring soon after consumption. Over time, this desirable outcome usually leads to increased liking for the taste of coffee itself. Coffee-related objects and events begin to stand out in the environment, drawing your attention to a passer-by sipping from a coffee cup, or easily spotting the street logo of your favourite coffee shop.

Attending to items in our surroundings relies heavily on eye movements (Kowler *et al.*, 1995; Corbetta and Shulman, 2002). Rewarded objects have been associated mainly with faster initiation and perturbation of eye movements towards the object (Milstein and Dorris, 2007; Hickey and Zoest, 2013), indicating a form of prioritization. Typically, in every-day life, visual selection involves saccadic eye movements and fixations, referred to as overt attention (Itti and Koch, 2000). It is, however, possible to direct attention covertly to a location in space without making an eye movement (Posner, 1978). In this thesis I mainly discuss overt attention involving eye movements (see **Chapters 2** and **3**); but occasionally also consider covert attentional selection (see **Chapters 3** and **5**). The role of reward in these attentional processes in humans is still under-examined. The dopamine neurotransmitter system in the brain is known to play an important part in the processing of reward and feedback, primarily in a reinforcement learning context (Frank, Seeberger and Reilly, 2004; den Ouden *et al.*, 2013; Sharp *et al.*, 2016). As such, a revealing way to study the brain and behavioural mechanisms associated with learning from feedback is the explicit manipulation of dopamine availability in this system.

In this thesis, I examine how rewarding objects affect the various properties of eye movements, such as saccade onset latency, curvature, and landing position, and whether reward at certain spatial locations produces similar effects on both overt and covert attentional selection. Given the central role of the dopamine neurotransmitter system in the processing of rewards, I also discuss the function of dopamine in learning from positive and negative events, and how visual representations are affected by dopamine and learning. This research involves the testing of

Parkinson's disease (PD) patients, who take daily dopamine medication to help replenish their low dopamine levels.

Prioritization

The brain has limited capacity for (visual) information processing. As a result, there is a *bottleneck* early in the visual stream when processing incoming information, since the majority of sensory input is irrelevant for the task at hand. A way of dealing with this bottleneck is to select certain parts of the visual input to be processed preferentially. This preferential processing may take numerous forms; for example, it may be based on a particular characteristic of an object such as the round shape, e.g. feature-based attention (Maunsell and Treue, 2006), or on a particular location in the environment, e.g. spatial-based attention (Soto and Blanco, 2004). Once attention has been allocated towards e.g. these round objects, the focus of attention may then be shifted from one object to another in a sequential, serial manner (Itti and Koch, 2000). Objects in the environment must therefore compete for attentional focus and prioritization. The *biased competition* model proposes that behaviourally relevant objects gain preference in this competition (Desimone and Duncan, 1995; Desimone, 1998). It also describes attention as an emergent feature of the neural mechanisms involved in resolving this competition. Neurons differentially encode specific locations in the visual field, based on where light from this location hits the retina, in what is known as the neuron's *receptive field* (RF). Neurons in early visual areas, such as the lateral geniculate nucleus (LGN) and primary visual cortex (V1) have small RFs and code for simple features such as orientation and contrast at a particular location. Objects presented closely together in physical space can therefore both be represented within the same RF, leading to competition. RFs in higher-order regions, such as the lateral occipital complex (LOC), posterior parietal cortex (PPC) and frontal eye fields (FEF) have larger RFs and can code for more abstract or conceptual properties that are in line with behavioural goals (see **Fig. 1**). Larger RFs in these high-order regions also lead to even greater competition, as more features are encapsulated within the same RF. Object properties relevant for a particular task can be represented in a more distributed manner across these regions, e.g. the color of a red ball may primarily be encoded in V4, with a representation of the ball object itself in LOC, as well as a memory of the last time you caught the ball encoded in areas such as the prefrontal cortex and hippocampus. Selective attention has been shown to coordinate these overlapping and distributed neuronal activations in order to resolve the competition for visual awareness (see Serences and Yantis, 2006 for a review). The voluntary

deployment of attention to a feature (or spatial location) both increases sensory gain to that feature and attenuates the neural response to distracting stimuli, thereby giving the attended stimulus a competitive advantage (Kastner *et al.*, 1998; Reynolds and Chelazzi, 2004; Serences *et al.*, 2004).

A prominent theory comprising the necessary components of prioritization is the *feature integration theory* (FIT) of attention (Treisman and Gelade, 1980). According to this, visual spatial attention may be described by a *preattentive* stage whereby parallel processing first occurs across the entire visual field, followed by a limited capacity stage in which attention is focused on only a few items at a time.

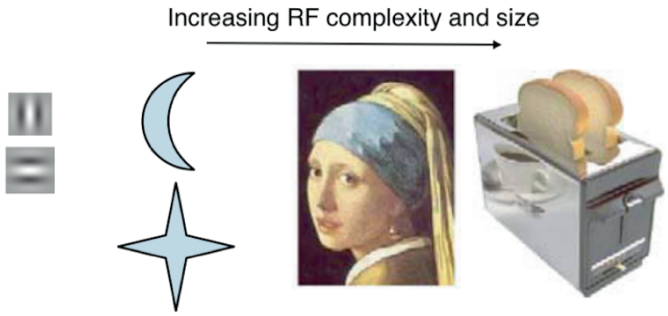


Figure 1. Features that require increasingly larger receptive fields as they are processed higher up the visual hierarchy. Figure from Serences and Yantis, 2006.

Two-dimensional maps representing locations in space have been developed to highlight the competitive process among objects or locations in space. These are a projected representation of *retinotopic maps* of visual space that exist mainly in visual cortex, and come in the form of *saliency* and *priority* maps. The preattentive stage of spatial attention is assumed to be governed by a saliency map that encodes the bottom-up visual saliency of items in the environment. Competition between neurons in this map leads to a single "winning" location that represents the most salient object (Itti and Koch, 2000). This object (location) becomes the target of the next eye movement, drawing it into focus. Once enough information has been gathered at this location by an observer, attention is shifted towards the next most salient object, with neurons encoding the previous location subsequently becoming inhibited, in a phenomenon referred to as *inhibition of return* (IOR) (Posner *et al.*, 1985). Models for such saliency maps have been presented in different forms

(Olshausen, Anderson and Van Essen, 1993; Wolfe, 1994; Treisman, 1998; Itti and Koch, 2000; see **Fig. 2**). Detail on the visual elements that contribute to computations of salience are highlighted in **Fig. 3**. Priority maps extend the concept of saliency maps by integrating additional goal-directed, control signals with physical saliency. Prioritization, the combined representation of salience and behavioural relevance, has been suggested to best describe neuronal firing in response to visual stimuli (Fecteau and Munoz, 2006).

Neuronal responses in the lateral intraparietal sulcus (LIP) of the PPC have been suggested to represent priority maps for visual selection (Bisley and Goldberg, 2010). LIP contains populations of target-selective neurons that respond to visual stimuli and have spatial receptive fields, responding selectively to stimuli that attract attention or gaze (Gottlieb and Kusunoki, 1998; Gottlieb *et al.*, 2014). Target selection responses in LIP neurons have also been shown to be influenced by goal-directed factors, such that if varying rewards are obtained for making saccades to specific locations, LIP responses scale with the value of the planned saccade (Sugrue, Corrado and Newsome, 2005; Kable and Glimcher, 2009). Similar to LIP, the FEF has been associated with target selection (Robinson and Fuchs, 1969; Schall, 1995; Thompson and Bichot, 2005) and is known to have a coherent map of visual space in oculomotor coordinates (Stanton, Bruce and Goldberg, 1993). FEF is also suggested to encode a unique combination of decision-related signals, such that neuronal responses are modulated by choice, visual properties, and reaction time of behavioural responses (Ding and Gold, 2012). These two brain regions are thus functionally well-positioned for the integration of stimulus saliency and goal-directed factors, by orienting the eyes to prioritized locations.

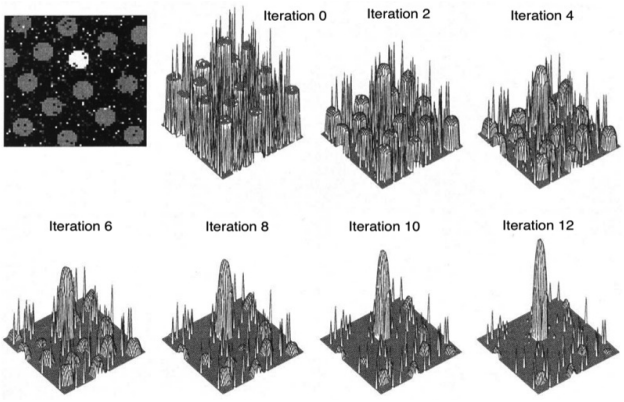


Figure 2. Example of a spatial saliency map. Figure from Itti and Koch, 2000.

What drives selection? The role of goals, saliency, and selection history

Two of the most comprehensively studied topics in attention research are goal-directed (also known as *top-down* or *endogenous*) attention and stimulus-driven (*bottom-up* or *exogenous*) attention. These topics are centered around the use of priority (goal-directed) maps and saliency (stimulus-driven) maps for guiding our attention. They are readily understood from a conceptual standpoint: we all know what it is like to have a mental picture of our house key in mind as we rush around before leaving in the morning (goal-directed), and what is like to have our attention suddenly grabbed by the flashing white light of an approaching cyclist (stimulus-driven). This exogenous capture of attention by salient, i.e. intense, or unexpected stimuli can occur in any sensory domain, e.g. hearing a sudden loud beep of our phone, feeling the sharp heat of a hot pan, smelling the pungent odor of our (least) favorite blue cheese.

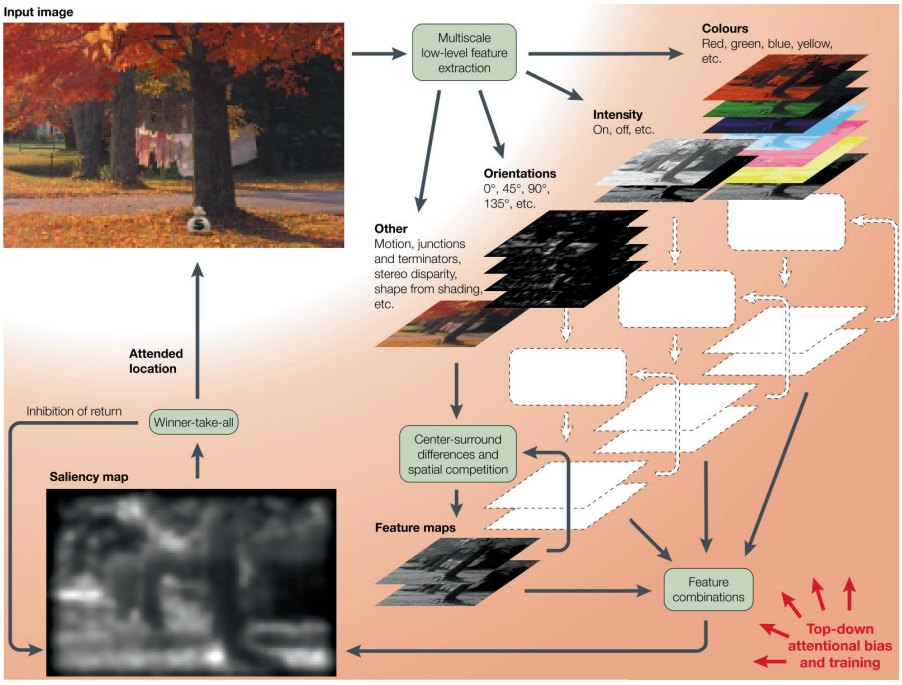


Figure 3. Overview of elements contributing to a saliency map of bottom-up visual attention. Figure from Itti and Koch, 2001.

When we use top-down factors to guide our attention, we draw on information prepared in advance to detect behaviourally relevant aspects of the environment, such as an object's location, motion or color (Eriksen and Hoffman, 1973; Ball and Sekuler, 1980; Posner, 1980). Attention applied in this way is usually volitional and under the control of the observer (Posner, 1980; Treisman and Gelade, 1980; Theeuwes, 2010). Bottom-up guidance occurs in a more automatic fashion, during which attention is captured by something that occurs suddenly or is distinctive e.g. a single red rose in a field of green grass, rather than in a field of colored tulips (Corbetta and Shulman, 2002).The interplay between these two types of attentional guidance depend on current goals, time pressure, and effort required. For instance, if I arrive at a train station to catch a train in five minutes, I have strong (successive) goals to look at the train schedule screen for the platform number, and immediately run towards the platform. In this situation, it is unlikely my attention will be captured by much else. If I have more time to spare, on the other hand, I might spot that colorful new poster outside of the Starbucks, advertising my favourite seasonal gingerbread latte. This in turn may lead to a new goal, at least for the immediate future.

As well as bottom-up and top-down attentional factors, in recent years there has been an accumulation of evidence highlighting the impact of selection history and experience on visual guidance, current choices and judgements (Awh, Belopolsky and Theeuwes, 2012; Braun, Urai and Donner, 2018; Talluri *et al.*, 2018). These newer theories describe how past selection history is combined with current goals and physical salience to form an integrated priority map (see **Fig. 4**; Theeuwes, 2018; for example, an observer's judgement about their environment is not only dependent on the current sensory input or prioritization, but is also biased by the history of preceding choices (Frund, Wichmann and Macke, 2014). Internally-driven, control signals can lead to a statistical influence of previous stimuli and responses on the current trial (Green, 1964). Such internal factors include attentional state, adaptation and learning, and inter-trial dependencies intimately linked with stimulus and response history (Frund, Wichmann and Macke, 2014).

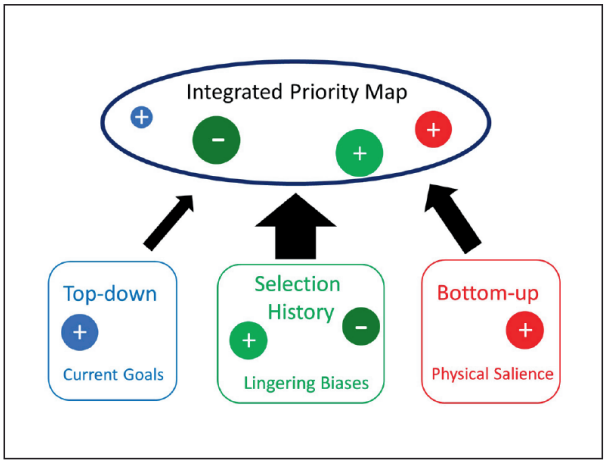


Figure 4. The combined role of goals, physical salience and selection history in prioritization and attentional guidance. Figure from Theeuwes (2018).

Repetitions of the same target stimulus have been shown to bias target selection (Brascamp, Blake and Kristjánsson, 2011; Meeter and Van der Stigchel, 2014) and to result in faster initiation of eye movements to the target (McPeek, Maljkovic and Nakayama, 1999; Becker, 2008). These effects are referred to as *intertrial priming*. Attempts to explain intertrial priming have culminated in two leading accounts: feature-weighting and episodic retrieval (Thomson and Milliken, 2013). Feature-weighting is largely intuitive; it puts emphasis on features associated with a target or positively-reinforced stimulus, and this weighting improves subsequent responses to that feature. This is a short-lived process that dies away quickly across time or trials (Maljkovic and Nakayama, 1994; Thomson and Milliken, 2012). The episodic retrieval account proposes that each trial is stored as a discrete episodic memory and that automatic retrieval of this memory representation affects performance on the current trial. When features of the current perception match those of the retrieved memory, performance is enhanced (Hillstrom, 2000; Huang, Holcombe and Pashler, 2004; Thomson and Milliken, 2013). In contrast to feature-weighting, this process can be context-dependent, and generally displays a more prolonged time course (Kruijne and Meeter, 2015). The episodic retrieval account has been suggested to play a substantial role in long-term priming during visual search; this priming occurs not when the feature is presented as a bottom-up singleton ("pop-out" stimulus), but when the primed feature is presented in conjunction with other features, i.e. in a specific context (Kruijne and Meeter, 2015). This finding corroborates an earlier line of

research on *contextual cueing*, a paradigm describing how stable layout information and/or object covariation is helpful in constraining our focus when an abundance of stimuli in our environment compete for attention (Chun, 2000). Overall, insights from intertrial priming research provide evidence for the importance of selection history on current choices.

Within the framework of a popular statistical method known as Bayesian inference, memories of past experiences can be referred to as 'priors' (Jeffreys, 1998). These priors are integrated with incoming sensory information, thereby influencing current decisions. In a recent perceptual decision-making study, participants performed a random dot motion discrimination task (Talluri *et al.*, 2018). The experiment differed from the standard random dot task format; instead of a single dot motion stimulus per trial, two separate random dot motion stimuli were presented successively. After the first, observers made a choice about the apparent motion direction of the pattern. It was found that these choices biased the sensory evidence accumulation process by selectively altering the weighting (gain) of subsequent evidence during the next pattern presentation. The authors claim that making intermittent choices in this way results in a recurrent interplay between decision-making and selective attention processes. In a separate study also using a random dot motion task, the authors report that when stimulus sequences are dominated by either repetitions or alternations, an individual's performance is predicted similarly well by the individual's degree of adjustment of history bias as by individual perceptual sensitivity (Braun, Urai and Donner, 2018). Other recent studies also highlight the role of *uncertainty* in sensory evidence and how this can influence behaviour (van Bergen *et al.*, 2015; Talluri *et al.*, 2018). For example, degradation of incoming sensory information can adjust the relative weighting on sensory evidence compared to prior experience. These findings emphasize the mutual reliance of these perceptual and selection history signals in decision-making and behaviour. It has recently been suggested that the integration of past experience with incoming sensory information is impaired in Parkinson's disease (Perugini, Ditterich and Basso, 2016; Perugini *et al.*, 2018), hinting at a mechanistic role for the basal ganglia, known to be affected by dopaminergic changes in Parkinson's disease, in this process.

Reward

In its most fundamental form, a reward satisfies a basic need or desire of an organism, e.g. the receipt of food or water when an organism is hungry or thirsty. This is considered a primary reinforcer, in that satiation comes instantaneously through the reward itself. A secondary

reinforcer, such as a monetary reward, does not provide an immediate benefit, but for humans indicates the opportunity for more personalized rewards at some point in the future. The behavioural and neural mechanisms underlying reward-driven feelings of pleasure have received much attention in the literature (Haber and Knutson, 2010; Panksepp, 2011; Damasio and Carvalho, 2013; Anderson and Adolphs, 2014; Berridge and Kringelbach, 2015), not least because of their implications in various types of addiction and in anhedonia, the inability to feel pleasure. The receipt of rewards has been linked to three main psychological constructs (Berridge and Robinson, 2003): (1) learning (both implicit and explicit knowledge), (2) emotion (implicit 'liking' and conscious feelings of pleasure), and (3) motivation (implicit 'wanting' and conscious goals). Motivational wanting processes likely govern the initial appetitive phase leading up to the receipt of a reward, while emotional liking processes dominate the following consummatory phase that satisfies the desire (Berridge and Kringelbach, 2015). A common form of learning, associative learning, describes how a person can pick up on relationships between stimuli or between stimuli and outcomes (Sutton and Barto, 1998; Niv, 2009). This may be *declarative*, in that one has a conscious recollection and can speak about the information acquired, or *procedural,* in which one learns habits or motor skills that usually cannot be consciously remembered or described. Later I will focus on the two main types of associative learning; Pavlovian and instrumental conditioning. The emotional aspect of receiving reward outlines a subjective conscious feeling of pleasure and a more objective 'liking' response. An example of such a hedonic 'liking' reaction is the particular and distinguishing facial expressions that can be commonly observed in humans and animals in response to either the taste of sucrose (sweetness) or quinine (bitterness) (Berridge, 2000; Steiner *et al.*, 2001). Finally, the motivational side of reward receipt concerns either a subjective conscious desire or an objective 'wanting' response, also called *incentive salience*. The notions of 'liking' and 'wanting' were commonly treated as an identical process in the literature, and indeed there are many overlapping aspects. Research delving into these constructs, however, has shown that the associated brain mechanisms can be teased apart under certain conditions (Berridge and Robinson, 1998; Wyvell and Berridge, 2000; Robinson and Berridge, 2003). Incentive salience converts sensory information about rewards and their cues into desired incentives. According to this view, the allocation of incentive salience to a sensory representation is what makes it a 'wanted' target of motivation (Robinson and Berridge, 2003). This 'wanting' response, unlike 'liking', is particularly affected by dopaminergic neurotransmission (Berridge and Robinson, 1998; Dayan and Balleine, 2002). Overall, the explicit and implicit aspects of these three reward constructs (i.e., learning, emotion, and motivation) differentiate how people can be, respectively, either aware their own

reactions to a rewarding stimulus, or can sometimes take actions in response to a stimulus without being aware of the stimulus itself or their own reaction, i.e. people can be either aware or unaware of their own learning/knowledge.

In the brain, responses to reward have been consistently found in the form of a reward prediction error (RPE), in dopaminergic neurons of the ventral tegmental area (VTA) and substantia nigra (SN) in the midbrain (Houk, 1995; Montague, Dayan and Sejnowski, 1996; Schultz, Dayan and Montague, 1997). Signals arising from these regions are projected to the basal ganglia (Beckstead, Domesick and Nauta, 1979; Surmeier *et al.*, 2007) and the prefrontal cortex (PFC) (Deniau, Thierry and Feger, 1980; Swanson, 1982). The orbitofrontal cortex (Schoenbaum, Roesch, M.R. and Stalnaker, 2006; Ostlund and Balleine, 2007) and anterior cingulate cortex (Botvinick, 2007; Hayden *et al.*, 2011) of the PFC have also been shown to carry information about expected outcomes and violations of those expectations.

Pavlovian and instrumental conditioning

Two distinct forms of learning have been established; Pavlovian (or classical) conditioning and instrumental (or operant) learning. Both types of learning rely heavily on *associations*, whether that be associations between stimuli, a stimulus and an action, or an action and an outcome (or indeed, all three together). Pavlovian conditioning (Pavlov, 1927) can be explained by reference to a *Pavlov's dog* anecdote; a primary reward such as food (an "unconditioned stimulus") leads to a salivating response in the dog ("unconditioned response"). If a bell (previously unassociated with anything) is then sounded prior to the presentation of the food, over a course of training, the bell itself starts to lead to an anticipatory, salivating response in the dog (see **Fig. 5**). The bell is referred to as a "conditioned stimulus"; the dog can use this stimulus to make predictions about future, behaviourally significant events. This mechanism will be discussed in more detail later in terms of a reward prediction error framework (see *Models of reinforcement learning*). It is important to realize that this type of stimulus-outcome association is made irrespective of the behaviour of the dog, i.e. he does not need to perform a particular action to be offered the food. Instrumental conditioning, on the other hand, requires a specific action or behavioural response in a given context in order to obtain a particular outcome (Dickinson and Balleine, 1994; Dayan and Balleine, 2002). Actions that are paired with a positive reinforcement are repeated, and those that lead to negative reinforcement, or punishment, become subsequently inhibited. This type of conditioning is closely

related to artificial intelligence models of reinforcement learning (Sutton and Barto, 1998, and see section *Models of reinforcement learning*).

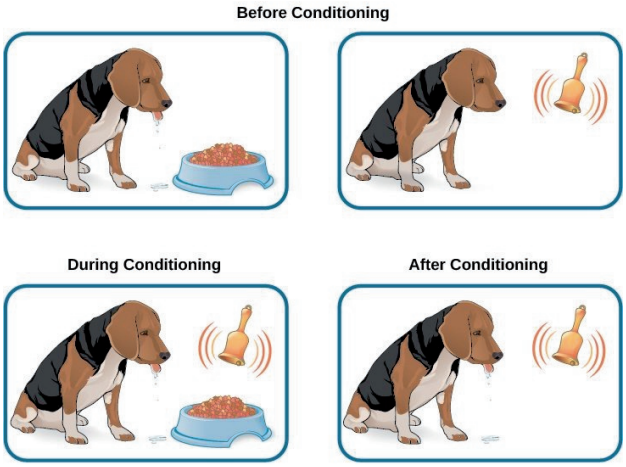


Figure 5. Pavlovian conditioning. Before conditioning, the dog salivates upon the sight of food (reward, a primary reinforcer). The dog may be outside and hear a bell ring in a different context. During conditioning, over an extended period, the bell is rung each time the dog is offered food. After conditioning, the dog salivates whenever the bell ("conditioned stimulus") is rung, now even in the absence of the food reward. Figure from website: www.meetcortex.com.

Attention and reward

The impact of reward on behaviour has been extensively studied in relation to how it affects a host of higher-level cognitive processes. For instance, in the motor/action domain it has been shown to improve response preparation (Mir *et al.*, 2011; Böhler *et al.*, 2014) and response vigor (Opris, Lebedev and Nelson, 2011). To assess how reward influences behaviour during cognitive conflict, researchers used a type of flanker task (Eriksen and Ericksen, 1974) and found increased conflict adaptation after rewarded trials (Braem *et al.*, 2012). Reward incentives have also been demonstrated to improve memory formation (Adcock *et al.*, 2006) and the retrieval of long-term memories (Wolosin, Zeithamova and Preston, 2012), by means of reward-related dopamine signals from the midbrain to the medial temporal lobe, a brain region strongly implicated in memory storage (Squire *et al.*, 1992). In approximately the last decade, researchers have begun to elucidate the influences of reward on visual selective attention. Higher perceptual sensitivity to peripherally-presented targets has been demonstrated for trials in which reward was available as an incentive (Engelmann J and L, 2007; Engelmann *et al.*, 2009). In most of the aforementioned experiments

reward was usually associated with good or correct performance, i.e. it was congruent with the behavioural goal of responding to a target. But what happens if the rewarded stimulus is not relevant for the task? i.e. can we maintain cognitive "top-down" control to perform a task correctly when distracting, rewarding stimuli are present in the environment? In this thesis, we explore this question by examining an array of oculomotor properties as participants attempt to ignore a task-irrelevant, but informative, distractor in the environment (see **Chapter 2**). Several previous experiments linking reward and selective attention have tackled this question, using intertrial priming paradigms (Della Libera and Chelazzi, 2006; Hickey, Chelazzi and Theeuwes, 2010). In their study, Della Libera & Chelazzi used a negative priming paradigm, consisting of separate *prime* and *probe* displays. The prime contained a task-relevant target stimulus, along with a distractor that should be ignored. The subsequent probe also contained a target and distractor stimulus; however, the probe target became the distractor from the prime display, with the same switch for the original distractor stimulus. High or low monetary reward was given after responses to the prime display. This negative priming led to lower accuracy and longer reaction times for the probe target, and this negative priming effect was greater when the prime target was associated with high compared to low reward. Another study made use of switching between target and distractor stimulus features from trial-to-trial in a more standard intertrial priming paradigm, using a version of the additional-singleton task (Theeuwes, 1992). The target was either a circle among similarly-colored diamonds or a diamond among similarly-colored circles, with a salient, differently-colored distractor also presented (Hickey, Chelazzi and Theeuwes, 2010). They found a cost, i.e. lower reaction times, when a highly rewarded target feature on the previous trial became a feature of the distractor in the current trial. An electrophysiological marker, known as the P1 component, assessed during this experiment using electroencephalography (EEG) was shown to have greater magnitude when the feature was associated with high but not with low reward. Source localization indicated that this occurred in extrastriate visual regions, such as the LOC, which is associated with the processing of objects and shapes (Grill-Spector, Kourtzi and Kanwisher, 2001). They also found that the magnitude of visual bias created by reward was predicted by the response to reward feedback in the anterior cingulate cortex (ACC), a medial frontal brain region known to receive dopaminergic inputs. Other research has similarly indicated greater attentional selection for particular perceptual features associated with reward, such as orientation (Serences, 2008) or spatial location (Hickey, Chelazzi and Theeuwes, 2014).

Several of the earlier studies examining the effects of reward incentive on visual selective attention were based on priming. However, negative priming effects might result from the fact that

reward feedback was provided throughout the experiment and therefore merely represents an inability to flexibly adapt to changing task-relevant features from one trial to the next, i.e. the effects of reward may only have been short-term and were mainly due to a deficiency in top-down control. A line of studies by Anderson and colleagues and subsequent studies from other labs tackled this issue by using separate learning and test sessions (Anderson, Laurent and Yantis, 2011; Anderson, Laurent and Yantis, 2011; Failing and Theeuwes, 2014; Le Pelley *et al.*, 2015). During learning, whenever a particular feature was a target, e.g. a red circle, correct responses to it were highly rewarded 80% of the time. In a subsequent test session, reward was no longer available, and the red circle was only ever presented as a distractor in some trials. This arrangement ensured that trial-to-trial flexibility of top-down control was not required and showed that contextually irrelevant stimuli that had previously been associated with reward do indeed capture attention (Anderson, Laurent and Yantis, 2011b).

Overt and covert attention

An important distinction can be made between two types of attentional orienting. During *covert* attention, the eyes are fixated at a particular location in space, but attention is allocated to some other location in the visual field. *Overt* attention is an explicit movement of the eyes to a specific location, in order to place that location at the center of attentional focus, i.e. in the eye's fovea, which processes incoming light in greatest detail. Early work on the extent of similarity between these two types of attention led to the development of the premotor theory, which postulated that covert attention is, in fact, the programming of an eye movement (Klein, 1980; Rizzolatti, 1983). Evidence for the theory comes from studies showing that structures involved in covert spatial attention are also involved in motor programming (Rizzolatti and Gallese, 1988; Rizzolatti, Riggio and Sheliga, 1994). In neuroimaging studies, covert and overt attention have been linked by a common neural architecture in the form of a largely overlapping, shared frontoparietal network in the brain (Corbetta, 1998; Beauchamp *et al.*, 2001; de Haan, Morgan and Rorden, 2008). The frontoparietal network encompasses the frontal eye fields (FEF) in the prefrontal cortex (PFC) and the lateral intraparietal sulcus (LIP) in the posterior parietal cortex (PPC). It has long been established that the parietal lobe plays an important role in the processing of sensory information, decision making, and spatial awareness (Critchley, 1953). An update to the premotor theory has since highlighted a separable effect of externally- and internally-generated cues on the extent of overlap of covert and overt shifts of attention (Belopolsky and Theeuwes, 2012). This research indicates that the shift in covert attention due to both exogenous and endogenous cues is always

accompanied by saccade preparation. However, for endogenous cues the saccade program to the attended location is suppressed shortly after the covert attentional shift has been completed. Contrary to the predictions of premotor theory, there have been several studies reporting the ability to endogenously attend to stimulus locations other than the saccade goal without disturbing the eye movement (Kowler *et al.*, 1995; Montagnini and Castet, 2007), and a lack of facilitation of visual perception for probes presented at a saccade goal (Hunt and Kingstone, 2003). A recent review highlighting the specific aspects of covert and overt spatial attention that adhere to or veer from the predictions of premotor theory, proposes that, overall, there is compelling evidence that in exogenously-driven contexts, e.g. when a salient stimulus is presented, covert attentional capture involves the activation of the oculomotor system (Smith and Schenk, 2012).

Oculomotor capture and reward

Extensive research on the interplay between stimulus-driven and goal-directed attention on oculomotor (eye movement) control has demonstrated bottom-up capture of the eyes by salient stimuli, in spite of task goals (Theeuwes and Burger, 1998b; Theeuwes *et al.*, 1999a; Godijn and Theeuwes, 2002). Exogenous and endogenous control of the eyes were subsequently reported to operate on two different timescales, with fast saccades occurring during stimulus-driven capture and slower saccades occurring during goal-driven attention (Zoest, Donk and Theeuwes, 2004). In terms of the underlying neurobiology associated with eye movements, the superior colliculus (SC) in the midbrain is responsible for the physical generation of a saccade (Sparks and Mays, 1981; Lee, 1988; Schall, 1991). In monkeys, activity in the LIP was found to reflect a general decision variable that was related to the likelihood ratio that the eyes would move to one of two possible locations (Gold and Shadlen, 2001). It has been suggested that LIP activity represents a 'common currency' for integrating information important to the selection of an eye movement, tying in with the more general concept of a 'saliency map" (Sugrue, Corrado and Newsome, 2005, and see earlier section *Prioritization*). Moreover, LIP target selection responses in monkeys have been shown to scale with the value of a planned saccade, when the value is determined by the magnitude, probability or delay of a reward given for making the saccade (Sugrue, Corrado and Newsome, 2005). The production and execution of a saccade involves the broadcasting of information from the visual system to the frontoparietal regions FEF and the LIP (see **Fig. 6**). These areas project to subcortical motor structures, including the basal ganglia, SC, and brainstem motor nuclei (Gottlieb *et al.*, 2014). The basal ganglia are an important reward-sensitive center in the brain (Corbett and Wise, 1980; Fibiger *et al.*, 1987). The caudate nucleus of the basal ganglia, in particular, has been implicated in

orienting the eyes to reward (Hikosaka, Nakamura and Nakahara, 2006). Neurons in the caudate nucleus (CN) are sensitive to spatial position differences in expected reward, leading to a bias in excitability between the superior colliculi so that eye movements to a highly rewarded location occur more quickly. It has been proposed that inputs to the CN carrying both spatial signals and dopaminergic reward-related signals are integrated (Hikosaka, Nakamura and Nakahara, 2006). Another study examining the functioning of both the CN and FEF found that the two regions showed similar reward modulation of visual responses that were spatially selective, but that visual responses specific to expected reward size were much more common in CN than FEF (Ding and Hikosaka, 2006). The tail of the CN was later suggested to play a role in the control of visuomotor skills; when positive reinforcement leads to specific goal-directed actions to achieve that outcome, these actions become repeated (see section on *Pavlovian and instrumental conditioning*), becoming automatic over time, resulting in the formation of a 'skill' (Yamamoto, Kim and Hikosaka, 2013).

In humans, numerous studies have investigated the impact of reward on several properties of eye movements. A study on the influence of the expected value of an object (the probability of receiving reward multiplied by the size of the reward, according to economic theory) on saccadic reaction time (SRT) showed a negative correlation between expected value and SRT, i.e. the time period between stimulus onset and the initiation of the saccade was shorter when the expected value of the stimulus was higher (Milstein and Dorris, 2007). In terms of the eye movement itself, saccade trajectories (curvature) were found to be disrupted by a reward-associated distractor even when participants expected the distractor, knew where it might appear, and tried their best to ignore it (Hickey and Zoest, 2013). Adding a twist to this type of paradigm, a separate study showed how distractors that were never rewarded in and of themselves but that signaled the availability of reward were found to attract the eyes, whereby the end point of saccades was located closer to the distractor for larger reward availability (Bucker, Belopolsky and Theeuwes, 2014). Another study also incorporating this altered setup in which distractors signaled reward availability found that distractors that signaled high reward led to greater oculomotor capture, i.e. a greater proportion of saccades, than those signaling low reward, and this was particularly evident for faster saccade latencies (Failing *et al.*, 2015).

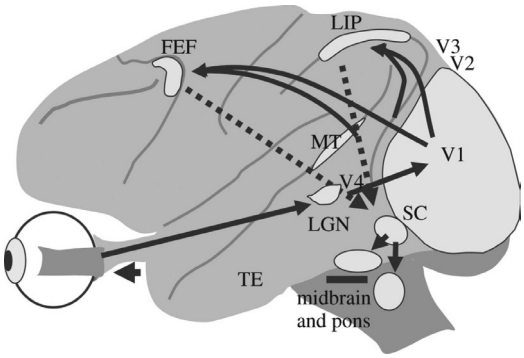


Figure 6. Oculomotor circuit for visually-guided saccade generation (monkey brain). Circuit extends from retina to primary visual cortex (V1) to the lateral intraparietal (LIP) area and frontal eye field (FEF). From cortex, information reaches the superior colliculus (SC), and from there to brainstem oculomotor areas in midbrain and pons. This is a simplified brain circuit, and does not show related circuits that include basal ganglia and the cerebellum. MT, middle temporal cortex; LGN, lateral geniculate nucleus; TE, anterior inferior temporal cortex. Figure from Wurtz, 2015.

Decision-making and learning

Every day we are confronted with alternative options to choose from, ranging from the relatively inconsequential, such as which flavor of tea to pick up at the supermarket, to more life-altering options, such as where to buy your new home or which career path to follow. We must therefore make decisions regularly, over both short and long timescales. Although decisions can be taken based on factual knowledge, such as 'knowing' not to walk across a busy road at rush hour, a large proportion of our decisions are *value-based*, in which a comparison is made between two or more options based on the (subjective) values those options have acquired over time. This has been formulated by economic and computational theory in terms of the encoding of expected *utility* (expected future reward) of the options to help select the one with the highest value (von Neumann and Morgenstern, 1944; Montague, Dayan and Sejnowski, 1996; Gläscher, Hampton and O'Doherty, 2009). A principle component of value-based decision making is learning. This involves the updating of option representations, valuation and actions, based on evaluation of the outcomes of previous choices (Rangel, Camerer and Montague, 2008). The study of value-based decision-making and learning has primarily been embedded in a computational framework, by constructing (optimal) models of how an agent learns from feedback and uses this learning to guide future actions.

Models of decision-making

In the extensive field of decision-making, there is a similar analogy to the bottom-up and top-down dichotomy presented earlier on the mechanisms of attention. This distinction has been nicely described by Daniel Kahneman in his book "*Thinking, Fast and Slow*" (Kahneman, 2011), in which he highlights the automaticity, speed and effortlessness of a *System 1*, in comparison to the deliberate, complex and effortful workings of *System 2*. In this framework, System 1 functions continuously in the background, orienting our attention, alerting us to threatening stimuli, and quickly recognizing (familiar) shapes and objects in the environment. In this way, it is heavily reliant on selection-history, driven by experience. System 2, on the other hand, is slow and effortful, requiring a lot of energy to perform complex computations. It therefore usually functions in something more akin to a "stand-by" mode, consuming little energy if possible, until required to take over. System 1 works constantly, passing impressions and intuitions on to System 2. Usually System 2 accepts these suggestions with few adaptations and, in this way, many of our everyday actions are facilitated primarily by System 1. When System 1 finds a problem too complex to solve, System 2 takes over. This interplay between System 1 and 2, like with the distinction in the attention literature, incorporates bottom-up and top-down sweeps of information. A prominent account of how such a process works is detailed by theories of the brain as a "predictive machine" (Clark, 2013). According to Clark, higher cortical levels of the brain make predictions of the future of the environment, i.e. form expectations based on previous experience. Incoming bottom-up sensory information is matched against these expectations, and any mismatch is coded in the form of a *prediction error*. Expectations in higher brain areas are adjusted according to relevant incoming information and new predictions are fed back to early sensory regions. In this way, prediction errors are projected up the cortical hierarchy, and "corrections" to the errors are projected back, in order to dampen or diminish these errors along the hierarchy. Later, I will elaborate on this view in the context of reinforcement learning models.

Drift diffusion model

In two-alternative forced choice (2-AFC) tasks, a decision is made to choose one out of two possible options. Decision-making models have been developed to describe an evidence accumulation process in which the stimulus or choice inputs can have various degrees of noisiness or "uncertainty" (Ratcliff, 1978; Busemeyer and Townsend, 1993; Gold and Shadlen, 2001). Once evidence accumulation reaches a boundary that represents one of the stimulus options, the

associated option is chosen (see **Fig. 7**). This accumulation process is sequential, i.e. information is sampled gradually across time. Sequential sampling models (SSM) can be divided into two categories: diffusion models and race models. Diffusion models are characterized by *relative* accumulation of evidence between two options, i.e. a single decision trace, in a manner that compares evidence for one option over the other at any given moment (Ratcliff and Rouder, 1998). The trace of this process passes a decision threshold (boundary) at a certain point, representing the decision to choose one of the options. Race models consider *independent* evidence accumulation for each option (Smith and Vickers, 1988). The first process to cross the associated decision boundary represents the option to be chosen. Here, and in **Chapter 2**, I will focus on diffusion models. Diffusion models have been used to simultaneously explain numerous empirical measures, including response times, accuracy, and the probability distributions for both correct and erroneous responses (Ratcliff and Rouder, 1998; Ratcliff and McKoon, 2008). These behavioural measures are inherently influenced by a so-called "speed-accuracy" trade-off, and thus, drift diffusion models (DDMs) combine these measures into one model described by several parameters. The parameters include a *drift rate* (v), which is the rate at which the process approaches a decision boundary, i.e. the amount of evidence accumulation per unit time (Ratcliff and Rouder, 1998). A high v signifies an easy decision and a low v signifies a more difficult decision. If the distance between the two boundaries, denoted by the boundary separation parameter (a), is small, then decisions can be made quickly, but errors may be easily made. If a is large, then decisions are more cautious and less prone to error, as it takes longer for the process to reach one of the decision boundaries. A prior bias (z) may also be present for one of the two options. For example, in the case of a choice between a toffee or a lollipop, one might already have a prior preference for the toffee and would be more inclined to choose that on any given day. Finally, the non-decision time parameter (t) describes the total time needed for encoding of the stimuli and the execution of the motor command, e.g. the manual button-press response.

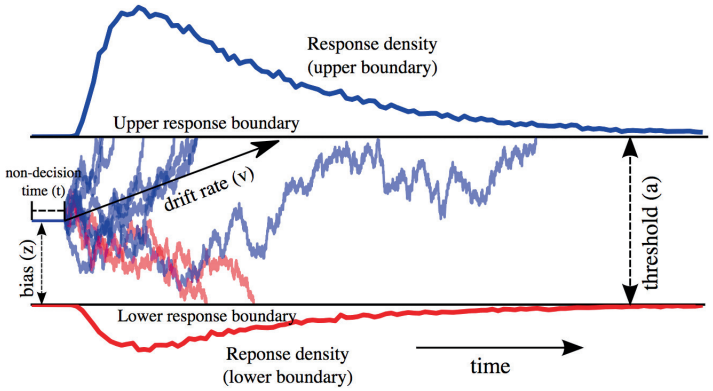


Figure 7. Parameters of a drift diffusion model. Two different distributions (blue and red) at each boundary represent decisions (e.g. responses) that are made for that option. Here, the blue option is chosen more often that red. This could be due to the blue option being the objectively correct option e.g. if it represents the correct direction in a random dot motion task, or the "better" option e.g. in value-based learning. Figure from Wiecki *et al*. 2013.

Hierarchical modelling

Hierarchical models are mathematical descriptions consisting of several parameters that describe a brain or behavioural process, arranged in some form of hierarchically dependence. The range of possible values of some parameters depend on the values of other parameters (Kruschke, 2015). In psychological models of behaviour, this type of modelling can be particularly useful for small sample sizes and low numbers of trials, since it provides a structure to constrain and inform parameter estimation. Hierarchical models can be naturally constructed in a Bayesian framework. Bayesian hierarchical models are particularly insightful, as for each parameter they provide a full distribution of the range of possible parameter values, called the *posterior distribution*. Psychological models not using such a structure generally either fit data for each subject separately, missing out on the statistical strength that can be garnered from similarities across subjects, or else fit models to a group as a whole as if all subjects are the same, thereby failing to take between-subject differences into consideration (Wiecki, Sofer and Frank, 2013). Accounting for such statistical weaknesses, a hierarchical structure facilitates the simultaneous estimation of both group level parameters and individual parameters, at different hierarchical levels (Wetzels *et al.*, 2010; Ahn, Krawitz and Kim, 2011; Lee, 2011; Steingroever, Wetzels and Wagenmakers, 2013). In such a model, it is assumed that individual subject parameters come from a group distribution. The more similar subjects are to each other, the lower the estimated variance in the group distribution

which, at the same time, puts greater constraints on parameter estimates of any given subject. The hierarchical drift diffusion model (HDDM) applies a Bayesian hierarchical framework to the standard DDM mentioned previously (Wiecki, Sofer and Frank, 2013), and has been shown to be particularly useful when the number of observations (sample size or number of trials) is small (Ratcliff and Childers, 2015). In this thesis, both the HDDM and a different form of hierarchical model are applied to data from learning tasks; in **Chapter 2**, the HDDM is applied to eye movement data, and in **Chapter 4**, a hierarchical Bayesian model is applied to behavioural responses in a reinforcement learning task in Parkinson's disease patients.

Models of learning

Decision-making and reinforcement learning processes are intimately linked. During decision-making, one option is chosen over another if it is considered to be correct or of greater value. Learning from outcomes informs decision-making; an agent learns to choose better options based on rewards and punishments received, in an instrumental manner (Thorndike, 1911; Skinner, 1938). Models of reinforcement learning are part of a *normative* framework in which behaviour is examined according to its hypothesized functions; in other words, given a certain goal, what are the computational (or neural) mechanisms associated with us behaving in an optimal or near-optimal way? (Niv, 2009). If such idealized accounts of learning mechanisms fall short in accurately describing behaviour, this is informative in and of itself, since it points towards constraints in our functioning and can highlight ways in which we do not perform optimally.

One of the most influential models of learning in animals and humans is the Rescorla-Wagner reinforcement learning model, which rests on the notion that learning takes place only when an event or reward does not match our expectation, i.e. the event was *unpredicted* (Rescorla and Wagner, 1972). This model involves updating the value or associative strength of a stimulus based on a *reward prediction error* (RPE). Predictions in such a model are made with regards to only the immediately anticipated reward or event. In this thesis, we use the Q-learning model to assess behaviour during learning in PD patients (see **Chapter 4**). The Rescorla-Wagner model was subsequently extended to include an estimation or prediction of the average total amount of reward expected to be received in the future, in what is known as *temporal difference* (TD) learning (Sutton and Barto, 1998). TD learning allows for associations between sequences of stimuli that eventually lead to reward, based on the probability of transitioning from one stimulus (or state) to the next (John O'Doherty *et al.*, 2003; Seymour *et al.*, 2004). This also involves a discounting factor

that accounts for an emphasis placed on more immediate rewards compared to those we receive further into the future. See the section *Neural correlates of learning* and **Fig. 8**, for a depiction of how this is implemented by neurons in the midbrain.

Model-free and model-based learning

The use of the word "model" in the terms *model-free* and *model-based* can lead to confusion. Computational learning models are used to characterize both model-free and model-based learning. The word "model" in these terms actually refers to how the brain tries to make sense of the world, either in a model-free way in which the brain does not have knowledge of how different actions can affect the state of the environment, or in a model-based way in which the brain has knowledge of how actions can affect transitions between states, informed by an "internal model" (Niv, 2009; Daw, 2011). Model-free evaluation is retrospective, integrating value across past reinforcements. In contrast, model-based evaluation is prospective, directly assessing available future possibilities. In this thesis, the reinforcement learning task we administer to PD patients (**Chapter 4**) probes model-free, but not model-based, learning.

Model-free and model-based learning has been proposed to account for the differentiation between habitual and goal-directed behaviour (Daw, Niv and Dayan, 2005). TD learning is a class of model-free learning, which relies on the idea of 'caching'; the association of a stimulus or action with a single, scalar value representing its long-term future value (Daw, Niv and Dayan, 2005). The dopaminergic response to feedback in the midbrain, described in more detail later, gradually shifts back in time to the stimulus or cue that led to (or was predictive of) that outcome (Montague, Dayan and Sejnowski, 1996; Schultz, Dayan and Montague, 1997). In this way, the value associated with a stimulus steadily becomes dissociated from the outcome itself. If there is a change in the type of feedback received for that stimulus, e.g. if it is no longer valuable, a model-free system is slow to pick up on the de-valuation, since its focus has been turned to the stimulus itself rather than the outcome. This highlights a potential mechanism underlying habitual behaviour. During model-based learning, one acquires knowledge of relevant environmental states and actions and, importantly, the transition structure between them (O'Doherty, Cockburn and Pauli, 2017). This involves flexibly making short-term predictions about the immediate consequences of each action in a sequence, allowing for multiple potential stimulus-action-outcome avenues to be encoded at the same time (Daw, Niv and Dayan, 2005). The flexibility of this process means that if a stimulus

becomes de-valued, its expected value can be much more easily updated to reflect this change in the environment. This allows for greater cognitive control and goal-directed behaviour.

Actor-critic model

Since much of our real-world learning is accomplished in an instrumental manner, optimal action selection is a fundamental component in obtaining rewards. Figuring out the optimal action in a given circumstance can be somewhat difficult when trying to deduce which particular actions were indeed the best for long-term outcomes, e.g. in a game of chess (Niv, 2009). This is referred to as *credit assignment;* inferring which actions were crucial to obtaining a reward (Barto, Sutton and Anderson, 1983; Sutton and Barto, 1998). In trying to solve the credit assignment problem, Barto et al. (1983) proposed a model consisting of two main components: an "adaptive critic element (ACE)" and an "associative search element (ASE)". The ACE used a fundamental TD-like (model-free) learning rule, upon which the TD learning model in its current form was established. This component evaluated different states of the environment. The ASE component was formed to select the correct action at each state. This ASE-ACE framework subsequently evolved into the widespread actor/critic models used today. An important aspect of actor/critic models is the incorporation of an explicit *policy*. This is a probability distribution over all available actions at each state. The policy is updated at every timestep via a policy learning rate and the current RPE. The critic element uses TD learning to update state values from trial-and-error experience with the environment, and the same TD RPE is used to train the actor element, which maintains and learns the policy (Niv, 2009).

Q-learning model

Distinct from actor/critic models involving two separate components and policy learning, the *Q-learning* model incrementally learns the *value* of state-action pairs (Watkins, 1989). This model is called "off-policy" because the RPE generated is based on the best possible future action, even if that action is not subsequently taken. According to this model, an agent gradually improves their evaluation of the quality of particular actions at particular states (Watkins and Dayan, 1992):

$$Q(S_t, a_t)_{new} = Q(S_t, a_t)_{old} + \alpha\delta_t$$

where $Q(S_t, a_t)$ is the Q-value of a given action at a given state, δ_t is the RPE, and α is the learning rate, i.e. a weighting placed on the RPE to update the value of a state-action pair. α lies in the range [0,1] and usually varies across agents (individuals) and across different contexts or tasks. In **Chapter 4**, we use the Q-learning model with two separate learning rates to emphasize differentiable learning according to RPEs induced by positive and negative outcomes.

A wealth of other learning theories has also been applied to participant data from learning or decision-making tasks, such as prospect theory in a risk aversion task (Sokol-Hessner *et al.*, 2009), reward-punishment or experience-weighted attraction in a probabilistic reversal learning task (Camerer and Ho, 1999; den Ouden *et al.*, 2013), and theories of model-based learning in a two-step task (Daw *et al.*, 2011; Wunderlich, Smittenaar and Dolan, 2012). It is common practice to fit a number of potential models to the data, and use model comparison procedures to assess the model that fits the data the best. The Q-learning model can be fit in a Bayesian hierarchical framework, as introduced earlier (see **Chapter 4** for such an implementation).

Neural correlates of learning

Signed (positive or negative) RPEs have been linked to the phasic firing of dopaminergic neurons in the midbrain (Montague, Dayan and Sejnowski, 1996; Schultz, Dayan and Montague, 1997, and see **Fig. 8**). These dopaminergic signals project to the ventral and dorsal striatum (nucleus accumbens and caudate nucleus/putamen, respectively). The neural correlates of reinforcement learning mechanisms can be studied in terms of cognitive maps (mental pictures of physical or abstract layouts), stimulus value, outcome identity, and the integration of these aspects (O'Doherty, Cockburn and Pauli, 2017). Several brain regions have been associated with the formation of cognitive maps (spatial and contextual) and the integration of these maps with outcome identity during learning; the hippocampus, the orbitofrontal cortex (OFC), and the posterior parietal (PPC). A robust spatial coding scheme has been found in the hippocampus, with the discovery of place cells (O'Keefe and Dostrovsky, 1971), grid cells (Hafting *et al.*, 2005), and recently, more conceptual, abstract maps (see Behrens *et al.*, 2018 for a review). Research on the behavioural effects of lesions of the OFC revealed its involvement in representing the current location of the animal in an abstract task space, especially when that state needed to be inferred or maintained (Schuck *et al.*, 2016). Neuroimaging studies have furthermore shown that the OFC may also signify outcome identity in response to stimuli that predict the outcome (Howard *et al.*, 2015).

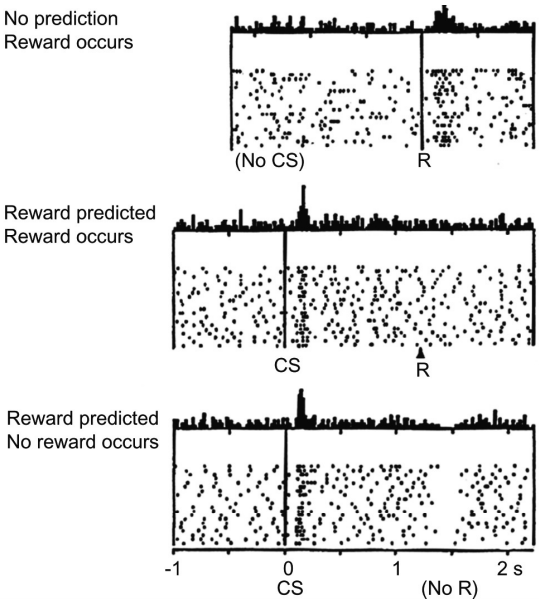


Figure 8. Changes in the firing of midbrain dopamine neurons depending on expectation and whether reward was delivered or not. When there is no prediction and reward is delivered, the neurons fire in response to the reward (top). If a conditioned stimulus (CS) is predictive of reward, the timing of the neuronal response moves back towards the CS itself (middle). If the CS subsequently becomes devalued (is no longer rewarded), there is a decrease in neuronal firing that coincides with the timing of the expected reward (bottom). Figure from Schultz (1997).

This representation could be a potential mechanism through which the expected value of a particular stimulus (the reward probability multiplied by the reward amount) can be estimated (O'Doherty, Cockburn and Pauli, 2017). The involvement of the PPC learning mechanisms has been highlighted in a study showing how it encodes information on the distribution of possible outcomes associated with an action and about the relative probability of obtaining an outcome according to whether a certain action was performed or not (Liljeholm *et al.*, 2011, 2013).

The aforementioned actor/critic models have been extensively linked to instrumental action selection and Pavlovian prediction learning in the basal ganglia (Barto, 1995; Houk, 1995; Joel, Niv and Ruppin, 2002). Three separate cortico-basal-ganglia loops have been suggested to be involved in such learning processes (Balleine, 2005; Niv, 2009). The "limbic loop" includes the regions such as the ventral striatum, basolateral amygdala and OFC, and has been implicated in

Pavlovian prediction learning and evaluation (Cardinal *et al.*, 2002; Holland and Gallagher, 2004). The "associative loop" includes the dorsolateral PFC and caudate nucleus, and has been associated with the goal-directed, instrumental system (Balleine and Dickinson, 1998; Yin, Ostlund and Balleine, 2005). Finally, the "sensorimotor loop" includes sensorimotor areas and the putamen; this has been suggested to play a role in stimulus-response, habitual behaviour (Yin, Knowlton and Balleine, 2004).

Neurotransmitter systems in the brain play an important role during learning and adaptive decision-making. A neurotransmitter is a chemical messenger that transports information from one neuron to another, via a tiny gap (*synaptic cleft*) between the neurons. Receptors on the postsynaptic side of the gap are activated by the neurotransmitter. Neurotransmitter signaling begins in small nuclei of the midbrain, broadcasting to widespread areas of the cortex, such as the prefrontal cortex and striatum, in response to changing environmental context, reward expectation and attentional demand. The most extensively researched neurotransmitters are dopamine, serotonin, and noradrenaline (norepinephrine). Dopamine and serotonin are both involved in learning from reinforcements (Frank, Seeberger and Reilly, 2004; Cools, 2006; Dayan and Huys, 2008; Crockett, Clark and Robbins, 2009) and flexibly updating beliefs when environmental changes occur (Boulougouris, Glennon and Robbins, 2008; Clatworthy *et al.*, 2009; den Ouden *et al.*, 2013). Attempts to disentangle separable roles for dopamine and serotonin in learning have indicated a particular sensitivity of dopamine to positive or rewarding experiences (Rutledge *et al.*, 2009; Voon *et al.*, 2010; Shiner *et al.*, 2012; Smittenaar *et al.*, 2012; den Ouden *et al.*, 2013) and serotonin to negative or punishing outcomes (Dayan and Huys, 2008; Crockett, Clark and Robbins, 2009; den Ouden *et al.*, 2013); however, this is still a controversial topic, as dopamine has also been implicated in the processing of negative outcomes only or in both types of learning (Cools, Altamirano and D'Esposito, 2006b; Bódi *et al.*, 2009; Palminteri *et al.*, 2009; Maril *et al.*, 2013). The noradrenergic system has been associated with general arousal and cognitive flexibility (Berridge and Waterhouse, 2003; Aston-Jones and Cohen, 2005; Bouret and Sara, 2005; Sara and Bouret, 2012), and is suggested to promote long-term synaptic plasticity (Harley, 2004). Increased noradrenergic activation during cognitive shifts can lead to rapid behavioural adaptation to changing environmental factors (Bouret and Sara, 2005). Furthermore, noradrenaline has been implied in the processing of "unexpected uncertainty" due to unanticipated changes in task context (Yu and Dayan, 2005). Although the noradrenergic neurotransmitter system is not primary to reinforcement learning, it may nevertheless play an important role under changing conditions, such as learning in volatile environments (Browning *et al.*, 2015).

Special case: How dopamine deficits affect learning in Parkinson's disease

Parkinson 's disease (PD) is a disease of the central nervous system. It is traditionally characterized by abnormal motor symptoms such as bradykinesia (slowness of movement), tremor and rigidity.

Patients typically present with such motor symptoms, and these remain the primary target of clinical interventions. Nevertheless, it has also been recognized that PD patients suffer from additional non-motor cognitive dysfunctions, including planning, flexibility, and learning from reinforcements (Alexander and Crutcher, 1990; Owen *et al.*, 1992; Cools *et al.*, 2001; Frank, Seeberger and Reilly, 2004). These dysfunctions have increasingly become the focus of study (Bosboom, Stoffers and Wolters, 2004). Underlying the changes in behaviour observed in PD are abnormalities of the mesolimbic dopamine system. PD is accompanied by a loss of neurons in midbrain dopaminergic nuclei (Fearnley and Lees, 1991), leading to decreased dopamine levels in target regions such as the striatum and frontal cortex (Lotharius and Brundin, 2002). Dopamine replacement medications, such as levodopa, attempt to restore dopamine signalling to 'healthy' levels, however, this can be accompanied by large individual variation in motor and cognitive responses to the different dopaminergic medications available. Long-term use of dopamine medications may furthermore lead to an imbalance in striatal dopamine in what is known as the dopamine overdose hypothesis (Swainson *et al.*, 2000; Cools *et al.*, 2001), whereby the more depleted dorsal striatum (occurring early in disease progression) is brought back up to 'healthy' levels, whereas the relatively unimpaired ventral striatum becomes 'overdosed' with dopamine.

In assessing reinforcement learning mechanisms in PD, two different cognitive tasks have generally been utilized; a probabilistic reversal learning (PRL) task (Cools *et al.*, 2001, 2007; Graef *et al.*, 2010) and a probabilistic selection task (PST) (Frank, Seeberger and Reilly, 2004; Bódi *et al.*, 2009; Shiner *et al.*, 2012). The main premise of the PRL task is that participants learn to switch to choosing a particular stimulus option (out of two options) based on received positive and negative feedback. One option is always substantially better than the other, with this contingency changing (reversing) multiple times throughout the experiment. Reversal-related switch behaviour is examined and is reflects the ability to inhibit previously rewarded actions (Costa *et al.*, 2015). PRL experiments have shown a substantially robust and counterintuitive effect of dopamine medication on performance of PD patients in this task; patients on medication were impaired in reversal shifting compared to patients off medication (Cools *et al.*, 2001), and this impairment was limited to those periods in which

reversals were signalled by unexpected punishment (Cools, Altamirano and D'Esposito, 2006a). This result has been explained in terms of the dopamine overdose hypothesis, as well as in terms of impaired tracking of dopaminergic bursts and dips in neuronal signalling to the striatum when patients are on medication (Frank, 2005). Dopaminergic dips are critical for learning from punishments, however the increase in baseline dopamine levels induced in patients on medication may lead to undetectable dips, e.g. in terms of dipping below a required threshold, thereby impairing the patients' ability to learn from negative reinforcements. This reversal-related activity by dopaminergic medication has been linked to changes in activity in the nucleus accumbens, the ventral part of the dorsal striatum (Cools *et al.*, 2007). In another common learning task, the PST, patients undergo separate training and test sessions. During training, they learn to choose the best option out of a pair of stimuli based on the feedback received, for three fixed pairs of stimuli. Each pair is associated with a particular reward contingency; for an 'easy' pair, one stimulus, when chosen, is rewarded 80% of the time, and the other stimulus is rewarded only 20% of the time (80:20 reward probability). The contingencies of the remaining two pairs are 70:30 and 60:40. In the subsequent test session, all stimuli from the training session are mixed up to make novel pairs of stimulus options. Feedback is no longer provided and participants have to make their choices based on what they learned from the training phase, which is assumed to have taken on some form of a hierarchy in terms of most to least rewarding stimulus. Dopamine-related differences due to medication status in PD have been most consistently found in the test phase of this task (Frank, Seeberger and Reilly, 2004; Frank, 2007) by means of an interaction between medication status and stimulus type; patients on (compared to off) medication are more likely to choose the overall best stimulus whenever it is presented, and patients off (compared to on) medication are more likely to avoid choosing the overall worst option whenever it is presented. Dopaminergic-mediated differences during learning in PD for this task have not been demonstrated, although, in line with findings from the PRL task, there has been an indication of negative feedback leading to a change (switch) in choice behaviour on the next trial less often in patients on compared to off medication (Frank, Seeberger and Reilly, 2004). Dopaminergic manipulation studies using this task in healthy people have implicated dopamine-related differences in the ventral striatum, by means of reward prediction error modulation of brain signalling (Pessiglione *et al.*, 2006; Jocham, Klein and Ullsperger, 2011).

Overview of imaging methods

Of particular interest in this thesis is the estimation of brain activations associated with parameters that describe behaviour according to a computational reinforcement learning model. The *Q-learning model* (see previous section) can provide per subject trial-by-trial estimates of RPE. These trial-by-trial model predictions can be compared to fMRI data timeseries, thereby revealing brain regions sensitive to RPE (John O'Doherty *et al.*, 2003; Bayer and Glimcher, 2005; Daw, 2011).

Functional MRI (fMRI) is used to measure brain activity, by taking advantage of increased neuronal demand for oxygen. When particular brain regions are involved in performing a computation, they require more oxygen, leading to increased blood flow to the region. Brain activity in fMRI is measured using blood oxygenation level dependent (BOLD) contrasts. Oxygen is transported in the body via red blood cells (hemoglobin). Hemoglobin has different magnetic properties depending on the amount of comprised oxygen (Ogawa *et al.*, 1990). The BOLD contrast is the ratio of oxygen-saturated hemoglobin (oxyhemoglobin) and hemoglobin lacking in oxygen (deoxyhemoglobin), with brighter parts of the contrast representing oxygen-rich regions. It is important to note that fMRI is an *indirect* measure of neuronal activity, as it highlights regions of increased oxygen demand across large ensembles of neurons.

Statistical modelling of fMRI experiments is performed using a general linear model (GLM). A user-defined timeseries is constructed that includes timings and durations of a particular task condition, e.g. the brief presentation of a car stimulus, and this is combined (*convolved*) with an HRF to estimate the expected BOLD response to stimulus presentation. For each brain voxel, this expected response is compared to the actual output fMRI data timeseries by means of the GLM. This also includes an error term to account for noise and unexplained variance. Solving the GLM requires a method of minimization (reducing the difference between the actual and modelled data) to give the best possible fit. From this, one can extract estimates or weightings (*beta* values) per voxel that denote the extent of activation of that voxel related to the processing a particular stimulus or event. Since this is carried out on a voxel-by-voxel basis, from a statistical point of view, it is necessary to correct for multiple comparisons across all voxels in the brain. Multiple nearby voxels need to show an effect in the same direction for a region to be considered statistically meaningful (Friston, Jezzard and Turner, 1994; Friston *et al.*, 1995; Worsley and Friston, 1995). In the literature, activation results are commonly reported as *univariate* activity using this method. In a later section, I give an overview of a separate *multivariate* analysis technique that makes use of patterns of activation across voxels.

Regions of interest

In addition to carrying out a whole-brain analysis, which requires conservative statistical procedures due the multiple comparison problem, many fMRI studies focus on specific regions of interest (ROIs) (Kanwisher, McDermott and Chun, 1997; Kourtzi *et al.*, 2003; Hickey and Peelen, 2015; Van Loon *et al.*, 2016). ROIs can be defined either *structurally* or *functionally*. Structural ROIs are usually based on the automated labeling of a subject's own anatomy (Fischl *et al.*, 2002) or on probabilistic group atlases (Hammers *et al.*, 2003; Eickhoff *et al.*, 2005; Pauli, Nili and Tyszka, 2018). Functional ROIs can be obtained by using a separate *localizer* scan, a separate task set up specifically to perform within-subject contrasts that reveal brain regions involved in a particular function, e.g. the processing of faces as opposed to any other visual objects. In this thesis, I use both functionally and structurally defined ROIs when assessing dopamine-related brain activation differences in PD patients during a reinforcement learning task (see **Chapter 4**), and also when examining grey matter volume changes in PD (see **Chapter 5**).

Quantification of the signal within ROIs can be done in several ways. The most common are done by performing an analysis based on the *hemodynamic response* or by extracting the *percent signal change* (PSC). When using the hemodynamic response method, the raw fMRI timeseries is used, i.e. before any GLM is performed, and the full hemodynamic response to each condition across the ROI is estimated (Poldrack, 2011). This can be done using a finite impulse response (FIR) model that calculates the response at each timepoint and makes no assumptions about the shape of the hemodynamic response (Dale, Fischl and Sereno, 1999), or by using other similar models that reduce the number of free parameters, thereby potentially leading to a more robust fit, e.g. with a Fourier basis set (de Hollander and Knapen, 2017). In the other method, PSC is calculated by dividing the BOLD response to a particular stimulus (or condition) by the overall mean of the BOLD time series. The BOLD response to a condition is based on the associated parameter estimate calculated from a GLM, as well as the baseline to maximum signal change during that condition.

Representations in the brain

Earlier, I described a *univariate* analysis as an individual-voxel technique, with averaged BOLD activation across an extended region of voxels indicating involvement of that region. Such spatial averaging can, however, blur out fine-grained spatial patterns (Kriegeskorte, Goebel and Bandettini, 2006). It is likely that (visual) information about a stimulus can be encoded as a combination of varying activations *across* voxels. For instance, it is possible for a particular brain region to be

similarly active during the visual processing of a car and chair stimulus, but for the *pattern* of activation across voxels within the region to adequately discriminate between the two types of stimuli (Haynes and Rees, 2006; Mur, Bandettini and Kriegeskorte, 2009), as revealed by a *multivariate pattern analysis* (MVPA), also referred to as "decoding". The implementation of such an analysis draws on methods from machine learning. A region that distinguishes between different stimuli or conditions is said to have informational content, in that it encodes distinctive representations of specific visual features, e.g. orientation (Kamitani and Tong, 2005; Harrison and Tong, 2009), objects (O'Craven and Kanwisher, 2000; Haxby *et al.*, 2001), or more abstract concepts, such as visual imagery (Albers *et al.*, 2013).

The most common method used for this has been *classification*, i.e. the classification or prediction of one of a set of discrete categories. Using this technique, model parameters are obtained that provide the most accurate prediction for new observations, allowing for the generalization from observed to unseen data (Poldrack, 2011). This is a type of supervised learning, in which data is separated into a training and test set. In common practice, a simple linear plane ("classifier") is applied to the training set in multi-dimensional space, i.e. across the voxels ("features"). The classifier learns to distinguish categories based on the relative weightings, e.g. BOLD activation, across the voxels. The same classifier is then applied to the test set and makes predictions about which category the new stimuli belong to. The output is classification accuracy, with a chance level of 50% in the case of a binary discrimination, e.g. between car and chair stimulus categories. In this thesis, I apply multivariate analysis techniques to classify positive and negative outcomes in PD patients (see **Chapter 5**).

Thesis Outline

In this thesis, I focus on the motivational and learning aspects of reward. In laboratory experiments that examine the impact of reward, differing monetary values e.g. large or small, can be used as a motivational incentive, i.e. participants try to perform the task well, driven by a feeling of 'wanting' to receive the monetary reward, as highlighted by the Berridge and Robinson framework (Berridge and Robinson, 2003).

In **chapter 2**, I investigate how a reward-signaling, but distracting, object in the environment affects eye movements to a target object, i.e. how does a task-irrelevant object interfere with a task goal? The distractor was associated with either high, low or no reward and

was placed at a short, medium or long (intra-hemifield) distance from the target. I found that these distractors adjust a host of oculomotor properties, specifically the amplitude, landing position and number of erroneous saccades to the distracting object. This influence was scaled by reward level and location, i.e. a distractor that signals high reward, placed near to the target, demonstrates the greatest ability to interfere with a task goal. These findings indicate that fast saccades are a good proxy for underlying processes that govern attention to reward and the quantification of these effects can be used in combination with other methods to examine the interplay between these phenomena.

In **chapter 3**, I examine whether reward learning from overt eye movements to specific spatial locations can influence covert attention to those locations in a similar manner, i.e. in the form of saccade latencies (overt attention) and reaction times (covert attention). Eye movements to high, low and no reward locations demonstrated increased prioritization of rewarded locations during learning. This did not lead to a similar improvement in target discrimination at the rewarded locations during the covert task. Instead, I found an impairment at the high reward location, as indicated by slowed reaction times. This suggests that there is a strong motivational link between reward and the trained response, overt eye movements in this case, leading to slowing or interference when the active response has been changed.

In **chapter 4**, I investigate the role of the dopaminergic system in learning from positive (rewarding) and negative outcomes, at both the behavioural and brain level. A reinforcement learning task was performed twice by patients with PD – one session took place while ON their standard dopaminergic medication, and the other in a clinically-defined OFF state, for which they withheld from taking medication. Patients OFF medication showed greater learning from negative outcomes. This was accompanied by an increased sensitivity in the OFF state to (positive or negative) RPEs in the dorsal striatum, as well as greater processing specifically of negative RPE in the caudate nucleus. A subsequent task based on the learned value of the stimuli from the training phase, revealed a medication-related interaction in behavioural responses to the most and least rewarding objects, as well as a similar medication-related interaction in BOLD PSC in the caudate nucleus.

In **chapter 5**, using data from the same patient sample as that in **chapter 4**, I assess the role of attentional capture during reinforcement learning, and how both attentional capture and learning measures may relate to changes in grey matter volume in different regions of the brain. In addition, I also examine multivariate activation patterns that represent positive and negative

outcomes in fronto-striatal and visual brain regions, and show dopamine-related differences in the way the brain distinguishes between these opposing outcomes.

Finally, in **chapter 6**, I give an overview of the main findings of this thesis and, combining these with other work, provide an account of how visual and reward-related mechanisms may be integrated to help examine reinforcement learning and decision-making processes.

Chapter 2

Effects of reward on oculomotor control

Adapted from

McCoy, B. & Theeuwes, J. (2016)

Effects of reward on oculomotor control

Journal of Neurophysiology, 116: 2453–2466

ABSTRACT

The present study examines the extent to which distractors that signal the availability of monetary reward on a given trial affect eye movements. We used a novel eye movement task in which observers had to follow a target around the screen while ignoring distractors presented at varying locations. We examined the effects of reward magnitude and distractor location on a host of oculomotor properties, including saccade latency, amplitude, landing position, curvature, and erroneous saccades towards the distractor. We found consistent effects of reward magnitude on classic oculomotor phenomena such as the remote distractor effect, the global effect, and oculomotor capture by the distractor. We also show that a distractor in the visual hemifield opposite to the target had a larger effect on oculomotor control than an equidistant distractor in the same hemifield as the target. Bayesian hierarchical drift diffusion modelling revealed large differences in drift rate depending on the reward value, location and visual hemifield of the distractor stimulus. Our findings suggest that high reward distractors not only capture the eyes, but affect a multitude of oculomotor properties associated with oculomotor inhibition and control.

INTRODUCTION

When we sample visual information from the world around us, we are constantly making saccadic eye movements. Between eye movements there are brief fixations when the image is relatively stable during which information is picked up. We are engaged in continuous eye movement sampling during most of our day when reading text, driving cars, viewing displays and interacting with the world in general. Following a saccade, the fixation of an interesting target image may represent a rewarding experience. The experience of receiving a reward is a primary driver of motivated behaviour among humans and animals alike (e.g., Pessoa and Engelmann, 2010). In the lab, when studying eye movements we usually use discrete-trial sequences of events in which observers have to move their eyes back to the center of the display after each saccade sequence (see for example the oculomotor capture paradigm; Theeuwes *et al.*, 1998, 1999b). In the current study, we investigated the effect of reward on oculomotor behaviour using a novel, continuous eye movement task in which observers had to follow a target around the screen. After each saccade, the next target appeared at a new location without the need to re-fixate at the center of the display. This continuous eye movement design ensured fast saccades, allowing the investigation of oculomotor control in a more realistic environment, eliminating the strong top-down component often seen in standard paradigms using a discrete-trials procedure.

It is well known that reward can improve motivation and learning (Thorndike, 1911). When participants know that performing at an optimal level helps them earn additional (monetary) rewards, they are motivated to perform better, resulting in faster and more accurate responses. For example, Milstein and Dorris (2007) showed that observers make faster eye movements when they can obtain a reward; moreover, saccadic eye latencies were negatively correlated with the relative expected reward value of the target. Similarly, Bucker and Theeuwes (2014) showed that reward-induced motivation improved visual orienting and demonstrated that high reward-induced motivation increased top-down control in attentional cueing.

Recently, however, the effects of reward have also been implicated in tasks that are less susceptible to reward-induced motivation. Several studies have demonstrated that an object that is associated with obtaining a reward may be prioritized for selection in an automatic bottom-up fashion, independent of motivation (Hickey, Chelazzi and Theeuwes, 2010; Anderson, Laurent and Yantis, 2011a, 2011b; Failing and Theeuwes, 2014, 2016; Jahfari and Theeuwes, 2017). By examining eye movements, Theeuwes and Belopolsky (2012) showed that even when no rewards could be obtained anymore, observers made more eye movements to a distractor that was previously associated with high reward compared to one associated with low reward. In their

study, observers made eye movements to either a vertical or horizontal bar appearing among shape stimuli, placed at equidistant clock positions on an imaginary circle. For each participant, depending on the orientation, a correct saccade to the vertical or horizontal bar led consistently to either a high or low monetary reward. In a subsequent separate test phase in which reward was no longer available, participants made an eye movement to one gray circle target appearing among red circles. On a subset of trials, an irrelevant vertical or horizontal distractor bar appeared with the onset of the target, at a previously unoccupied location. Theeuwes and Belopolsky found that as the training phase progressed participants became quicker to initiate a saccadic eye movement to the target when it was associated with high reward compared to low reward, demonstrating a motivation-induced speeding up of saccades. Crucially however, during the following test phase, participants made significantly more eye movements to the previously high rewarded stimulus onset (horizontal or vertical bar) in comparison to the previously low rewarded stimulus, even when these distractors were no longer relevant for the task at hand (see also Hickey and Zoest, 2013). Further studies have demonstrated that eye movements are also made to distractors that signal the availability of reward on a trial, even when making an eye movement to the distractor instead of the target results in the loss of reward (Failing *et al.*, 2015; Le Pelley *et al.*, 2015).

The current study was designed to systematically investigate the effect of reward-associated distractors on oculomotor behaviour. Saccadic eye movements are often used as a tool to help uncover the neural bases of reward choice. Two important brain areas involved in oculomotor processing are the lateral intrapartietal area (LIP) and frontal eye field (FEF). These have projections to subcortical regions, including the basal ganglia (BG) and superior colliculus (SC). The initiation and execution of saccadic eye movements is controlled by the SC, with average responses of the entire population of active neurons determining the direction, amplitude and velocity of a saccade (Sparks and Mays, 1981; Lee, 1988; Schall, 1991). Both LIP and FEF have spatial receptive fields and respond selectively to stimuli that attract attention (Thompson and Bichot, 2005; Bisley and Goldberg, 2010). Visual representation in LIP is rather sparse, with only very salient or behaviourally relevant objects strongly represented (Gottlieb and Kusunoki, 1998). In primate research, when a monkey has to choose between two visually distinct objects each associated with different rewards, single unit recording show that the firing rates of neurons in LIP are proportional to the relative expected subjective value of the objects (Sugrue, Corrado and Newsome, 2004). The BG are needed for voluntary control of body movements, including saccadic eye movements (DeLong and Georgopoulos, 1981). The caudate nucleus (CN), a structure within the BG, has also been associated with reinforcement learning, reward-based decision making, and

the orienting of the eyes to rewarding stimuli (Hikosaka, Nakamura and Nakahara, 2006; Schönberg *et al.*, 2007). The substantia nigra pars reticulata (SNr) of the BG plays an important role in selecting an appropriate action, by releasing the tonic inhibition it holds over the SC, the final station before execution of an eye movement. Disinhibition of the SC by the BG is therefore suggested to occur via a CN-SNr-SC circuit (Hikosaka *et al.*, 2000; Gottlieb *et al.*, 2014).

The goal of the current study was to determine the extent to which distractors associated with reward affect oculomotor control in a continuous eye movement task. In different conditions, distractors associated with reward were presented simultaneously with the target at specific locations near or far from the target location, either within or across visual hemifield. The color of the distractor signaled the availability of the amount of reward available in each trial; however, participants were never explicitly informed of this color-reward contingency. By not informing participants of the association, the top-down component of immediately attending to the distractors was not available. This posed a more natural setting for learning from experience, targeting the concept of reward-learning and history effects in the framework proposed by Awh and colleagues (Awh, Belopolsky and Theeuwes, 2012). Crucially, if participants made a saccade to the distractor, no reward was obtained on that trial. In other words, participants had to suppress making a saccade to the distractor as doing so would have a negative impact on their total reward payment.

There are several classic eye movement effects that have been reported with the typical discrete-trial paradigm. We assess how distractor-related reward contingencies affect these classic eye phenomena while using a continuous eye movement task. First, we examined the role of reward-associated distractors on the *remote distractor effect* (Walker *et al.*, 1997). This effect demonstrates that distractors presented at remote locations increase saccade latency but have little effect on landing position. Second, we examined *the global effect*, which implies that distractors presented in close spatial proximity to the target modulate the landing position while having little effect on saccade latency (Coren and Hoenig, 1972; Findlay, 1982). Third, we investigated *errors* in oculomotor control, referred to as *oculomotor capture* (Theeuwes *et al.*, 1998, 1999b), representing those trials in which observers make saccades towards the distractor instead of the target. These erroneous saccades are assumed to occur in a bottom-up, automatic fashion, against the intentions of the observer (Meeter, Van Der Stigchel and Theeuwes, 2010). Fourth, we examined how reward-associated distractors affected *saccade deviations*. It is known that depending on the condition saccades may either deviate towards or away from distractors (see van der Stigchel and Theeuwes, 2005). Fifth, we wanted to establish the specific role of distractors when presented equidistant

within and across hemifields. It is well known that the neural representation of a target is diminished by the presence of nearby distractors (Reynolds, Chelazzi and Desimone, 1999). Stimuli presented simultaneously are not processed independently but compete for neural representation in a mutually suppressive way (Chelazzi *et al.*, 1998; Torralbo and Beck, 2008; Hickey and Theeuwes, 2011). Finally, we determined whether the effects of reward could be fitted by means of a Bayesian hierarchical drift diffusion model, allowing us to determine how reward value and distractor location are represented by saccade latencies and erroneous behaviour in a probabilistic framework.

MATERIALS AND METHODS

Participants

Twenty-four students of the VU University Amsterdam (mean age 24.8 years ± 3.8 years S.D.) with reported normal or corrected-to-normal vision gave written informed consent to take part in the study. The sample included 8 males (mean age 26.0 years ± 5.3 years S.D.) and 16 females (mean age 24.3 years ± 2.8 years S.D.). Participants were given fixed monetary compensation of €9 for undergoing eye tracking and received a reward bonus of up to €9.90 (mean €7.27 ± €0.93 S.D.) based on their performance during the task.

Apparatus

The experiment and stimuli were created using OpenSesame software for Windows (Mathôt, Schreij and Theeuwes, 2012), based on a PsychoPy backend. Stimuli were presented on a Lacie electron22 blueIV monitor (1024 x 768 resolution, 100 Hz refresh rate), with viewing distance held constant at 60 cm using a chin-rest. Participants were tested in a sound-attenuated, dimly-lit room. Eye movements were recorded using an Eyelink 1000 Tower Mount eye tracker (SR Research Ltd., Canada) with 1000 Hz temporal resolution and <0.01° gaze resolution. An automatic algorithm detected saccades using minimum velocity and acceleration criteria of 35°/s and 9,500°/s^2, respectively. Data were collected and analyzed for the right eye only.

Stimuli

Participants had to follow a black unfilled target circle (0.8° radius) around the screen (grey background, CIE: x = 0.313, y = 0.329, 55.834 cd/m^2), while ignoring a salient unfilled colored distractor circle of the same size. The distractor during the practice session was red (CIE: x=.501, y=.347, 28.20 cd/m^2). During the task, the distractor color was either blue (CIE: x=.190, y=.200,

28.30 cd/m^2), green (CIE: x=.265, y=.490, 27.70 cd/m^2), or orange (CIE: x=.480, y=.465, 27.70 cd/m^2).

Task design

Calibration of the eye tracker was carried out at the beginning of each participant's session. The calibration procedure tested participants' gaze positions for nine fixed locations on the screen, to ensure accurate tracking of the eyes. Participants first completed a practice block of the experimental task consisting of 75 trials. The block began with drift correction at the center of the screen. Drift correction is an automatic procedure carried out by the eye tracker to make sure an observer's eyes have not "drifted" far from the center (as a result of head motion, for example), making the original calibration no longer valid. If the drift correction error is too large, the experimenter is prompted to recalibrate before proceeding. After a variable interval (50 – 250 ms) a target (black circle) appeared at a position on an imaginary circle with an eccentricity of 4.5 visual degrees. A red distractor circle appeared simultaneously with the target, also at this eccentricity. The distractor could take one of three positions relative to the target; at a clockwise polar angle of 30°, 120°, or 180°. Participants were instructed to make a quick saccade to the target circle. During the practice, the cutoff time for the eye to land at the target was 2000ms. This was reduced in the experimental task for each participant, depending on their individual saccade latency (the time between stimulus onset and initiation of a saccade) during the practice session. Since people have different baseline eye movement latencies, the 85th percentile of each participant's unique saccade latency distribution was taken after the practice block and used as a latency cutoff for subsequent experimental trials, so that participants would continue making quick saccades throughout the experiment. Although ensuring fast saccades, the 85th percentile is not too strict, allowing the vast majority of saccades to be considered in time. This method of accounting for differences in baseline latencies has been used in previous similar experiments (Bucker *et al.*, 2015; Le Pelley *et al.*, 2015). Correct trials were those made to the target region of interest (1.6 visual degrees; twice the radius of the stimulus) within the cutoff time. If a first saccade was made to the distractor region of interest (within 1.6°), a sinewave 50ms beep was heard, signaling an incorrect trial. Once the trial was completed the next trial began with the previous target black circle as the new starting point, presented alone on the screen for a random interval of 50 – 250 ms. Before beginning each trial, it was necessary for the eye to be within the previous target (new starting point) ROI. This constituted the continuous nature of the design. After the practice session, participants completed the experimental task of 1800 trials, split into 20 short blocks of 90 trials each. Each block began with central fixation and drift correction. Trials in the main experiment followed the same

procedure as in the practice, with the additional element of reward association; see Figure 1 for the trial sequence and timings.

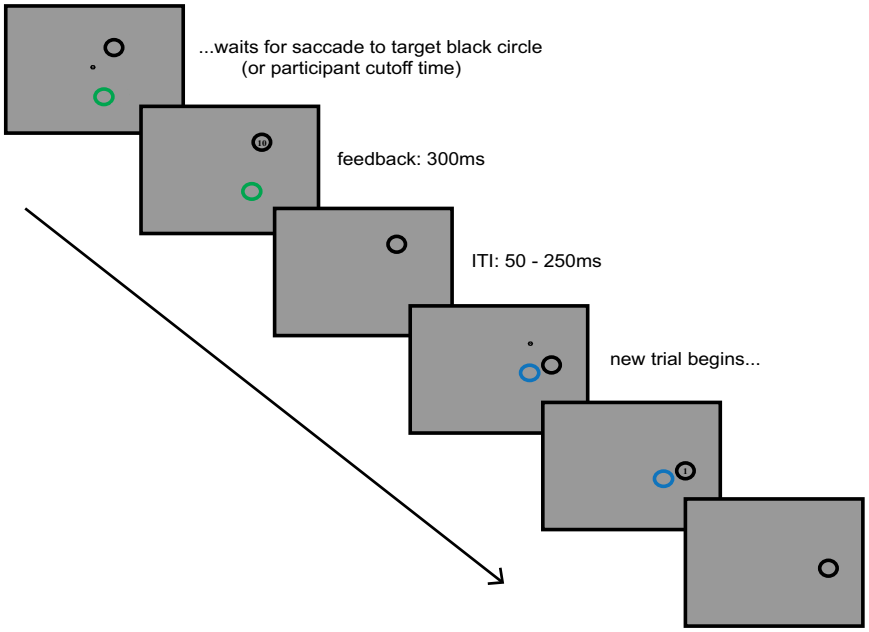


Figure 1. Trial sequence. Participants made a saccade to the target black circle, following it to the next location in every trial. In each trial a colored distractor also appeared. Once a correct saccade to the target was made, feedback was presented (10, 1 or no points). An inter-trial interval of 50-250ms was followed by a new trial. The small black dots appearing here are for illustration purposes only, to show the starting point for each trial. Participants continuously made an eye movement from one black circle to the next across trials. Sizes are not to scale, they are enlarged in the figure for clarity.

Each trial contained a distractor which led to feedback of either high reward (10 points), low reward (1 point) or no reward, associated with the color of the distractor. The blue, green and orange distractor colors were counterbalanced for reward level across participants. When a participant made a correct saccade to the target, feedback was presented for 300ms within the target circle as a '10' or '1' (size 1.08° visual angle) or else nothing appeared, indicating no reward. Participants were not informed of the distractor color-reward association. Importantly, if an incorrect saccade was made to the distractor, a beep was heard and no reward was received for that trial. If participants were slower than the cutoff time, an 's' (1.08°) appeared in the target circle.

The end of feedback presentation constituted the end of the trial and the next trial began in the same way as during practice. Each trial therefore required one saccade to the next target, with the presentation of feedback ending the trial. Figure 2 shows examples of conditions used in the experiment; three possible distractor colors and three distractor positions relative to the target (only the stimulus presentation screens are shown in the figure). All combinations of these colors and positions made up the full set of conditions in the experiment. In order to ensure that the target and distractor did not end up in positions off the screen, a template of possible target positions was created. This took on a lattice structure, with target positions located on one of the four main diagonals from the starting point of each trial (45°, 135°, 225°, or 315° polar angle), always at 4.5° eccentricity. At the end of each block participants received feedback on mean reaction time and points received for that block. At the end of the experiment their total accumulated points was converted into money and displayed on the screen ((points/1000)*1.50 euros).



Figure 2. Trial conditions. For each participant each distractor color signaled the availability of either high, low or no reward for that trial. Distractors could be placed at either 30, 120, or 180 polar degrees from the target. All distractor location-reward combinations appeared equally often for each participant. Sizes are not to scale. The small black dot never appeared in the experiment; it is shown here to illustrate the starting point of the trial.

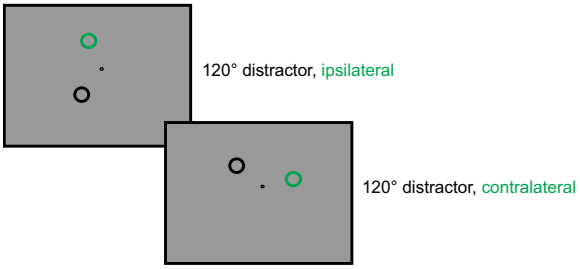


Figure 3. Hemifield display for equidistant ipsilateral and contralateral distractor trials.

Data Analysis

Data were processed and analyzed using Data Viewer (version 2.3.1, SR Research Ltd., Ontario, Canada), Python (version 2.7) and SPSS (version 23.0, SPSS, Inc., Chicago, US) software. Trials from the practice session were excluded and the following properties of all first eye movements in the experimental session were analyzed: saccade latency, erroneous saccades towards the distractor, saccadic amplitude and landing position, and curvature towards the distractor. Saccade latency was described as the time interval between the simultaneous presentation of the target and distractor and the initiation of the saccadic response. Data were analyzed for first saccades, containing no blinks, which started from within 1.6° of the starting point of each trial (twice the circle radius) and landed within 1.6° of the target or distractor circle. Saccade latencies less than 80ms (anticipation errors) and greater than each participant's individual cutoff time (see *Task design*) were excluded from further analysis. If there was any erroneous drift of the eyes within a block, resulting in a portion of hanging trials (in which the script could not allocate the eye as being at either the target or distractor location), then that block of trials was omitted from further analysis. Unless otherwise stated, variance in all analyses is reported as standard error of the mean (SEM).

Latency quartiles

Latency quartiles were calculated for erroneous saccade trials for each individual participant. The percentage of those erroneous trials was calculated per reward condition, taking quartiles from each participant's unique saccadic latency distribution across each reward condition. Thus, percentages of erroneous saccades to high, low and no reward distractors were calculated for the fastest 25%, 25-50%, 50-75%, and the slowest 25% of saccade latencies.

Global effect

The midpoint between the target and distractor centers was taken for all 30° trials. The absolute distance from this point to the landing position of a correct eye movement to the target was expressed as a percentage representing landing closer to or further from the midpoint (positive or negative respectively), with respect to the center of the target.

Saccadic curvature

Saccadic curvature was analyzed by taking all correct first saccades to the target which contained more than 10 samples, i.e. saccades lasting longer than 10ms. The 'main sequence' linear relationship between saccade magnitude and duration (Bahill, Clark and Stark, 1975) predicts that correct saccades in this task, with an amplitude of approximately 4.5°, should last longer than this duration. The calculation of saccadic curvature was based on the peak deviance method (Smit and Van Gisbergen, 1990; Doyle and Walker, 2001; Nummenmaa and Hietanen, 2006). The angular deviation of the saccade path for each 1ms sample was calculated, relative to a straight line from the starting point of the saccade to the saccade endpoint. The sample with the maximum angular deviation (in either a positive or negative direction) was taken as the peak angle. The perpendicular distance from this sample point to the straight line joining the start and end point of the saccade was deemed the peak deviance, *P*. The curvature *C* was thus calculated by dividing the peak deviance *P* by the distance of the straight line from the saccade start to end point, otherwise known as the amplitude, *A* (i.e. $C = P / A$). This correction for amplitude was necessary to account for differences in curvature for given differences in amplitude (Smit and Van Gisbergen, 1990). The curvature was then expressed as a positive (towards the distractor) or negative (away from the distractor) unitless entity.

Visual hemifield

We analyzed the special case of hemifield given a fixed target-distractor distance, i.e. those trials in which the distractor was presented at 120° from the target. Out of all 120° distractor trials, approximately half were ipsilateral and half were contralateral to the target. The physical distance between the target and distractor was the same for ipsilateral and contralateral trials (see Figure 3 for an example of ipsilateral and contralateral distractor trials).

Hierarchical drift diffusion modelling (HDDM)

Drift diffusion modeling (DDM) was also carried out on the data. This technique has been used extensively to explain behavioural data in choice experiments involving two competing items

(Ratcliff and Rouder, 1998, 2000). Since eye movements to a target in the presence of a distractor require a motor plan to the target location, while inhibiting a competing location, tasks of this nature may be described as a speeded decision process at a fundamental level. The DDM is a sequential sampling method that assumes decisions are made by a noisy process that accumulates relative information over time from a starting point to one of two decision boundaries. Once evidence has passed the boundary threshold a decision has been made, and the appropriate motor response is planned accordingly. Since the evidence gathered in aiding the decision is noisy, the process is stochastic and the state of the system is determined probabilistically. The probabilistic nature of this process has led to the recent proposal of a Bayesian approach to the DDM, the hierarchical drift diffusion model (HDDM) (Wiecki, Sofer and Frank, 2013). Parameters of the HDDM allow simultaneous estimation of subject and group parameters, whereby individual subjects are assumed to be drawn from a group distribution. It is therefore possible to use HDDM to infer further information about eye movements under different conditions at the group level, above and beyond standard frequentist analyses of the data.

The speed at which the evidence accumulation process approaches one of two boundaries (e.g. choosing a red or blue item) is called the drift-rate (v) and denotes the relative evidence for or against a particular response. The drift rate is a function of the value difference between two items, so a relatively high drift rate indicates an easy decision, with a lower drift rate indicating a more difficult decision. In the current experiment, each trial required a quick eye movement towards the target. Given that the angle and reward value of the distractor changed across trials, we expected to see differences in the drift rate, representing the disturbing influence of the distractor on the control of saccades to the target. A second important parameter in the HDDM is the separation between the two boundaries (a). If the boundaries are close together, then a noisy accumulation process reaches one of the boundaries quickly, but it is highly likely that some errors are made (a liberal decision). If the boundary separation is large, it indicates the need for greater evidence accumulation before a decision is made, but once a decision has been reached it tends to be more accurate (a cautious decision). Since the current experiment required quick eye movements, with no changes in instruction or stimulus luminance across trials, we hypothesized that the boundary separation would remain constant throughout the task, and that the boundaries would be close together, resulting in a risk of errors. To perform HDDM on our data we used the open source HDDM python toolbox created by Wiecki and colleagues (Wiecki, Sofer and Frank, 2013). Regarding other parameters of the model, we allowed variation in the starting point z of the decision process (z lies in range [0,1]; $z = 0.5$ suggests no apriori bias in the starting point). Non-decision time T_{er} is

the combined time for stimulus encoding (before the decision process begins) and response execution once the decision has been made. We kept T_{er} fixed primarily for practical reasons, as allowing too many parameters in the model to vary can lead to unnecessary complexity in the design. We assumed any between- or within-individual variation in T_{er} to be of a lesser magnitude and importance than differences in other model parameters across our experimental conditions. The behavioural data provided to the full model comprised all correct and error trials from all participants and included response times (ms), response outcome (correct to target or incorrect to distractor), relative angle of the distractor (30°, 120°, or 180°) and reward value indicated by the distractor (high, low or no reward). As with the frequentist data, trials in which the first saccade was anticipatory or late, contained blinks, or did not end at the target or distractor location were not included in the analysis. An additional hemifield model was also run, in which only the 120° distractor trials were taken and labelled as either ipsilateral or contralateral to the target. Again, response times, response outcome, and reward value were taken for each trial, as in the full model, but with hemifield used instead of angle for the hemifield model. For each model 10,000 samples were drawn, the first 1,000 discarded ("burn in"), and every fifth sample from the remaining samples was taken and used to estimate the model parameters. These figures were chosen to ensure convergence of the model, which uses Markov chain Monte Carlo (MCMC) sampling techniques. Since the chains can be erratic at the beginning, the large number of discarded burn in samples means this noisy initial period does not influence the final parameter estimation.

RESULTS

Exclusions

Due to inaccurate drift correction, one block of 90 trials was removed for seven participants, and two blocks were removed for two participants. All 1800 experimental trials were analyzed for the remaining 15 participants. Of this data 78.51% (SD 6.63%) of first saccades were made to the target, 9.47% (SD 4.29%) were erroneous saccades to the distractor, and the remaining 12.02% (SD 5.96%) were to a location that was considered neither target nor distractor. Of these trials 0.75% (SD 1.13%) were discarded due to blinking during the saccade. 1.60% (SD 1.05%) of trials were removed due to saccade onset latencies of less than 80ms, and 6.03% (SD 2.35%) were discarded due to saccade arrival latencies greater than the participant's individual timeout (mean 322 ms ± 40ms SD).

Saccadic latency

The mean saccadic latency of saccades across all participants and conditions was 166 ± 19ms. A repeated-measures ANOVA was carried out on saccade latency of correct saccades to the target, with distractor angle (30°, 120° or 180°) and reward type (high, low, or no) as factors (Figure 4). This revealed a highly reliable main effect of angle ($F(2,46) = 92.72$, $p < .001$, $\eta_p^2 = .80$). Saccades were initiated much faster to the target when a distractor was located at 30° than at either 120° (162.53 ± 3.45 ms vs. 178.47 ± 3.66 ms respectively, $t(23) = 10.35$, $p < .001$) or 180° (180.45 ± 4.05 ms, $t(23) = 9.59$, $p < .001$). Saccade latencies were also quicker when the distractor was at 120° than at 180° ($t(23) = 3.20$, $p = .004$) There was no main effect of reward on latencies, or any interaction between angle and reward ($p > .1$). However, planned paired-samples t-tests on the180° distractor condition revealed a reliable difference in saccadic latency between high and low reward distractors, with slower saccadic latencies to the target when a high reward distractor appeared at the mirror opposite location (180°) compared to a low reward distractor at that location (181.60 ± 4.12 ms vs. 179.54 ± 3.99 ms, $t(23) = 3.45$, $p = .002$). There were no significant differences between high and no reward or low and no reward distractor trials ($p > .1$). A two-tailed paired-samples t-test on saccade latencies to the target versus saccades to the distractor, regardless of angle or reward condition, revealed that saccadic latencies to the distractor were considerably faster (158.92 ± 3.34 ms vs. 173.815 ± 3.56 ms respectively, $t(23) = 8.527$, $p < .001$).

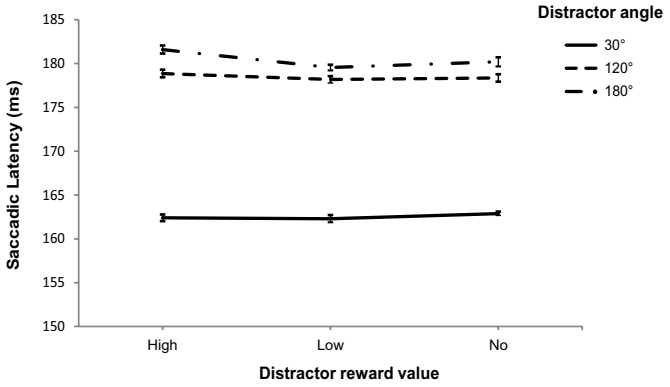


Figure 4. Saccade latency to the target. Error bars show ±1 SEM normalized for within-subject design (Loftus and Masson, 1994; Cousineau, 2005).

Saccadic amplitude

Saccade amplitude is a measure of the size of the saccade in visual degrees. Correct saccades landing directly at the center of the target were expected to have an amplitude of 4.5°, as this was the eccentricity of the starting point to target in every trial. Since it was necessary to begin and end a saccade in a small pre-defined target region of interest for the experiment to continue, measurements of amplitude were taken as the size of the saccade in correct trials, without adjustments for the actual start or end point. Amplitude analysis was carried out on the deviance from the center of the target, with a positive result indicating an overshoot and a negative result indicating an undershoot to the target center. A repeated-measures ANOVA on saccade amplitude deviance with angle (30°, 120° or 180°) and reward type (high, low, or no reward) as factors showed a main effect of angle ($F(1,23) = 8.87$, $p = .001$, $\eta_p^2 = .28$), see Figure 5. The deviance of the target saccade was more negative when a distractor was located at 30° from the target than at either 120° (-.17° ± 0.06° vs. -.10 ± 0.06° respectively, $t(1,23) = 2.65$, $p = .01$) or 180° (-.08 ± 0.06°, $t(1,23) = 3.35$, $p = .003$). Furthermore, saccadic amplitude to the target was significantly shorter when the distractor was at 120° compared to 180° ($t(1,23) = 2.15$, $p = .04$). The ANOVA also revealed a main effect of reward on saccade amplitude deviance ($F(1,23) = 5.86$, $p = .005$, $\eta_p^2 = .20$). At each distractor angle, the high reward distractor resulted in greater undershoot than both a low reward distractor (-.15 ± 0.06° vs. -.10 ± 0.06° respectively, $t(1,23) = 2.85$, $p = .009$) and no reward distractor (-.10 ± 0.06°, $t(1,23) = 3.02$, $p = .006$). There was no difference in amplitude between low and no reward distractor trials ($p > .1$). Overall, eye movements fell slightly short of the target center on average, irrespective of distractor reward or angle.

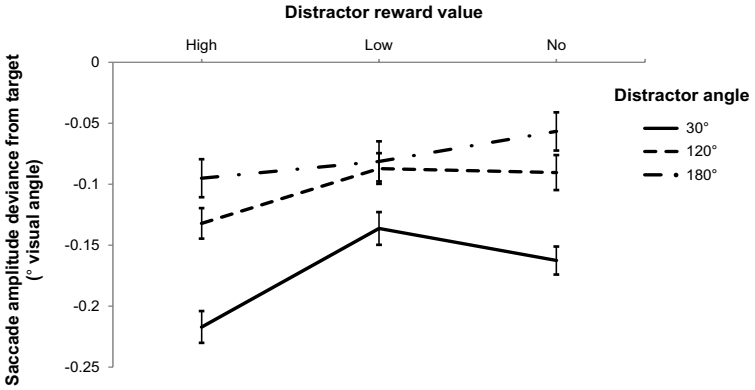


Figure 5. Saccade amplitude deviance from the target. Negative deviance indicates an undershoot in the saccade. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005).

Erroneous saccades

10.11% (SD 4.57%) of all allowed trials (excluding blinks, anticipatory and slow responses) were erroneous saccades to the distractor circle. The percentage of the total erroneous saccades made per participant was taken for each distractor angle and reward combination. The higher the reward–association of the distractor and the closer it was to the target, the more it erroneously captured the eyes, as revealed by a repeated-measures ANOVA (Figure 6). The ANOVA showed both a main effect of distractor angle ($F(2,46) = 13.20$, $p < .001$, $\eta_p^2 = .37$) and a main effect of reward level ($F(2,46) = 6.16$, $p = .004$, $\eta_p^2 = .21$). A distractor at 30° resulted in more erroneous saccades than a distractor at either 120° (14.88 ± 1.04% vs. 10.45 ± .50%, $t(1,23) = 3.03$, $p = .006$) or 180° (8.00 ± .72%, $t(1,23) = 4.00$, $p = .001$). The 120° distractor also led to more errors than the 180° ($t(1,23) = 3.70$, $p = .001$). High reward distractors captured the eyes significantly more than distractors associated with no reward (14.33 ± 1.15% vs. 8.38 ± 0.77%, $t(1,23) = 3.57$, $p = .002$), with a trend for high compared to low reward distractors (14.33 ± 1.15% vs. 10.62 ± 1.02%, $t(1,23) = 1.83$, $p = .08$). There was no reliable difference in erroneous saccades to the low over no reward distractor ($p > .1$).

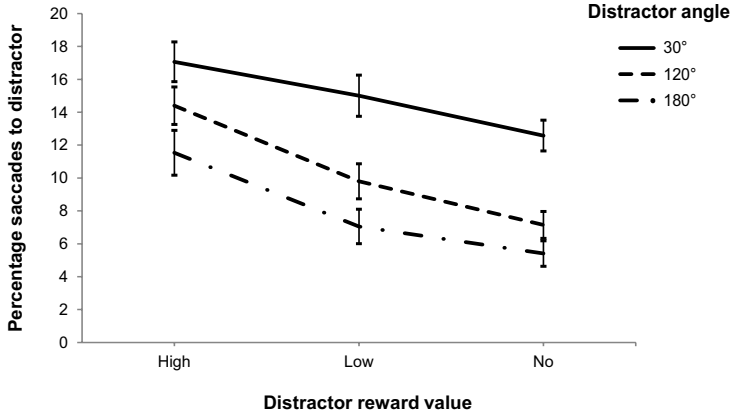


Figure 6. Percentage erroneous saccades to distractor. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Distribution of erroneous saccades across latencies

Percentages of erroneous saccades to the distractor were calculated for the latency quartiles of each individual participant's unique latency distribution per reward condition (Figure 7), as described in *Methods*. A 3x4 repeated-measures ANOVA with reward (high, low or no) and latency bin (fastest to slowest quartiles) revealed a main effect of reward regardless of latency bin ($F(2,46) = 6.15$, $p = .004$, $\eta_p^2 = .21$), with the high reward distractor leading to significantly more erroneous saccades than the no reward distractor across all bins ($t(23) = 3.57$, $p=.002$), and a trend for high over low reward distractors ($t(23) = 1.83$, $p = .08$). A main effect of latency bin was found and expected, given that the data was first split into quartiles ($F(3,69) = 9.20$, $p < .001$, $\eta_p^2 = .29$). There was no reward x latency bin interaction; the shape of each reward level data was similar across bins, suggestive of a simple additive effect of reward on the distribution of the data, especially for high and low reward. It is important to note here that even the mean latency of the slowest saccade quartile was still considerably fast (186.45 ± 4.61 ms, for the high reward condition), which could explain the relatively stable effect of reward across time bins, with no interaction later in time.

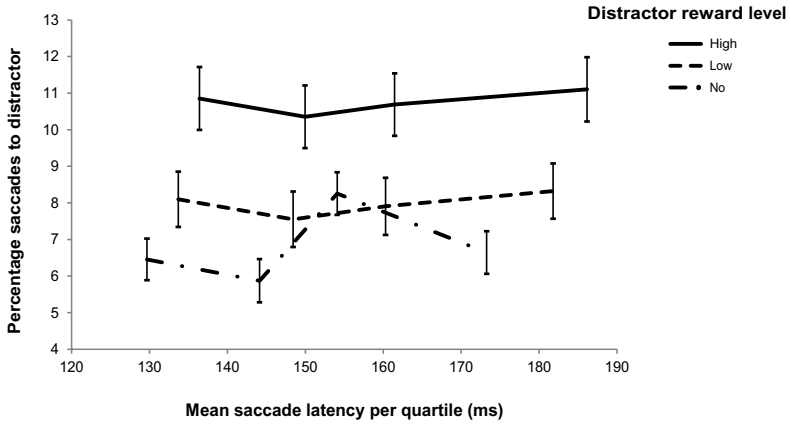


Figure 7. Percentage saccades to distractor per quartile latency bin. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Global effect

The global effect is usually observed when two stimuli are in close proximity, specifically when the polar angle between them is approximately ≤ 30°. Thus, correct 30° distractor trials were analyzed for differences in landing position (see *Methods* for approach). For each participant these were split into either fast or slow saccades, according to the median saccadic onset latency for that participant across all correct 30° distractor trials. A positive percentage indicates a landing position closer to the target-distractor midpoint, thereby indicating a landing position closer to the distractor. A repeated-measures ANOVA was performed on these data, with latency (fast or slow) and reward (high, low, or no) as factors (Figure 8). This revealed a main effect of latency ($F(1,23) = 19.90$, $p < .001$, $\eta_p^2 = .46$) , as well as a main effect of reward ($F(2,46) = 6.78$, $p = .003$, $\eta_p^2 = .23$) , with no interaction between the two. Eye movements were closer to the target-distractor midpoint (thus closer to the distractor) when a high reward distractor was presented compared to both a low reward (5.44 ± 2.37% vs -0.65 ± 3.10%, $t(23) = 3.05$, $p = .006$) and no reward distractor (-0.487% ± 2.36%, $t(23) = 4.21$, $p < .001$), regardless of latency. There was no difference in target landing position between low and no reward distractor saccades ($p>.1$). Faster saccades resulted in a landing location closer to the distractor (7.76 ± 3.31% vs -4.89 ± 2.13%, $t(23) = 4.461$, $p <.001$), irrespective of reward level.

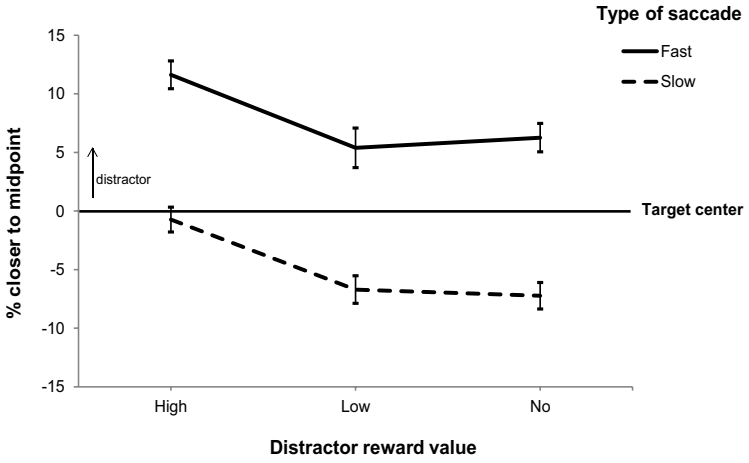


Figure 8. Percentage closer to target-distractor midpoint for fast and slow saccades. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Saccadic curvature

As described in *Methods*, we employed the peak deviance approach to analyze curvature of correct saccades to the target. Curvature was calculated for the 30° and 120° distractor trials only, as it was not possible to define curvature as being either towards or away from the 180° distractor. A repeated-measures ANOVA with angle and reward level showed a significant main effect of angle ($F(1,23) = 29.98$, $p < .001$, $\eta_p^2 = .556$), see Figure 9. The smallest distractor angle of 30° led to less curvature towards the distractor than 120°, irrespective of reward value (0.03 ± 0.01 vs. 0.05 ± 0.01, $t(23) = 5.48$, $p < .001$). There was also a significant interaction between reward and angle ($F(2,46) = 3.70$, $p = .03$, $\eta_p^2 = .14$), influenced mainly by the low reward distractor at 120° resulting in greater curvature than the high and no reward distractors. This effect is described in more detail in the next section investigating the effects of visual hemifield on oculomotor capture.

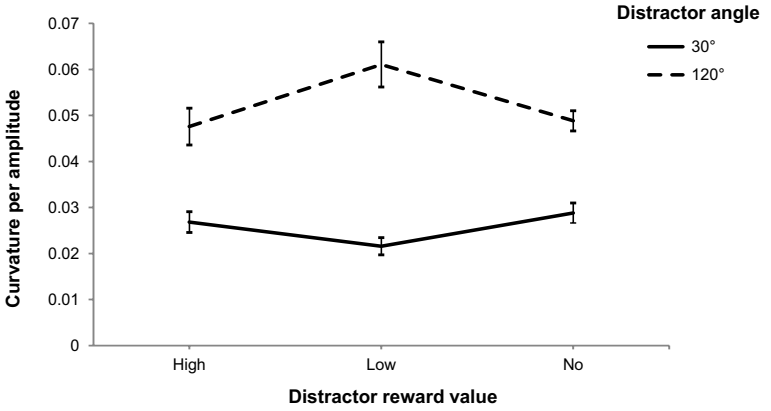


Figure 9. Curvature of saccades towards distractor, for correct saccades to target. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Learning

Given that participants were not explicitly informed of the distractor color-reward contingency, we expected this to be learned as the experiment progressed. For learning analyses the experiment was divided into five main blocks (each containing four mini-blocks, to make up the complete 20 mini-blocks of the experiment). Whether participants learned the association between color and reward magnitude was expected to be reflected in the number of erroneous saccades to the distractor; if an association is learnt one expects the eyes to be captured more by a high value distractor relative to a distractor with lower value. To that end we performed a repeated-measures ANOVA on the percentage of erroneous saccades with distractor angle (30°, 120° or 180°), reward value (high, low or no reward) and block (1-5) as factors. This revealed a main effect of reward ($F(2,46) = 6.15$, $p = .004$, $\eta_p^2 = .21$) and crucially, a reward x block interaction ($F(8,184) = 2.60$, $p = .01$, $\eta_p^2 = .10$), see Figure 10. A main effect of angle ($F(2,46) = 13.00$, $p < .001$, $\eta_p^2 = .36$) and an interaction between angle and block ($F(8,184) = 2.11$, $p = .037$, $\eta_p^2 = .08$) was also visible in erroneous saccades. Two-tailed paired-samples t-tests on reward value for individual blocks, regardless of angle, revealed no significant differences in reward level in the first two blocks, with substantial differences emerging in the third block, particularly between high and no reward (3.03% ± 0.36% vs. 1.89% ± 0.25%, $t(1,23) = 2.35$, $p = .028$). This difference in erroneous saccades towards high over no reward-signaling distractors remained reliable for the rest of the experiment

(2.80% ± 0.26% vs. 1.27% ± 0.16%, t(1,23) = 3.98, p = .001 and 3.54% ± 0.51% vs. 1.40% ± 0.21%, t(1,23) = 3.70, p = .001 for blocks 4 and 5 respectively). These results clearly indicate that participants learned the association between distractor color and reward magnitude, causing them to progressively make more erroneous saccades to distractors signaling high magnitude reward.

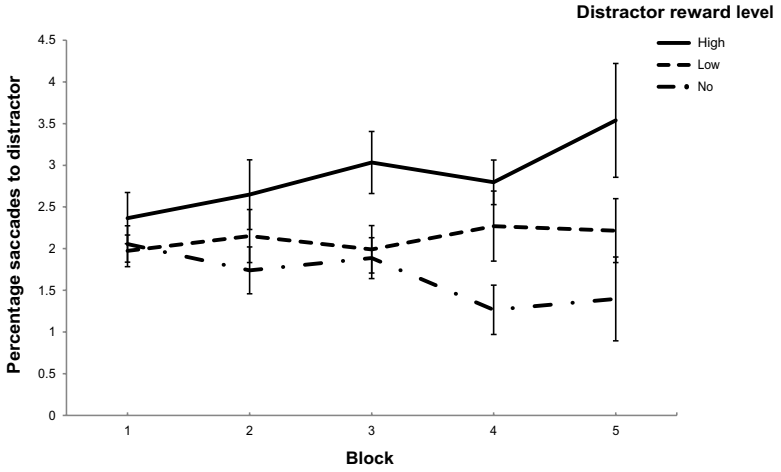


Figure 10. Learning of the distractor color-reward association; percentage erroneous saccades to the distractor across blocks. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Summary of the effects

The results presented above reveal faster saccade latencies for the smallest distractor angle (30°) compared to larger angles (120° and 180°). Distractor reward level only appeared to play a small role at the greatest distractor angle (180°), indicating a slowing down in saccade latency to the target when a high compared to low reward distractor was present. The greatest undershoot in saccade amplitude was seen for the 30° distractor. At each distractor angle, the high reward distractor led to greater undershoot of the target than low and no reward distractors. The greatest oculomotor capture (indexed by erroneous saccades) was also seen for the 30° distractor, followed by the 120° and 180° distractor respectively. There was greatest reward capture by the high reward distractor, at each distractor angle. This effect was robust across all latency quartiles, although all saccades in the task were considerably fast. The global effect was observed, with the eyes landing closer to the distractor for the fastest half of saccades. The landing position was also

closest to the distractor when a high reward distractor was present compared to low and no reward distractors, regardless of saccade latency. Saccadic curvature was greater for the larger compared to smaller distractor angle (120° vs. 30°), with distractor reward value having no differentiating effect on this property. The learning of the distractor color-reward association was reflected in the percentage of erroneous saccades made to the distractor throughout the course of the experiment; participants increasingly made more erroneous saccades towards the high than no reward-signaling distractor as the experiment progressed.

Visual hemifield

Although the physical distance between the target and distractor was the same, the relative hemifield of the 120° distractor led to differences in saccadic latency of correct saccades, percentage of erroneous saccades to the distractor, and saccadic curvature. Total percentages of ipsilateral and contralateral hemifield trials were similar (16.27% vs 16.08% of all experimental trials, p>.1).

Saccadic latency

When a 120° distractor was presented in the contralateral hemifield to the target, there was a significant increase in saccadic latencies towards the target, as compared to ipsilateral hemifield presentation. A repeated-measures ANOVA with hemifield (ipsilateral or contralateral) and reward levels (high, low, no) as factors demonstrated this main effect of hemifield on saccadic latency ($F(1,23) = 20.783$, $p < .001$, $\eta_p^2 = .475$), see Figure 11. Saccadic latencies for the different reward levels were affected in a similar way across hemifield ($p > .1$). A two-tailed paired-samples t-test on saccade latency of erroneous saccades to the distractor when the distractor was in the ipsilateral versus contralateral hemifield, regardless of angle or reward condition, showed that saccadic latencies to the distractor were considerably faster when the distractor was in the contralateral hemifield (158.19 ± 3.23 ms vs. 170.50 ± 4.52 ms for contralateral vs. ipsilateral distractors, $t(23) = 4.321$, $p < .001$).

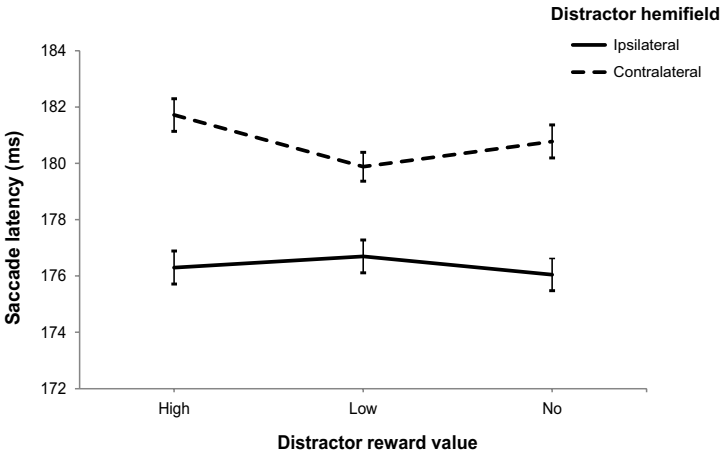


Figure 11. Saccade latency to target for ipsilateral and contralateral distractor hemifield. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Erroneous saccades

A repeated-measures ANOVA on erroneous saccades with hemifield and reward level as factors revealed a similar pattern of erroneous saccades as seen in Fig 6 (see Figure 12). Distractors presented at 120° showed a main effect of reward level on the proportion of saccades made to the distractor, both when the distractor was presented ipsilateral and contralateral to the target ($F(2,46) = 8.67$, $p = .001$, $\eta_p^2 = .27$). Furthermore, distractors in the contralateral hemifield resulted in a significantly greater proportion of saccades towards them than distractors presented in the ipsilateral hemifield ($F(1,23) = 23.79$, $p<.001$, $\eta_p^2 = .51$).

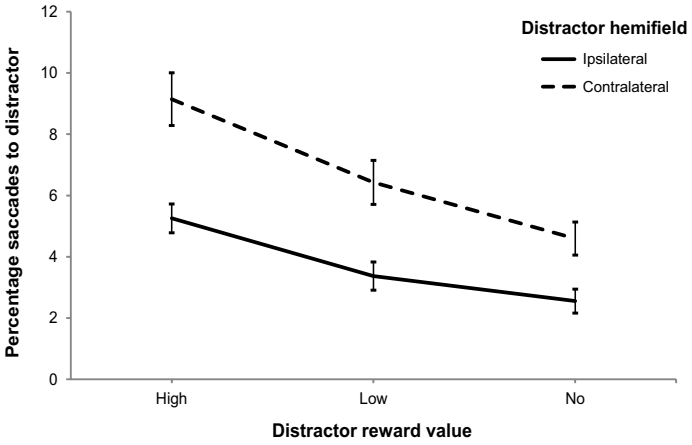


Figure 12. Percentage saccades to distractor for ipsilateral and contralateral distractor hemifield. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Saccadic curvature

A repeated-measures ANOVA on curvature with hemifield and reward level as factors showed that, although the physical distance was the same, trials in which a distractor was presented in the ipsilateral hemifield to the target led to considerably more curvature than trials with a distractor appearing in the contralateral hemifield ($F(1,23) = 6.34$, $p = .02$, $\eta_p^2 = .22$), as seen in Figure 13. This effect was present for all reward levels.

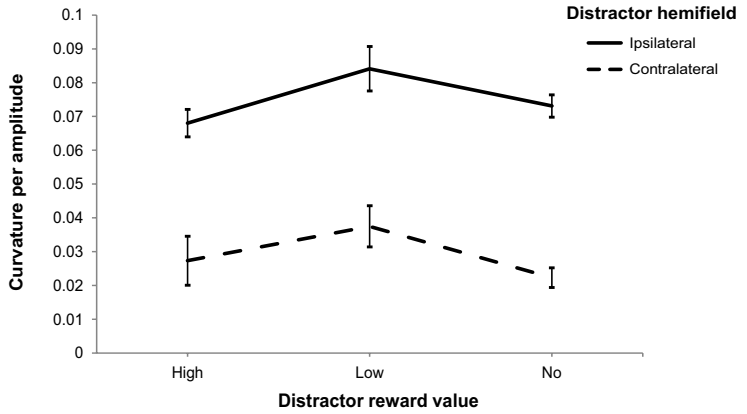


Figure 13. Curvature of target saccades towards distractor for ipsilateral and contralateral distractor hemifield. Error bars show ±1 SEM normalized for within-subject design (Cousineau 2005; Loftus and Masson 1994).

Summary of visual hemifield effects

Distractors presented remote from a target (at 120°) but in the contralateral hemifield resulted in slower saccades to the target, with more erroneous saccades towards the distractor than when distractors were presented at the same physical distance from the target but in the ipsilateral hemifield. Ipsilateral presentation of distractors instead led to greater curvature towards the distractor than contralateral distractors. These findings (more capture and slower saccades) suggest that a remote distractor in the contralateral hemifield has a much larger effect on oculomotor behaviour than a distractor presented at the same distance within the same hemifield. Once the saccade is initiated, there seems to be more competition from the ipsilateral relative to the contralateral distractor as evidenced by the larger curvature for ipsilateral distractors. Note that the effect on latency and curvature does not depend on reward value; while the amount of capture does.

Bayesian hierarchical drift diffusion modeling (HDDM)

Parameters for the HDDM models were estimated according to the details in *Methods*. In a similar approach to the frequentist analyses, one model was based on all data ("full model") and another was based on data collected only for the 120° distractor angle condition, in order to assess the effect of hemifield ("hemifield model").

General parameters

Table 1 shows the results for all parameters of each model (see *Methods* for description), including boundary separation (a), non-decision time (T_{er}), starting point bias (z) and drift rates (v). Due to decisions made regarding the noise parameter in the creation of the HDDM package, the a, z, and v parameters must be multiplied by 0.1 when making comparisons to standard DDM estimates in the literature (e.g., Ratcliff and Rouder, 1998). Similarly, it is also necessary to multiply the z parameter by a to get a comparable starting bias. As expected, the boundary separation a (0.91 and 0.77 for full and hemifield model, respectively) was smaller than in other tasks analyzed with a DDM (e.g., Ratcliff, Van Zandt and McKoon, 1999), since the current task was a relatively easy, fast-paced eye movement paradigm. The starting point z ($z*a$ = 0.23 and 0.22 for the full and hemifield model, respectively) indicated a bias closer to the error boundary, meaning errors were fast compared to correct responses. This is in line with the frequentist results comparing saccade latencies of correct and errors trials. The Bayesian full model estimate of non-decision time T_{er} suggests that approximately 90 ms were spent on combined stimulus encoding and motor response planning for the full model (Table 1).

Drift rate

Posterior probability distributions of drift rate were estimated for each distractor angle and reward combination of the full model (Fig. 14) and for each hemifield and reward combination of the hemifield model (Fig. 15). Hypothesis testing was done directly on the posteriors of the conditions, assessing the probability that one posterior distribution is statistically different from another. Tables 3–6 show these probabilities for comparisons between full model distractor angles and reward levels (Tables 3 and 4) and between hemi-field model distractor hemifield and reward levels (Tables 5 and 6). The term P(X < Y) gives the probability that the posterior distribution of the drift rate for condition X is less than that of condition Y. Comparison between angles in the full model largely reflects earlier analyses on saccadic latencies; there is a high certainty that the drift rate for larger angles (180° and 120°) is statistically lower than that for a smaller angle (30°) at each reward level, meaning that quicker decisions were made in the small-angle condition (Fig. 14 and Table 3). The greatest differences between reward levels were seen particularly in the 180° case [Fig. 14 and Table 4; P(high < no)= 0.875, P(high < low) = 0.816], which was also highlighted in the results on saccadic latency using frequentist analysis. However, the HDDM model additionally indicated differences in reward at smaller angles, with P(high < no) = 0.860 and P(high < low) =

0.781 for the 120° condition, and P(high < no) = 0.687 for the 30° condition. The results suggest slower drift rates (slower decision latencies) in those trials where the high-reward distractor was present, compared with both low- and no-reward trials, in every distractor angle condition. The hemifield model clearly shows slower drift rates when distractors were presented in the contralateral compared with ipsilateral hemifield [Fig. 15 and Table 5; P(contra < ipsilateral) = 0.974, 0.931, and 0.866 for high-, low-, and no-reward distractors, respectively]. This consolidates the saccadic latency results for hemifield found earlier. However, as with the HDDM full model, the hemifield model similarly estimated substantial differences in drift rate for the different reward levels that were not seen in the frequentist analysis (Table 6). When distractors were presented at 120° in either hemifield, the drift rate was lower when the distractor was highly rewarded, especially compared with no-reward distractor trials. This reward effect was stronger when the distractor was presented in the contralateral hemifield [P(high < no) = 0.951 and 0.824 for contralateral and ipsilateral hemifield, respectively]. Results from both HDDM models therefore indicate greater competition from high-reward distractors at larger angles to the target, specifically those presented in the contralateral hemi- field, resulting in longer deliberation time before an eye movement is made.

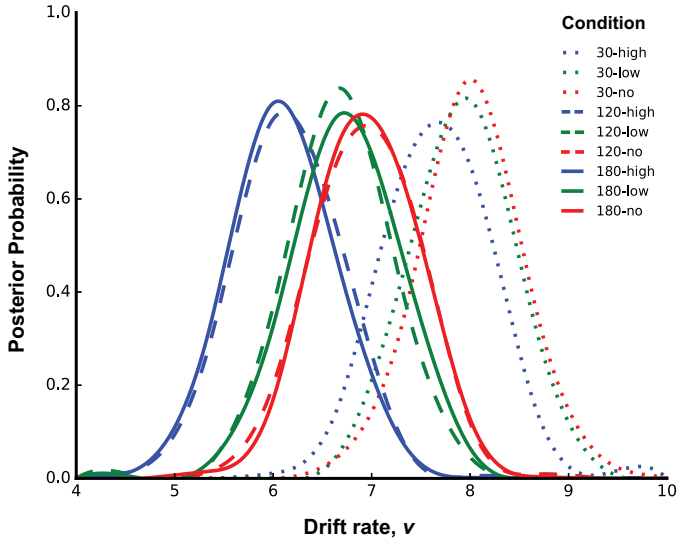


Figure 14. Full HDDM model showing posterior distributions for drift rate *v*, per distractor angle (30°, 120°, or 180°) and reward (high, low, or no) condition.

Model Parameters	Mean
α	0.912
z	0.250
T_{er}	90
v (30.high)	7.593
v (30.low)	7.828
v (30.no)	7.910
v (120.high)	6.237
v (120.low)	6.729
v (120.no)	6.940
v (180.high)	6.191
v (180.low)	6.781
v (180.no)	6.956

Table 1. Full model parameters of the HDDM. The boundary separation, *a*, starting point bias, *z*, and drift rates, *v*, may be multiplied by 0.1 to make parameters comparable to standard DDM estimates in the literature (e.g. Ratcliff and Rouder 1998). The factor of 10 is due to the choice of a scaling parameter used by HDDM.

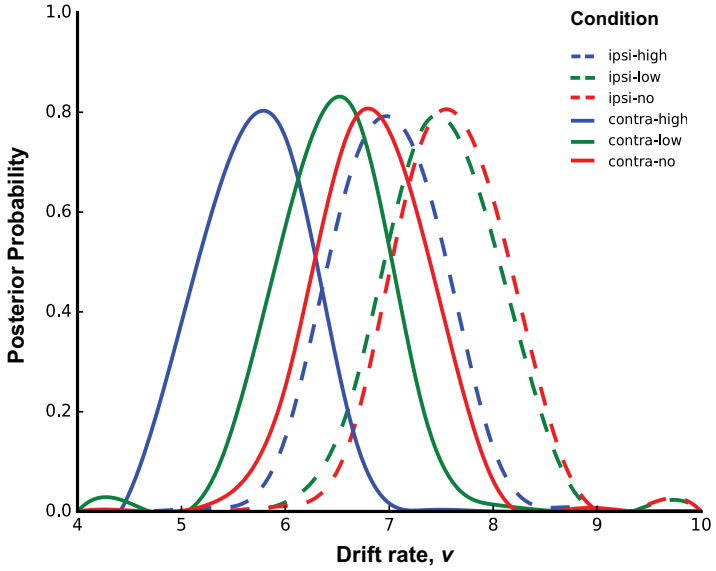


Figure 15. Hemifield HDDM model showing posterior distributions for drift rates *v*, per distractor hemfield (ipsilateral or contralateral to target) and reward (high, low, or no) condition.

Model Parameters	Mean
α	0.766
z	0.291
T_{er}	108
v (same.high)	6.980
v (same.low)	7.454
v (same.no)	7.533
v (opp.high)	5.832
v (opp.low)	6.525
v (opp.no)	6.843

Table 2. Hemifield model parameters of the HDDM. The boundary separation, *a*, starting point bias, *z*, and drift rates, *v*, may be multiplied by 0.1 to make parameters comparable to standard DDM estimates in the literature (e.g. Ratcliff and Rouder 1998). The factor of 10 is due to the choice of a scaling parameter used by HDDM.

Reward	Angle	Probability (X < Y)
High	X: 180°	0.984
	Y: 30°	
	X: 180°	0.532
	Y: 120°	
	X: 180°	0.977
	Y: 30°	
Low	X: 180°	0.944
	Y: 30°	
	X: 180°	0.459
	Y: 120°	
	X: 180°	0.960
	Y: 30°	
No		
	X: 180°	0.924
	Y: 30°	
	X: 180°	0.489
	Y: 120°	
	X: 180°	0.928
	Y: 30°	

Table 3. Probabilities that drift rates are different across conditions for the full model: angle comparison. Probability (X < Y) gives the probability that the posterior distribution of the drift rate for condition X is less than that of condition Y. A value of 0.5 indicates largely overlapping traces. The closer the result is to 1, the more certain that X and Yare statistically different (condition X is slower to reach the boundary than condition Y). Legend also applies to Tables 4–6.

Angle	Reward	Probability (X < Y)
30°	X: high	0.687
	Y: no	
	X: high	0.650
	Y: low	
	X: low	0.548
	Y: no	
120°	X: high	0.860
	Y: no	
	X: high	0.781
	Y: low	
	X: low	0.626
	Y: no	
180°		
	X: high	0.875
	Y: no	
	X: high	0.816
	Y: low	
	X: low	0.614
	Y: no	

Table 4. Probabilities that drift rates are different across conditions for the full model: reward comparison.

Reward	Hemifield	Probability (X < Y)
High	X: contralateral	0.974
	Y: ipsilateral	
Low	X: contralateral	0.931
	Y: ipsilateral	
No	X: contralateral	0.866
	Y: ipsilateral	

Table 5. Probabilities that drift rates are different across conditions for the hemifield model: hemifield comparison.

Hemifield	Reward	Probability (X < Y)
Ipsilateral	X: high	0.824
	Y: no	
	X: high	0.781
	Y: low	
	X: low	0.569
	Y: no	
Contralateral	X: high	0.951
	Y: no	
	X: high	0.870
	Y: low	
	X: low	0.700
	Y: no	

Table 6. Probabilities that drift rates are different across conditions for the hemifield model: reward comparison.

Model comparison and quality

Because a number of models could explain the observed data, we sampled four separate models and compared the deviance information criterion (DIC) obtained for each. DIC is a hierarchical modeling generalization of the AIC (Akaike information criterion) and BIC (Bayesian information criterion). Since we provide the HDDM with information regarding different experimental conditions (e.g., distractor angle and reward level), we can choose which explanatory variables to include in the model. Important model parameters such as drift rate and boundary separation can also vary or remain fixed. Thus, testing numerous models with differing combinations of these variables and parameter limitations allows us to compare and assess which model best fits the data. DIC results in Table 7 confirm that the model that allowed a varying drift rate according to angle and reward condition (our "full model") explained the data best, as indicated by the lowest DIC. The additional model allowing for changes in the boundary separation as well as the drift rate was worse at explaining the data than any of the other drift rate only models.

Model	DIC
v: angle only	*-126723.253*
v: reward only	*-126340.070*
v: angle and reward	*-127307.558*
v: angle and reward; α: angle and reward	*-120430.048*

Table 7. DIC results for four models. Values are deviance information criterion (DIC) for 4 models. v means the drift rate is allowed to change according to the indicated condition; a means the boundary separation may change.

Model convergence

The R-hat Gelman-Rubin statistic was used to assess chain convergence of the model (Gelman and Rubin, 1992). This involved running multiple chains and checking whether they all converged to the same stationary distribution. The statistic then compared between-chain variance to within-chain variance. R-hat values for the three parameters a, Ter, and v were 1.000, 1.001, and 1.000, respectively. These values were all below 1.1, indicating successful convergence of all chains. Visual inspection of the trace plots for parameters a, Ter, z, and all v suggested good mixing of the chains, which moved easily around the parameter space.

Quality of model fit

Posterior predictive checks may be used to assess whether the model can reproduce important patterns in the data. New data were simulated from the posterior of the fitted model, and statistics on these data were compared with the model data to see that a key pattern has been replicated. Using the post_pred_gen() function from the HDDM toolbox, a different data set for 500 parameter values was simulated from the posterior of the fitted model. Numerous summary statistics were then compared (including accuracy, mean of upper/lower boundaries, standard deviation of upper/lower boundaries, and quantiles) and deemed reliable if the comparison fell within the 95% credible interval. All summary statistics comparisons between the respective fitted full or hemifield model and simulated full or hemifield model were credible.

DISCUSSION

We systematically examined the effect of reward-signaling distractors on oculomotor control in a novel, continuous, fast-paced eye movement task. The results show that classic properties of oculomotor behaviour such as the *global effect* and the *remote distractor effect* are affected by distractor reward value. Increased erroneous saccades to the distractor (i.e., oculomotor capture) occurs for distractors of increasing value, even though making these erroneous saccades was detrimental to reward pay-out (see also Failing *et al.*, 2015; Le Pelley *et al.*, 2015 for a similar effect). Even though moving the eyes to the distractor resulted in the omission of reward, observers could not seem to prevent it, providing strong evidence that the reward-signaling distractor captured the eyes against the top-down intentions of the observer. This finding is in line with the theory of Awh and colleagues (Awh, Belopolsky and Theeuwes, 2012) describing the additional component of reward history effects on visual responses, above and beyond the standard theories of endogenous and exogenous influences. Bayesian hierarchical drift diffusion modeling, informed by both individual and group parameters, was used to highlight differences in saccade latency distributions and errors in oculomotor control, depending on the reward value associated with a distractor in the environment. Taken together these findings suggest that even in the very simplest contexts of stimulus processing, reward opportunities and the relative location of reward-signaling distractors play an important role in oculomotor control.

Classic eye movement effects

We replicated several classical eye movement effects with distractors that signal reward. First, we demonstrated that the *remote distractor effect* (Walker *et al.*, 1997) is sensitive to reward value; a high reward-signaling distractor presented at the opposite side of the target (at 180 degrees) had a larger effect on saccade latency than a low reward-signaling distractor. Second, we showed that *the global effect* was modulated by reward value. For distractors presented close to the target (at 30 degrees) we saw that the saccade amplitude (Figure 5) and saccade landing position (Figure 8) were strongly modulated by reward value with shorter saccade latencies, landing closer to the distractor for high versus low reward value. Reward value did not modulate saccade latencies for 30 degree distractors (see Figure 4). Third, reward value had a large effect on *errors in oculomotor control*; more saccades were made to the distractors when they signaled high reward value than when they signaled low or no reward value (Figure 6). We also show a larger increase in errors across blocks for high reward-signaling distractors (Figure 10), demonstrating a learning effect as participants gradually made the distractor color-reward association, in spite of its detrimental

effect on total reward payment. Previous studies in which participants were explicitly informed of the distractor color-reward relationship show a relatively stable difference between erroneous saccades to high compared to low reward-signaling distractors across the whole experiment (Failing *et al.*, 2015; Le Pelley *et al.*, 2015). The current study demonstrates that people can learn this stable association for themselves. Since the distractor color was never mentioned and looking to the distractor was counterproductive to succeeding in the task, this approach eliminated a top-down, instruction-based explanation for these erroneous saccades.

Saccade amplitude and landing position

Distractors signaling reward affected saccade amplitude and landing position (c.f., the global effect). The undershoot in amplitude to the target center observed when a distractor was present was amplified by a high reward distractor at each location. Few studies have looked into the effects of reward on saccadic amplitude. Those studies that have assessed amplitude investigated it in relation to rewarded or non-rewarded eye movement directions (Shimo and Hikosaka, 2001; Takikawa *et al.*, 2002), reporting a decrease in the variation of amplitude for the rewarded direction but with no significant changes in mean amplitude between these conditions. The current study shows that there are reward-related changes in the amplitudes of target-directed saccades in the presence of a reward-signaling distractor. A similar effect was also seen in the landing position of saccades to the target; distractors associated with high reward led to saccade endpoints closer to the distractor than low or no reward distractors. This was the case for both fast and (relatively) slower saccades, with fast saccades in the presence of high reward resulting in a landing position closer to the distractor than in any other condition. A previous study investigating the effect of reward value on saccades to two target circles placed in close proximity to each other revealed a shift in the center of gravity towards the more highly valued stimulus (Bucker *et al.*, 2015). We found that the high reward value of a distractor also led to a shift in the landing position towards the distractor even though in the current study (unlike Bucker *et al.*, 2015) a saccade to the distractor resulted in losing the reward for that trial.

Brain mechanisms

As mentioned in the introduction, neurons in LIP represent salience and object-value information, encoding the location of attention-worthy objects in the visual scene (Gottlieb *et al.*, 2014). LIP output can be fed directly to the SC, but also to the CD in the BG. The CD acts as a gateway for gaze control, projecting to the SC via the SNr (inhibitory connections). Decreases in SNr spiking activity

leads to the "release" of the SC, essentially disinhibiting it (Hikosaka and Wurtz, 1983). It has been suggested that the head and body of the CD encode reward values flexibly, relying on short-term memory (Kawagoe, Takikawa and Hikosaka, 1998; Sato and Hikosaka, 2002), whereas the CD tail (CDt) is responsible for long-term value memory, when objects are associated with low or high reward consistently and stably (Yamamoto, Kim and Hikosaka, 2013). It has recently been shown that increased CDt neuronal firing resulting from stably high rewarded objects leads to the inhibition of SNr neurons, whereas reduced CDt firing from stably low rewarded objects (in a relative setting) leads to excitation of the SNr, resulting in SC disinhibition or inhibition, respectively (Yasuda, Yamamoto and Hikosaka, 2012). Monkeys in this study gradually acquired a preference for making saccades to high-valued over low-valued objects quickly and automatically, even when a reward was not immediately delivered anymore. We believe distractor reward-signaling encoding in the current study operates via a similar mechanism, whereby the CDt-SNr-SC circuit is trained inadvertently to acquire a preference for the high-value signaling color, in spite of its effect on reward payment.

Competitive integration

Many oculomotor effects reported here are consistent with the competitive integration account of saccadic programming, as described by Godijn and Theeuwes (Godijn and Theeuwes, 2002) (see also (Trappenberg *et al.*, 2001; Mulckhuyse, Van Der Stigchel and Theeuwes, 2009; Meeter, Van Der Stigchel and Theeuwes, 2010). This model assumes that exogenous (e.g. sensory information) and endogenous (voluntary) inputs are integrated in the SC on a common saccade map. When a target and distractor are close together in the visual field, the combined activation on the SC saccade map converges to a local maximum somewhere between these two neighboring locations. However, if the target and distractor are far apart, there is mutual inhibition until attention is directed towards the target location, leading to lateral inhibition of the remote distractor below baseline activation. The competitive integration account establishes a time-course for the activation of the SC saccade map (Godijn and Theeuwes, 2002), suggesting that in the earliest phase there is activation at the salient distractor location, with corresponding deactivation at all other locations including the target. This is followed by a rise in activation of the target location, so that both target and distractor locations are similarly active. In later stages, the target location becomes most active, passing a threshold for saccade execution to occur (see Hanes and Schall, 1996; Heitz and Schall, 2012 for description of such a threshold in the FEF), while the distractor (and fixation) locations fall below baseline for successful inhibition. The mechanisms by which endogenous signals are

integrated with sensory information have been investigated in a recent study of the monkey SC (Dorris, Olivier and Munoz, 2007), in a task in which the target location is fully predictable (an endogenous signal). A large early peak in neural activity was found in the SC for erroneous saccades to the distractor, while correct saccades to the target were represented by an early sub-threshold peak and a larger supra-threshold peak later in time corresponding to the eye movement. Our finding demonstrating faster saccade latencies to the distractor than to the target is in line with this time-course description. This behaviour is also similar to results from anti-saccade tasks, in which participants refrain from making an eye movement to a target and instead make a saccade in the opposite direction (Munoz and Everling, 2004), thereby relying on top-down inhibition of an automatic saccade to the target. Furthermore, the finding that fast saccades land closer to the distractor than slower saccades (Figure 8) is consistent with the competitive integration model. Crucially, this effect is stronger for high value relative to low value distractors, indicating that the high value distractor reaches threshold faster.

Visual hemifield effects

The effects of reward were strongly modulated by visual hemifield; relative to ipsilateral, distractors presented contralateral to the target caused more oculomotor capture (Figure 12) and had a larger effect on saccade latencies, with slower saccades to the target for contralateral versus ipsilateral distractors (Figure 11). This suggests that the competition evoked by the distractor was larger when presented within the hemifield that did not overlap with the hemifield of the target. This result is consistent with the notion that there is more suppression of the distractor activation when a target and distractor are presented within the same hemifield (with more overlapping neural representation) than between hemifields. Indeed, when presented in the contralateral hemifield the distractor activation may rise faster to threshold than within hemifield since the mutual inhibition between hemifields is assumed to be much less than within hemifield (Hickey and Theeuwes, 2011). This can be explained by the topography of a number of oculomotor-related brain regions, such as the SC and SNr. These structures contain maps with spatial representation mainly for the contralateral visual hemifield (Hikosaka *et al.*, 2000). Since objects in separate hemifields are encoded in separate populations of neurons, there is thus less mutual inhibition across hemifields than within hemifield. We also found that visual hemifield played a role in the remote distractor effect; in addition to demonstrating increased saccade latencies when a distractor was placed at the mirror opposite location to a target (Walker *et al.*, 1997), we showed that the remote distractor effect exists for increasing distance from the target (increased latencies for 180

degrees compared to 120 degrees), and also for distractors in the contralateral hemifield to the target compared to an equidistant distractor in the same hemifield. Visual hemifield effects on target processing have been investigated in several attentional tracking and visual search tasks (Luck *et al.*, 1989; Alvarez and Cavanagh, 2005). These studies concluded that hemifield independence is a signature of multifocal spatial selection, sometimes referred to as two separate "spotlights of attention", one for each hemifield. The reduced capacity to track and process items within the same hemifield has been explained in terms of competition between neighboring locations. In line with competitive integration, in which there is greater competition between items presented in the same hemifield (e.g., Stoermer, Alvarez and Cavanagh, 2014), we also show greater curvature for distractors presented at a remote location ipsilateral to the target compared to those presented equidistant in the contralateral hemifield. Overall, we demonstrate that a remote distractor in the contralateral hemifield has a large effect on saccade *planning* (slower latency, erroneous saccade endpoint) and that a remote distractor in the ipsilateral hemifield has a large effect on saccade *execution* (greater curvature). This has been explained in terms of the distinct or overlapping neural representation of space for each hemifield in oculomotor-related brain areas.

Hierarchical drift diffusion modeling

Bayesian HDDM was used to analyze inter- and intra-individual differences and similarities in saccade latencies and errors. The hierarchical framework of HDDM allowed us to overcome some limitations associated with frequentist analysis, such as the assumption that all subjects are the same and are copies of an "average subject" (Wiecki, Sofer and Frank, 2013). The HDDM instead uses both individual and group parameters to inform and constrain each other. This model revealed a much more robust effect of reward on saccadic latencies, while also taking erroneous behaviour into account in the estimation. Figure 14 shows that at each distractor angle the high reward distractor elicited a posterior probability distribution centered at the lowest drift rate (slowest latency) compared to low and no reward distractors. Like the frequentist results, this is most apparent at the largest distractor angle of 180 degrees, however HDDM also provides evidence for reward differences at smaller angles, particularly for high reward compared to low and no reward. This model further confirmed the large-effect results of distractor location found in the frequentist analysis, namely a substantially higher drift rate (faster latencies) at smaller angles. The HDDM hemifield model presented in the current study emphasizes differences in eye movement latencies when reward-signaling distractors appeared ipsilateral or contralateral to the target. The model clearly shows a slower drift rate (slower saccade latency) for contralateral distractors.

Furthermore, the effect of reward value can be seen in both ipsilateral and contralateral domains, with high reward distractors substantially slowing down the initiation of an eye movement. Results from this model suggest greatest capture of attention by remote, high reward-signaling distractors appearing contralateral to a target. Since the HDDM drift rate is not analogous to only saccade latencies, but also takes the distribution of errors into account, it may be seen as a complimentary technique to standard frequentist analyses. The incorporation of saccade latencies and errors, within and across individuals, into a single framework makes HDDM a powerful tool in analyzing decision-making behaviour, even for a fast, continuous eye movement task as presented in the current study.

Acknowledgements

We thank Eduard Ort for inspiration and help with the experimental paradigm.

Chapter 3

Overt and covert attention to location-based reward

Adapted from

McCoy, B. & Theeuwes, J. (2018)

Overt and covert attention to location-based reward

Vision Research, 142: 27-39

ABSTRACT

Recent research on the impact of location-based reward on attentional orienting has indicated that reward factors play an influential role in spatial priority maps. The current study investigated whether and how reward associations based on spatial location translate from overt eye movements to covert attention. If reward associations can be tied to different locations in space, and if overt and covert attention rely on similar overlapping neuronal populations, then both overt and covert attentional measures should display similar spatial-based reward learning. Our results demonstrate that location- and reward-based improvements in one attentional domain do not lead to similar improvements in the other. Specifically, although we found similar improvement at differentially rewarded locations during overt attentional learning, this translated to the greatest slowing at the high reward location during covert attention. We interpret this as the result of an increased motivational link between the high reward location and the trained eye movement response acquired during learning, leading to a relative slowing in manual response during covert attention when the eyes remained fixated and the saccade response was suppressed. In a second experiment, participants were not required to keep fixated during the covert attention task and we no longer observed relative slowing at the high reward location. Furthermore, the second experiment revealed no covert spatial priority of rewarded locations. We conclude that the transfer of location-based reward associations is intimately linked with the reward-modulated motor response employed during learning and can interfere with spatial priority when presented with an alternative attentional context.

1. INTRODUCTION

Current understanding of visual selective attention indicates that both covert and overt attention (attending to a location in our periphery or making an eye movement to that location) are linked by a common neural architecture, by means of a shared frontoparietal network (Corbetta, 1998; Beauchamp *et al.*, 2001; de Haan, Morgan and Rorden, 2008). The extent of this overlap has, however, been controversial. The premotor theory of attention proposed by Rizzolatti (Rizzolatti *et al.*, 1987) postulated that covert attention is akin to the programming of an eye movement. Evidence in support of this theory comes from studies reporting higher detection accuracy at a target location coinciding with the endpoint of a saccade (Deubel and Schneider, 1996; Doré-Mazars, Pouget and Beauvillain, 2004), better target detection at a saccade goal even when explicitly directed to attend somewhere else (Hoffman & Subramaniam, 1995), and a gradual build-up of attention at a saccade goal, reaching a peak immediately prior to saccade onset (Doré-Mazars, Pouget and Beauvillain, 2004; Deubel, 2008). Investigation of the frontal component of the oculomotor network, the frontal eye fields (FEF), has revealed disruption of saccades and shifts in spatial attention after applying transcranial magnetic stimulation (TMS) or microstimulation to FEF neurons (Beckers *et al.*, 1992; Moore and Fallah, 2001). Furthermore, microsaccades or 'fixational eye movements' (Martinez-Conde, Macknik and Hubel, 2004; Martinez-Conde, Otero-Millan and Macknik, 2013) have been observed during covert visual search and possibly reflect covert attentional shifts (Engbert and Kliegl, 2003; Otero-Millan *et al.*, 2008), supporting the idea of a common oculomotor neural underpinning for overt saccades and fixational eye movements in covert attentional settings.

However, there have also been numerous studies which report findings not in line with the predictions of premotor theory, including the ability to endogenously attend to stimulus locations other than the saccade goal without disturbing the eye movement (Kowler *et al.*, 1995), and a lack of facilitation of visual perception for probes presented at a saccade goal (Hunt and Kingstone, 2003). In a recent review, Smith and Schenk (Smith and Schenk, 2012) suggest that the main consistent finding from studies on premotor theory is that only exogenous (stimulus-driven) attention is dependent on saccade preparation (Henik, Rafal and Rhodes, 1994; Smith and Ratcliff, 2004; Sereno *et al.*, 2006). In their study examining the time-course of exogenous and endogenous effects, Belopolsky and Theeuwes (2012) found that although saccade preparation accompanied shifts in covert attention due to both exogenous and endogenous (goal-driven) cues, for endogenous cues the saccade program to the attended location was suppressed shortly after a covert attentional shift had been completed. Smith and Schenk (2012) propose that an alternative

to premotor theory, the biased competition account of visual attention (Desimone and Duncan, 1995; Desimone, 1998), may provide a more appropriate framework to incorporate the empirical findings garnered from assessing the validity of premotor theory. In the biased competition account, competition between neural representations is integrated across sensory and motor systems, converging on a single 'winning' representation. Physically salient items in the environment usually have a strong representation, but competition may also be biased towards less physically salient stimuli by endogenous factors such as current goals in working memory (Soto *et al.*, 2008). The lateral intraparietal area (LIP) in the parietal lobe and the FEF have been implicated in target selection in both covert attention and saccades, comprising selective spatial receptive fields (Thompson and Bichot, 2005; Bisley and Goldberg, 2010). LIP in particular has been proposed to act as an integrated priority map of top-down and bottom-up signals for behaviourally relevant stimuli (Bisley and Goldberg, 2010).

Attentional orienting has generally been described in terms of exogenous and endogenous control (Desimone and Duncan, 1995; Corbetta and Shulman, 2002; Theeuwes, 2010; Chelazzi *et al.*, 2013). In recent years, it has been suggested that these two forms of attentional orienting do not fully account for the behaviour and biases observed in the attentional orienting literature (Awh, Belopolsky and Theeuwes, 2012). Previous research has shown that equally salient stimuli associated with reward can capture attention and the eyes, even when this contradicts current selection goals (Hickey, Chelazzi and Theeuwes, 2010; Anderson, Laurent and Yantis, 2011b, 2011a; Failing and Theeuwes, 2014; Le Pelley *et al.*, 2015; McCoy and Theeuwes, 2016). In 2012, Awh and colleagues developed a framework incorporating past selection history with existing models, leading to an integrated priority map for attentional control (Awh, Belopolsky and Theeuwes, 2012).

In line with the selection history component described in the model of Awh et al. (2012), behavioural research has shown attentional orienting to the location of a non-salient cue that had acquired value through reward learning (Failing and Theeuwes, 2014). Similarly, eye movements have been observed to land closer to high compared to low reward-signaling distractors (Bucker *et al.*, 2015; McCoy and Theeuwes, 2016). It has recently been suggested that reward learning of particular locations relies upon spatial priority maps, specifically when multiple potential targets compete for attention (Chelazzi *et al.*, 2014). In the study of Chelazzi and colleagues, locations in space were first trained with reward associations, i.e. responding to a target at a particular location consistently led to a greater chance of high reward. In a subsequent test phase, letter or digit targets appeared at one, two or none of the possible spatial locations, with distractor non-

alphanumeric characters at the remaining locations, and participants had to detect and identify the target alphanumeric stimuli. A competitive advantage was found for targets presented in spatial locations previously associated with high compared to low reward. Participants could also correctly report two targets more often when the targets appeared in opposite visual hemifields, supporting previous findings that the two hemispheres can process information in parallel (Luck *et al.*, 1989; Sereno and Kosslyn, 1991; Alvarez and Cavanagh, 2005). In a previous study of ours we have also shown that visual hemifield plays a role in visual attention, indexed by both saccade latency and the number of erroneous eye movements made to a distractor (McCoy and Theeuwes, 2016). In that experiment, a target and distractor were presented simultaneously and participants had to make a quick eye movement to the target. When the target and distractor were placed in separate visual hemifields along the vertical axis, saccade latencies to the target were slower and more erroneous eye movements were made to the distractor compared to when the target and distractor were presented in the same hemifield. These findings support the biased competition model of selective attention (Desimone and Duncan, 1995), and suggest that there is less competition between than within hemifields.

Other studies also examining the effects of spatial-based reward on covert attention have reported a decrease in manual response time to targets presented at a previously rewarded location (Hickey, Chelazzi and Theeuwes, 2014; Stankevich and Geng, 2014). Specifically, Hickey et al. found that attention was primed to return to a target location associated with high reward in the preceding trial, and was biased away from the location that was previously occupied by a distractor. Stankevich and Geng (2014) showed that spatial probabilities and reward associations have independent and additive effects on manual responses in a target discrimination task, reporting a decrease in reaction time for a highly rewarded target at a highly probable location. The aforementioned studies all assessed location-based reward effects on covert attention with manual responses. However, very little research has been carried out in humans on how location-specific reward affects eye movements to those locations, or on how location-based reward mechanisms might transfer from overt to covert attention. In one relevant study different groups of participants had to perform a task that required overt responses in one group or covert responses in the other group (Camara, Manohar and Husain, 2013). The results showed similar reward effects across the two groups, namely that when a particular location was associated with high compared to low monetary reward (training phase), participants later freely choose the location that had previously been more often associated with the high reward. They also found that distractors captured gaze (overt group) or increased errors (covert group) more when they appeared at the high compared to

low reward location in the subsequent test phase. Although this study shows similar influences across the two attentional domains, individual participants always used the same response mode in both phases, with no within-participant transferal across attentional settings. Thus, it does not provide insight into how an individual's learning of location-based reward translates from overt to covert attention.

The present study was therefore designed to investigate whether and how reward associations based on spatial location translate from overt eye movements to covert attention within individual participants. We hypothesized that if reward associations can indeed be tied to different locations in space, and if overt and covert attentional orienting depend on overlapping neuronal populations, representing the integration of exogenous, endogenous and reward-related factors, then both overt and covert attentional measures should display similar spatial-based reward learning. Specifically, we expected this to be evident by reduced saccade latency and manual reaction time (RT) to stimuli presented at locations associated with higher reward value. Due to previous research on the effect of visual hemifield, we designed the experiment to maximally separate attentional allocation towards the high and low reward locations by placing them in opposite visual hemifields. In this way, attention to locations should be independent, and any effects can be assumed to be driven by the reward value at that location. We tested our hypotheses using two different task contexts: one learning phase in which saccades were made towards a salient stimulus presented at locations associated with high, low, or no reward, and one pre-training baseline and post-training test phase in which participants fixated at the center and carried out a covert visual discrimination task with stimuli presented at these same locations. The only consistent parameter across tasks in this experiment was the relative spatial positions of the stimuli, i.e., all stimulus features were different across tasks. In this way, we wished to determine the entirely spatial nature of reward learning across the two types of attentional orienting in different task contexts.

2. EXPERIMENT 1

2.1. MATERIALS AND METHODS

2.1.1. Participants

Twenty-four participants (9 female; mean 23.6 ± 3.1 years old) with normal or corrected-to-normal vision gave written informed consent to take part in the study. The experiment was approved by the Scientific and Ethical Review Committee of the VU University Amsterdam and carried out in accordance with the Code of Ethics of the World Medical Association (Declaration of Helsinki). Participants were given fixed monetary compensation of €9 for undergoing eye tracking and received a reward bonus of up to €9.90 based on their performance during the training phase.

2.1.2. Stimuli and apparatus

The experiment and stimuli were created using OpenSesame (version 2.9.7) software for Windows (Mathôt, Schreij and Theeuwes, 2012). We set the software to use a legacy back-end built on top of PyGame. Stimuli were presented on a 22-inch Samsung SyncMaster 2233RZ monitor (1680 x 1050 pixels resolution, 120 Hz refresh rate). Participants were tested in a sound-attenuated, dimly-lit room. The viewing distance was held constant at 70 cm using a chin rest. Eye movements were recorded using an Eyelink 1000 tracker (SR Research Ltd., Canada) with 1000 Hz temporal resolution and a 0.2° of visual angle spatial resolution. This tracker consists of an infrared video-based tracking system to compute the pupil center and pupil size of both eyes. Data were collected and analysed for the right eye only.

2.1.3. Training phase

Training of eye movements (overt attention) constituted the main part of the experimental session, lasting approximately 35 minutes including breaks. In this task participants had to make a quick eye movement to a target location (see Fig. 1). Each trial began with participants focusing on the central fixation-cross (0.09° x 0.09°) presented on a grey background (rgb = (100,100,100), 5 cd/m^2), pressing the spacebar to start the trial. Immediately when the trial was initiated, four dark grey circles (rgb = (54,54,54), 19 cd/m^2) appeared at four locations equally spaced on an imaginary circle with a radius of 6.6°. At this point, participants continued to remain fixated at the center. After a short, jittered interval (200 – 800ms), three out of the four circles turned red (rgb = (255,0,0), 34 cd/m^2), with only one circle remaining the original grey color. The fixation cross disappeared at the instant these circles turned red. Participants were instructed to make a quick

eye movement to the remaining grey circle, within 500ms of the change in stimulus color. They heard an incorrect low beep tone (20ms, 250 Hz sine wave) if they did any of the following: moved their eyes from central fixation before the circles turned red, looked first to a red circle before looking to the target grey circle, or failed to make a saccade to the grey circle within the 500ms time limit. In this way, participants were trained to make a quick and direct eye movement to the target location. Participants first completed a baseline practice block, with no feedback other than when the trial was incorrect. They subsequently completed three blocks in which rewarding feedback was also presented at specific locations. Of the four target locations, two locations had a special rewarding status and two were used as control locations where no reward was ever given. Of the two reward locations, one was associated with a high reward 80% of the time, and the remaining 20% of trials corresponded to low reward. The other reward location adhered to the opposite contingency; low reward feedback was presented on 80% of the trials, with high reward on 20%. Feedback was presented as yellow text '10' or '1' (0.5°/0.25° x 0.5°, rgb = (216,216,0), 82 cd/m^2) at the center of the circle to which an eye movement was made. This corresponded to a monetary payment of 10 cents or 1 cent per high or low reward trial respectively. There was no feedback given at the two remaining control (no reward) locations. Participants were informed that they would lose 1 cent for an incorrect response. This deduction was implemented to encourage participants to always look to the target location for each trial, even if no reward was to be expected at that location. The target grey circle was presented at each location 30 times per block, leading to 120 trials in total per block. Participants completed one baseline block in which no reward was given, followed by three reward blocks. They received general feedback at the end of each block, indicating their accuracy and average saccadic latency. They were also presented with their accumulated earnings ¾ of the way through each block. In order to ensure subjects were staying attentive to the reward feedback, a question appeared pseudo-randomly after some trials asking how much they had received that trial "10c, 1c, or nothing?", with instructions on which arrow key to press for each answer. This question was posed only after correct trials, up to eight times per block. In each block the question appeared no more than twice for each possible target location, so that attention was not inadvertently drawn to any one particular location. Finally, due to results from previous research on competitive effects across visual hemifield, the high and low reward locations were always presented in opposite visual hemifields to each other, and were counterbalanced across participants in all possible combinations of opposite hemifield locations (see Fig. 2).

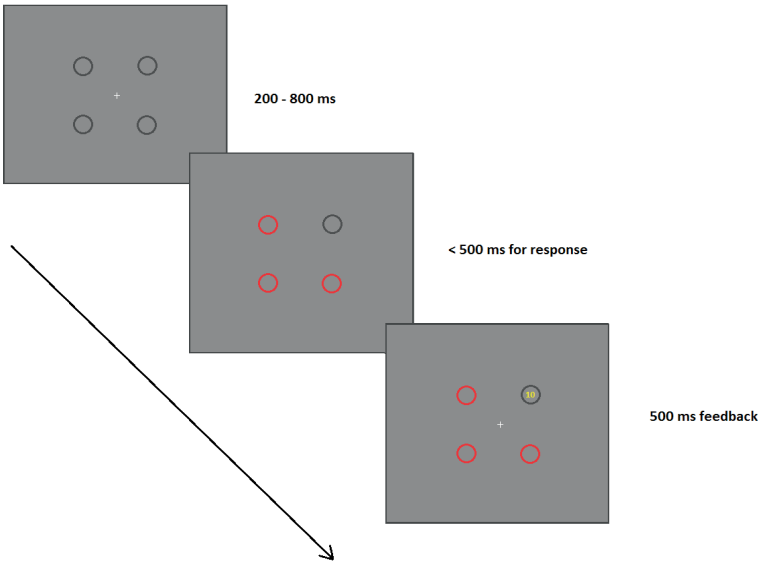


Figure 1. Training phase trial sequence.

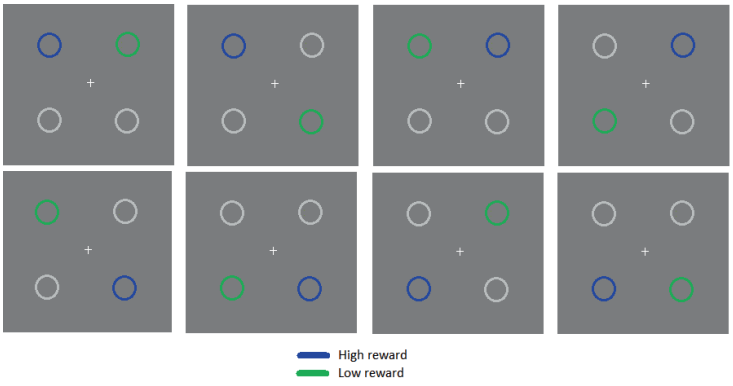


Figure 2. Reward-location contingences. Each picture represents a reward-location contingency for a given participant. These eight contingency options were repeated three times across the 24 participants. The light grey circles represent the control locations where no reward was given.

2.1.4. Test phase

The test phase was a covert attention task completed twice by each participant, once before and once after the training phase (see Fig. 3). Each test session lasted approximately 10 minutes. The first test acted as a baseline, to obtain participants' button-press response times (RTs) to targets at specific spatial locations before any reward manipulation had occurred. Trials were presented in a random order in both sessions. The same instructions were provided on screen before each session, explicitly informing participants that no reward would be earned during this part. Participants were instructed to maintain fixation on the central fixation-cross during all trials, and to make no eye movements to the stimuli. The trial began with a short random interval (200 – 800 ms) showing only the fixation cross (0.09° x 0.09° visual angle) presented on a grey background (same as training phase: rgb = (100,100,100), 5 cd/m^2). A pre-mask display with four yellow figure-eight templates (0.67° x 0.91°, rgb = (216,216,0), 82 cd/m^{2}was then presented, with the stimuli equally spaced at four spatial locations on an imaginary circle with a radius of 6.6°, i.e., the stimuli were presented at four fixed locations along the principal diagonal axes, at exactly the same locations as those used in the training phase. After another short, jittered interval (200 – 800 ms), lines from these stimuli disappeared, leaving the letters 'C', 'F', 'H', and 'P'. These letters were visible on the screen for 200 ms. The target letter was 'C', to which participants had to make a button press response: pressing the left arrow key when the gap was on the left of the letter (a mirrored 'C'), or the right arrow key when the gap was on the right of the letter (a standard 'C'). Participants had 1000ms from the appearance of the target to make their response, while remaining fixated. They were instructed to give a response as quickly and accurately as possible. Participants heard a low beep tone (20ms, 250 Hz sine wave) whenever they gave an incorrect response or failed to respond in time. Trials were self-paced, so participants could begin the next trial whenever they wanted by fixating and pressing the spacebar. The target was presented 30 times at each of the four locations, with 15 of these presentations requiring either a left or right arrow response, leading to 120 trials in total in each of the baseline and test sessions.

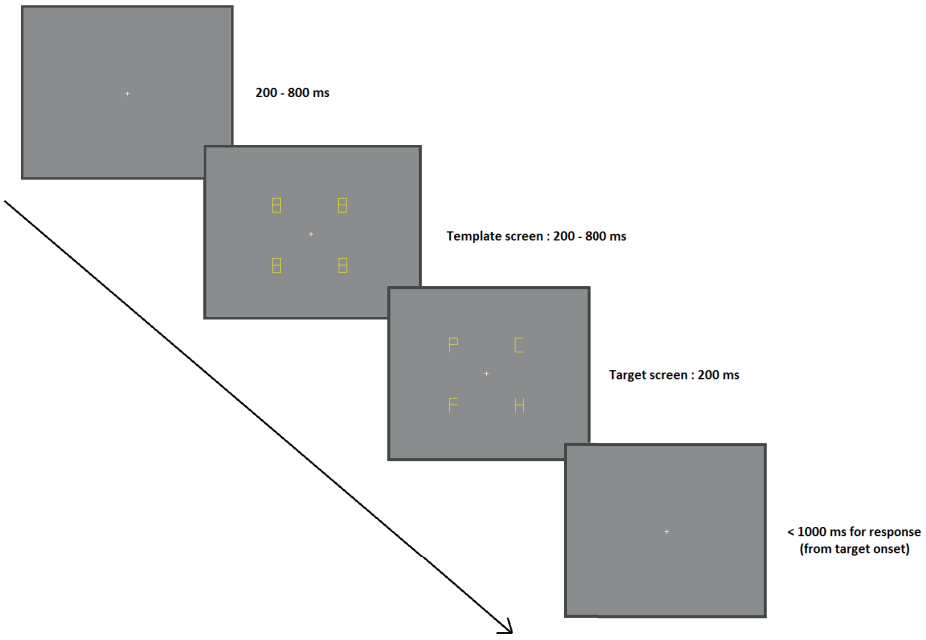


Figure 3. Baseline and test phase trial sequence in Experiment 1.

2.1.5. Data Analysis

Data were analyzed using custom-made Python (version 2.7) scripts, Data Viewer (version 2.3.1; SR Research) and SPSS software (version 23.0; Chicago, IL). For the training phase, saccade latencies of the first eye movement of each trial were analyzed. An eye movement was considered a saccade when eye velocity exceeded 35°/s or eye acceleration exceeded 9500°/s^2, and the saccade was deemed complete whenever the velocity/acceleration fell below that threshold. Saccade latency was described as the time interval between the display stimuli turning red (leaving one target grey circle) and the initiation of the saccade response. Data were analyzed for correct first saccades, containing no blinks, which started from within 3° of fixation and landed within 3° of the target location. Saccade latencies less than 80ms (anticipation errors) and greater than 500ms (timeout) were excluded from further analysis. Latencies greater than 3 standard deviations from the mean were removed from the analysis.

Manual response times (RTs) were collected during the test phases. The RTs were calculated as the difference between the time the pre-mask turned into letter stimuli and the time of the button press. Data were analyzed for correct trials with no blinks and no saccades made outside of the fixation region of interest (3° visual angle), with RTs greater than 200ms (responses < 200ms were considered anticipation errors) and RTs less than 1000ms (timeout). RTs greater than 3 standard deviations from the mean were removed from the analysis. Validity of the main findings were corroborated in a Bayesian framework using Bayes factors with the *BayesFactor* package (version 0.9.12-2) in R (R Development Core and Team, 2015). To assess whether there was any effect of duration of the pre-mask (stimuli resembling the figure eight) on manual RTs in the test phase we carried out a generalized linear mixed effects regression in R using the *lme4* package (Bates *et al.*, 2014). The analyzed data included all correct trials for each participant in which the target was presented at the high or low reward location. RT, the dependent measure, was modeled as an inverse Gaussian distribution, with reward location and phase (baseline or test) included as fixed effects, pre-mask duration as a mean-centered covariate and subject as a random intercept. Further models were established to include interaction terms, to assess changes across phase and whether pre-mask duration differentially affected RTs to rewarded locations. An ANOVA was carried out on the original main effects model and each new model including interaction terms, for model fit comparison. Akaike information criterion (AIC) and Bayesian information criterion (BIC) for each model were also compared to see which one best explained the data (more negative AIC and BIC represent a better model).

2.2. RESULTS

2.2.1. Exclusions in overt task

For the 24 participants who took part in the training session, an average of 0.86% (SD 1.07%) of participant's trials were discarded due to saccade onset latencies of less than 80ms, and 3.97% (SD 4.12%) were discarded because of saccade onset latencies greater than the timeout of 500ms (participants were given feedback indicating the error for these trials). 0.32% (SD 0.67%) of trials were removed due to blinking either immediately before or during the first saccade. Of the remaining trials, 80.07% (SD 12.28%) of first saccades landed within 3 degrees of the target, and 19.30% landed outside of this region. Only those first saccades landing within the target region were used for further analysis.

2.2.2. Reward learning in overt task

Mean saccadic onset latencies were calculated per participant per block (one baseline block and three reward blocks). Latencies for the reward blocks were obtained according to whether the target was at the high reward, low reward or no reward (control) spatial locations for that participant (see Fig. 2 for participant-specific spatial-reward contingencies). Saccade latencies across reward blocks and locations are presented in Fig. 4. Saccade latency improvements at each location were calculated by subtracting the mean latency at that location in each reward block from the mean latency of all trials in the first baseline block in which no reward was given. Data were pooled across the two control locations at which reward was never given. A repeated-measures ANOVA was carried out on these improvements across blocks, with reward location (high, low, or no reward) and reward block (1-3) as factors. This revealed a main effect of reward location ($F(2,46)=3.78$, $p = .03$, $\eta^2 = .141$) and a reward location*block interaction ($F(4,92) = 3.17$, $p = .017$, $\eta^2 = .121$). Probing this interaction, pairwise comparisons (Bonferroni corrected for multiple comparisons) were made between each pair of high, low and no reward locations in each block. This revealed a significant difference between the high and no reward locations in the final block ($t(23)=3.32$, $p = .009$), with a 15 ms reduction in saccade latency to the high relative to no reward locations (22 ms vs 6 ms for high versus no reward locations, respectively). There was also a reliable difference in improvement between the low and no reward locations in the final reward block ($t(23)=3.31$, $p = .006$), with a 18 ms reduction in saccade latency to the low relative to no reward location in the final block ($t(23)=3.46$, $p = .006$; 24 ms vs 6 ms for low versus no reward locations respectively). No differences were found between the high and low reward locations in any block (all $p > .1$). Participants were presented with the question 'how much did you receive in the last trial' on average 18 (SD 3) times across the three reward blocks, to ensure they paid attention to the feedback given. This was not a constant number across participants as it depended on participants responding correctly to trials, while ensuring as far as possible that the question was presented a similar number of times for the four locations and dispersed evenly across each block. Participants responded correctly to 91% (SD 12%) of these questions.

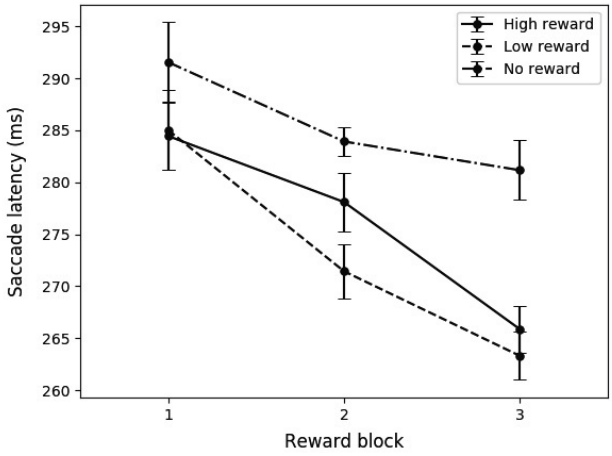


Figure 4. Experiment 1 training phase: saccade latency per block for each of the rewarded locations. Error bars are ± SE normalized for within-subject design (Cousineau, 2005; Loftus & Masson, 1994).

2.2.3. Exclusions in covert task

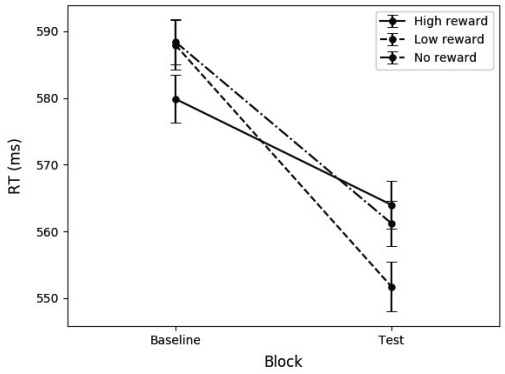
The test phase was designed to be relatively simple to promote good performance so that the RTs of a sufficient number of correct trials could be analyzed across locations. Indeed, accuracies for the pre-training baseline and post-training test phase reflected this, with correct response to 93.16% (SD 9.52%) and 96.84% (SD 3.14%) of trials in each phase respectively. There were no manual RTs occurring quicker than 200ms in either the baseline or test phase. 5.07% (SD 8.66%) and 1.25% (SD 2.87%) of all trials from the baseline and test phases respectively were removed due to RTs greater than the timeout of 1000ms. 8.33% (SD 9.02%) and 3.44% (SD 4.54%) of trials were removed due to blinks during the first fixation of a trial. 1.84% (SD 2.14%) and 1.39% (SD 2.49%) of trials were discarded due to any saccades made outside of the fixation zone (> 3° of visual angle). Exclusions are reported as percentage of all trials, however there may have been overlap in the reason for removal of a given trial, e.g. one trial may have been both incorrect and contained blinks. After these exclusions, trials with RTs greater than 3 SD were removed, leaving 88.19% (SD 10.29%) and 93.26% (SD 5.00%) of all trials from each phase available for analysis.

2.2.4 Improvements in covert task

Reaction times in the covert task at each location before and after training are presented in Fig. 5a. RT improvement at each location can be seen in Fig. 5b. As in the training phase, RTs for the no reward locations were pooled in each of the baseline and test phases, and baseline to test phase improvement was calculated from these averages. A repeated-measures ANOVA was carried out on the change in RT from the baseline to test phase with reward location (high, low, or no reward) as a factor. We found a significant main effect of reward location ($F(2,46) = 5.439$, $p = .008$, $\eta^2 = .191$). Bonferroni-corrected pairwise comparisons on the gains at each of the high, low and no reward locations revealed a robust difference in RT gain between the high and low reward locations, with a relative slowing at the high reward location (15.87 ms versus 36.20 ms for improvement at the high and low reward location respectively; $t(23) = 3.03$, $p = .018$). There was a smaller and non-significant difference between the high and no reward locations, with a relative slowing at the high location (15.87 ms vs. 27.19 ms for high and no reward locations respectively; $t(23)=1.72$, $p > .1$).The difference between low and no reward locations was also not significant ($t(23)=1.76$, $p >.1$; 36.20 ms vs. 27.19ms for low and no reward locations respectively). To ensure the feasibility of collapsing RTs across the two no reward locations and that they followed a similar pattern, one of the no reward locations was taken as that location in the same left/right visual hemifield as high reward, and the other as the location in the same hemifield as low reward. These were different locations depending on the subject-specific spatial-reward contingencies (Fig. 2). Improvements in RTs to these two no reward locations were almost identical, with no differences in gain between locations (27.23 ms vs. 27.14 ms for location in the high reward hemifield vs. location in low reward hemifield respectively; $t(23) = .015$, $p = .988$).

An additional analysis using Bayes factors was carried out on these test phase data, to help confirm the validity of the main finding. A Bayesian paired-samples t-test was conducted on the difference in RT gain (from baseline to test phase) between the high and low reward locations. This was implemented using a standard Cauchy distribution with a scale of $\sqrt{2}/2$ as the prior for the alternative hypothesis. A Bayes factor of 7.52 was found in favor of the alternative relative to the null hypothesis, indicating that it was over seven times more likely that RTs to the high relative to low reward locations were different.

(a)



(b)

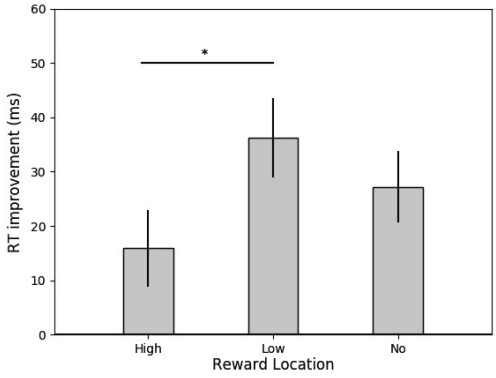


Figure 5. Experiment 1 test phase. **(a)** Manual RTs for each of the rewarded locations, before and after the training phase, and **(b)** improvements in manual RTs for each of the rewarded locations. Error bars are ± SE normalized for within-subject design.

2.2.5. Improvements for relative reward locations

Due to the counterbalanced design of this experiment it was possible to split the test phase data in half, with 12 participants in one group who received the high and low reward at locations spaced 180 polar degrees (°) apart (in opposite visual hemifields along both the vertical and horizontal meridian), and the remaining 12 participants in another group with the high and low reward locations spaced 90° apart (in opposite visual hemifields along the vertical axis only). In each

subset, the no reward locations were thus perfectly matched with respect to the reward locations, e.g. in the 180° condition, the two no reward locations were also 180° apart. We conducted a repeated-measures ANOVA on the change in manual RTs from baseline to test phase in the covert attention task with reward location (high, low, or no) as a within-subject factor, and angle between high and low reward (90° or 180°) as a between-subject factor (see Fig. 6). This revealed a reward location* group interaction ($F(2,44)=4.86$, $p = .012$, $\eta^2 = .181$). Bonferroni-corrected pairwise comparisons on the difference in RT improvements between the high and low reward location across the two groups revealed significantly more slowing at the high compared to low reward location for the group that viewed those locations 180° apart during training ($t(22)=3.50$, $p = .006$; with a high versus low difference of 3.23ms and 37.44ms for the 90° and 180° group, respectively). This high vs. low RT difference in the 180° group was very robust ($t(11)=4.56$, $p<.001$). There was also a significant high vs. no difference in the 180° group ($t(11)=2.79$, $p=.032$).

We again used Bayes factors to corroborate these findings. A Bayes factor analysis of the RT differences between the high and low reward location of the 180° group showed very strong evidence in favor of substantial slowing at the high relative to low reward location, with this alternative hypothesis 120 times more likely to explain the data compared to the null hypothesis (BF=120). Analysis of the difference between the high and no reward locations also provided evidence in support of a relative slowing at the high reward location (BF=7.99).

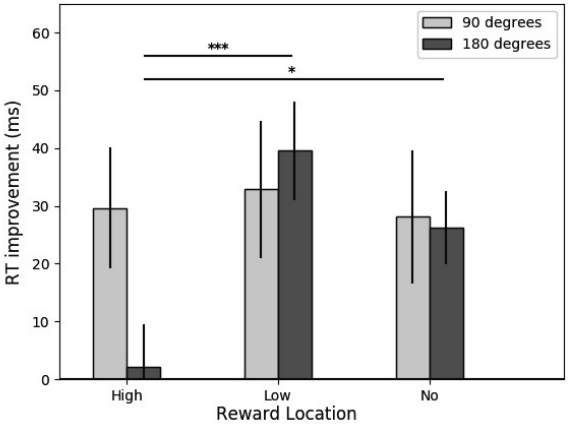


Figure 6. Experiment 1 RT improvements at each location for participants with high and low reward locations separated by 90° or 180°. Error bars are ± SE normalized for within-subject design.

2.2.6. Effect of pre-mask duration

Given the design of the test phase, it was possible that the presentation of the pre-mask display before the target screen influenced attentional allocation to the different locations prior to target onset. For example, if participants attended preferentially to the high reward location during pre-mask presentation, longer pre-mask durations may have led to inhibition of return to that location, resulting in longer manual reaction times if the target were subsequently presented there. The pre-mask screen was presented for a random jittered interval on each trial. In order to test its influence on RTs, we ran a generalized linear mixed effects regression model on trials in which the target was presented at either the high or low reward location. This was carried out on data from both the baseline and test phase, with a fixed-effect variable encoding the task phase (see *Methods* for full description of analysis). In addition to phase and reward location as fixed main effects variables, and pre-mask duration as a covariate, the inclusion of a phase*reward location interaction term resulted in a better model fit (improved model fit over main effects model: $\chi^2(1) = 10.60$, p = .001, with lower AIC and BIC). This analysis revealed a main effect of pre-mask duration, with longer durations resulting in longer RTs (t = 2.69, p = .007). It also confirmed the previous critical finding, namely a reward location*phase interaction (t=3.26, p =.001), and the strong main effect of phase on RT reduction (t=8.00, p < .001). Including a pre-mask duration*reward location interaction term did not improve model fit compared to the main effects only model ($\chi^2(1) = .14$, p = .712, with higher AIC and BIC). Overall, we have shown that the pre-mask duration influenced manual RTs during the covert attention task, but did not interact with the observed effect of reward location on RTs to targets presented at the high and low reward locations.

2.3. DISCUSSION

This experiment was designed to determine whether reward-based biases of attention to locations in space would persist across different tasks involving overt or covert attention. During an eye movement training phase, observers learned which locations were associated with obtaining reward. Saccade latencies became faster to rewarded than to non-rewarded locations across this phase. The outcome of the covert test phase was, however, the opposite to what was expected: manual RTs to targets presented at the high reward location were slower than to all other locations. If acquired biases of spatial attention are persistent and generalize across stimuli and task contexts as argued by Chelazzi et al. (2014), we should have found benefits for directing attention to the high reward location relative to other locations. Specifically, Chelazzi et al. actually found no differences

in RT to high and low reward locations during learning, but still found spatial priority for the high reward location in the test phase when there was greater competition for attention between locations. Our finding implies that reward-based training involving eye movements does not result in the shaping of a "general" spatial priority map, at least not a map that persists across tasks and motor outputs.

Since participants had to keep fixated throughout the test phase, we interpret the relative slowing at the high reward location as the result of greater suppression of the learned, reactive eye movement response. This inhibition affected the high more than the low reward location due to increased competition between those locations in the test phase, since the target was located by covert visual search, compared to the exogenous, bottom-up capture of the salient target in the training phase. An increased motivational link between the high reward location and the saccade response therefore manifested in an increased manual RT during covert attention. Such inhibition can arise from increased competition in neurons tuned to specific spatial locations in brain regions associated with fixational eye movements and the generation of saccades - the superior colliculus (SC), frontal eye fields (FEF), and lateral intraparietal sulcus (LIP). Evidence in support of this interpretation comes from research on the anti-saccade task. In this task, participants are required to look to the mirror position of a target upon its presentation, i.e. a target presented at 4 degrees directly to the left of fixation requires a saccade to a location 4 degrees to the right. Studies using this task have been carried out in an effort to dissociate stimulus encoding from saccade preparation (Hallett, 1978; Amador, Schlag-Rey and Schlag, 1998; Bell, Everling and Munoz, 2000). In a review by Munoz and Everling (Munoz and Everling, 2004), the anti-saccade task is described in terms of an incongruent stimulus-response mapping. When the stimulus and response are compatible, the appearance of the stimulus automatically activates the correct saccade response. However, when they are incongruent the automatic response has to be aborted and the correct response (moving the eyes away) must be prepared. Similar to evidence from manual response tasks on congruency effects, such an abort mechanism takes time and results in longer reaction times for incongruent responses (Hommel and Prinz, 1997).

In their review, Munoz and Everling (2004) describe the neural mechanisms of two important processes involved in the anti-saccade task; the first and relevant one for the current study is the suppression of the automatic pro-saccade to the target stimulus. This suppression arises from the interplay between neurons in the rostral and caudal SC. These two populations of neurons were initially termed ‘fixation’ and ‘saccade’ neurons respectively, as they displayed reciprocal activity during fixation and eye movements (Munoz, Pelisson and Guitton, 1991; Munoz

and Wurtz, 1993a, 1993b). More recently, a number of studies have found both rostral and caudal SC to be involved in attending to a target (Cavanaugh and Wurtz, 2004), target selection (Carello and Krauzlis, 2004), and saccade trajectory (Gandhi and Keller, 1999). Emerging evidence indicates that fixation is an equilibrium state in which activity distributed across the left and right SC determines gaze direction (Goffart, Hafed, & Krauzlis, 2012), with the SC containing a single map of behaviourally relevant goal locations (Hafed & Krauzlis, 2008). Fixational eye movements or microsaccades can therefore result from transient imbalances between fluctuating target position activity in the SC (Hafed & Krauzlis, 2008; Hafed, Goffart & Krauzlis, 2009). Although the SC drives saccade execution, and has been extensively researched in relation to saccade threshold and initiation (Sparks, Holland & Guthrie, 1976; Pettit et al. 1999; Wurtz & Goldberg 1972), it has also been acknowledged that several other brain areas play important roles in saccade processes including the LIP, FEF, supplementary eye field (SEF), dorsolateral prefrontal cortex (DLPFC) and the basal ganglia, forming recurrent connections with each other and projecting their outputs to the SC (Leigh and Zee, 1999; Schall and Thompson, 1999; Glimcher, 2003; Hikosaka, Nakamura and Nakahara, 2006). Indeed, Munoz and Everling (2004) postulated that it is likely that the signals arising in the SC and FEF, during the suppression of the automatic pro-saccade in the anti-saccade task, come from the SEF, DLPFC or the SNr. The SNr is particularly relevant in the current study, as it is modulated by the caudate nucleus (CN) of the basal ganglia, a crucial brain area for reward processing (Nakahara, Amari and Hikosaka, 2002; Hikosaka, Nakamura and Nakahara, 2006; Nakamura and Hikosaka, 2006b; Yasuda, Yamamoto and Hikosaka, 2012). Furthermore, it has been suggested that the tail of the CN encodes both what an object is and where it is located in space (Yamamoto *et al.*, 2012). Our additional analysis on learning of relative reward locations indicates that the RT slowing at the high compared to low reward location can be explained mainly by those participants who experienced the reward associations at mirror-opposite locations (180° apart) during reward learning. As such, the neuronal responses encoding saccades to those two locations are assumed to be maximally independent, and potentially demonstrate a bias towards the high reward location on the SC map of behaviourally relevant goal locations.

In summary, we propose that the larger value associated with the high reward location increased activity in the SC neurons encoding its location, thereby making successful inhibition of the learned, automatic saccades to that location more effortful and time-consuming. The manual response to the target was subsequently delayed relative to responses to other locations associated with lower value.

3. EXPERIMENT 2

In Experiment 1, participants were instructed to stay fixated throughout the test phase. The experiment was specifically designed in this way to ensure covert attention to the target stimulus was fully spatial-based relative to fixation. In Experiment 2, we wanted to gather evidence regarding our subsequent interpretation that it was this specific imposition that led to the relative slowing at the high reward location. We therefore carried out a very similar experiment, but removed the requirement to maintain fixation in the baseline and test phase. The target screen was again presented for 200 ms; short enough so that participants could not fixate the target while it was visible on the screen. Participants fixated the central cross only to begin a trial and did not hear a beep (which indicated an incorrect response in Experiment 1) if they moved their eyes. Experiment 2 was therefore designed to address two questions: 1) whether the relative slowing at the high reward location in Experiment 1 was due to suppression of reactive saccadic eye movements and 2) whether overt spatial-based reward learning leads to similar improvements when covertly attending these locations, when the eyes are free to move.

3.1. MATERIALS AND METHODS

3.1.1. Participants

Twenty-four new participants (15 female; mean 26 ± 6 years old) with normal or corrected-to-normal vision gave written informed consent to take part in the study. All procedures and compensation were the same as in Experiment 1.

3.1.2. Stimuli and apparatus

The experimental setup was almost identical to that used in Experiment 1 (Figs. 1 & 3), except for the presentation order of stimulus screens in the baseline and test phase. In this experiment, the target stimulus screen was presented first for 200 ms, *before* the "pre-mask" figure-eight template screen appeared. The pre-mask thus became a standard mask in this experiment and stayed on the screen until the end of the trial. The order of the stimulus screens was reversed to minimize eye movements before the target screen appeared; if the original order remained, the onset of the pre-mask might have encouraged eye movements towards the stimuli, meaning that covert attention at the moment of target onset might not necessarily have been considered spatial-based in relation to

fixation. The location-reward associations were fully counterbalanced across participants as shown in Fig. 2.

3.1.3. Data analysis

Data analysis was carried out in the same way as in Experiment 1. An additional analysis was carried out here on eye movement behaviour during the test phase. Participants were not given explicit instructions regarding eye movements in this experiment, aside from needing to fixate to begin a trial, so we therefore observed a range of saccade and microsaccade behaviour and strategy across participants. Based on extensive research in recent years on the role of microsaccades in visual attention (Otero-Millan *et al.*, 2008; Goffart, Hafed and Krauzlis, 2012; Martinez-Conde, Otero-Millan and Macknik, 2013), we included saccades of any size in this analysis. To keep the analysis as clean as possible we took only one saccade for a given trial according to the following criteria 1) it was the first saccade of any size made after stimulus presentation, occurring before the manual response was made, 2) the saccade had to move in the direction of the specified location (high, low, or no reward) from saccade start point to end point, with an end point in that quadrant of the screen, and lastly 3) the target stimulus during the trial was presented at the specific location being analyzed. In this way we calculated the number of first eye movements made to the high, low and no reward locations when the target stimulus was presented at that location.

3.2. RESULTS

3.2.1. Exclusions in overt task

For the 24 participants who took part in the training session, an average of 0.53% (SD 0.70%) of participant's trials were discarded due to saccade onset latencies of less than 80ms, and 3.53% (SD 2.73%) were discarded because of saccade onset latencies greater than the timeout of 500ms (participants were given feedback indicating the error for these trials). 0.79% (SD 1.60%) of trials were removed due to blinking either immediately before or during the first saccade. Of the remaining trials, 82.37% (SD 9.67%) of first saccades landed within 3 degrees of the target, and 17.63% landed outside of this region. Only the first saccade of each trial landing within the target region was used for further analysis.

3.2.2. Reward learning in overt task

As in Experiment 1, saccade latency improvements at each location were calculated by subtracting the mean latency at that location in each reward block from the mean latency of all trials in the first block in which no reward was given (absolute saccade latencies for reward blocks can be seen in Fig. 7). A repeated-measures ANOVA with reward location (high, low, or no reward) and reward block (1-3) as factors revealed a main effect of reward location ($F(2,46)=4.24$, $p = .02$, $\eta^2 = .156$). There was no interaction between reward and block ($p > .1$), demonstrating a consistent effect of reward from early in the reward blocks. Pairwise comparisons (Bonferroni corrected) were made between each pair of high, low and no reward locations. This showed a consistent difference between the high and no reward locations across blocks ($t(23)=2.80$, $p=.03$; with a high vs. no reward improvement in saccade latency of 5.95 ms vs. -3.25 ms respectively). There were no significant differences between the high and low reward locations or between the low and no reward locations (all $p>.1$). Participants were presented with the question 'how much did you receive in the last trial' on average 18 (SD 2) times across the three reward blocks, to ensure they paid attention to the feedback given. Participants responded correctly to 89% (SD 15%) of these questions.

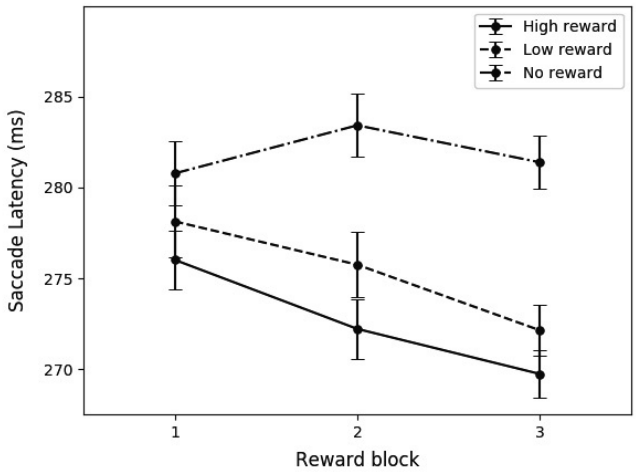


Figure 7. Experiment 2 training phase: saccade latency per block for each of the rewarded locations. Error bars are ± SE normalized for within-subject design.

3.2.3. Exclusions in covert task

Accuracies in the pre-training baseline and post-training test phases were 82.85% (SD 14.84%) and 86.04% (SD 12.36%) respectively. No trials were discarded due to manual RTs quicker than 200ms in either the baseline or test phase. 3.13% (SD 3.51%) and 1.08% (SD 1.83%) of all trials were removed due to RTs greater than the timeout of 1000ms. 5.17% (SD 9.04%) and 4.72% (SD 6.70%) of trials were discarded due to blinks during the first fixation of a trial. Finally, 17.15% (SD 14.84%) and 13.96% (SD 12.36%) of trials were removed for an incorrect button response. After these exclusions, trials with RTs greater than 3 SD were removed, leaving 78.82% (SD 14.44%) and 81.49% (SD 11.94%) of all trials from each phase available for full analysis.

3.2.4. Improvements in covert task

RT improvement from baseline to test phase at each location is shown in Fig. 8. A repeated-measures ANOVA was carried out with reward location (high, low, or no reward) as a factor. We found no main effect of reward location $F(2,46)=.05$, $p=.951$), with very similar improvements at all locations (26 ms, 28 ms and 25 ms improvements for high, low and reward locations respectively). To ensure similarity in responses to no reward, one of the no reward locations was again assigned as the location in the same hemifield as high reward, and the other as the location in the same hemifield as low reward. There were again no differences in RT gain between the two no reward locations (26 ms vs. 25 ms for no reward location in the high vs. low reward hemifield respectively, $p > .1$).

As in Experiment 1, here we also used Bayes factors to confirm the test phase results. This analysis is especially useful in this case, as Bayes factors can provide evidence not only for an alternative hypothesis, but also for the null hypothesis. A Bayesian paired-samples t-test was conducted on the difference in RT gain (from baseline to test phase) between the high and low reward locations. Comparing the alternative hypothesis (that these two variables were different) to the null hypothesis, we obtained a BF of 0.22. Thus, we found evidence in favor of the null hypothesis (BF = 4.57), with the data over four times more likely under the null hypothesis than the alternative.

Finally, since we found a large difference between manual RTs to the high and low reward location specifically in the 180° group in Experiment 1, we also analyzed the RTs separated by group in Experiment 2. A repeated-measures ANOVA was performed on RT improvement with reward location (high, low, or no) as a within-subject factor, and angle between high and low reward (90° or 180°) as a between-subject factor. In contrast to Experiment 1, this did not show

any interaction between reward and group (F(2,44)=.05, *p*=.95), indicating that relative improvements at the high, low and no locations did not differ between groups.

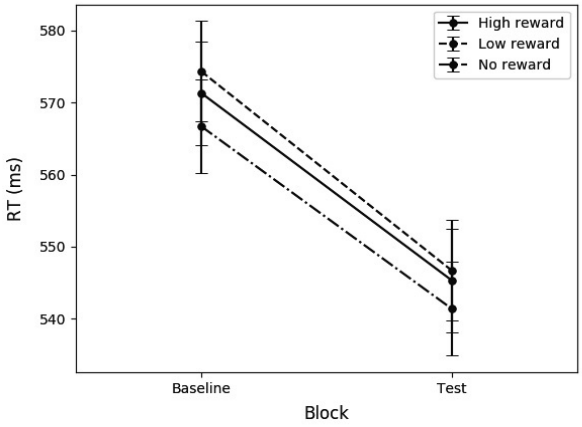


Figure 8. Experiment 2 baseline/test phase results. Manual RTs for each location, before and after the training phase. Error bars are ± SE normalized for within-subject design.

3.2.5. Saccades in covert task

Since participants were not given explicit instructions regarding eye movements during the baseline and test phase of Experiment 2, aside from the need to fixate to begin each trial, it was possible to analyze changes from the baseline to test phase in the number of eye movements made to specific locations. Subjects differed greatly in strategy, some making very few saccades across the whole phase and others executing many eye movements (an across-subject average of 0.46 saccades per trial were made in the baseline phase (SD 0.40) and 0.43 saccades per trial in the test phase (SD 0.52)). Saccades also varied in size (from micro to standard saccades), occurred several or few times per trial, and could be executed at any point in the trial. Therefore, to keep this measure as clean as possible we analyzed only one saccade for a given trial according to the criteria outlined in the *Data Analysis* section 3.1.3. 7.43% of all trials in the baseline phase (SD 8.64%) and 5.62% of all trials in the test phase (SD 6.66%) contained a first saccade to one of the locations when the target was presented there. The number of saccades towards each location in each phase was calculated as a percentage of the total number of first saccade trials in that phase. Since Experiment 2 was established to explain the outcome of the test phase in the first experiment, we

focused on the 180° group here to examine changes in eye movements that may help in explaining the observed pattern in Fig. 6. A repeated-measures ANOVA with reward location as a factor showed a marginally significant main effect of reward location ($F(2,22) = 3.40$, $p = .05$, $\eta^2 = .236$), with a 10.05%, 4.77% and -5.74% change from baseline to test phase in saccades made to the high, low and no reward locations respectively. Thus overall, there was the greatest numerical increase in saccades made towards the high reward location and a numerical decrease in saccades towards the no reward locations. A similar analysis of the 90° group did not show any significant changes in saccades between locations ($p > .1$).

3.3. DISCUSSION

The second experiment was established to answer two questions, the first of which addressed the outcome of Experiment 1; did the requirement to remain fixated and suppress saccades lead to the relative slowing seen at the high reward location? In Experiment 2, participants were free to move their eyes when the target was presented, and we observed that this no longer led to slowing at the high reward location, with participants showing similar RT gains to all locations after training. In Experiment 1, we found the relative slowing at the high reward location during the test phase was especially robust in the subset of participants who were trained with the high and low reward locations separated by 180°. In Experiment 2, a different group of participants underwent exactly the same reward location training, but here we observed no RT differences to any location, regardless of the relative separation between high and low reward during training. Although this experiment was not established to specifically examine eye movements and microsaccades during the test phase, it was still possible to analyze saccade behaviour of participants. In recent years, numerous studies on fixational eye movements have been providing evidence for the role of microsaccades in visual attention, indicating a microsaccade-saccade continuum rather than two discrete neuronal populations for fixation and saccade generation (Otero-Millan *et al.*, 2008; Hafed, Goffart and Krauzlis, 2009; Krauzlis, Goffart and Hafed, 2017). Our results demonstrate the greatest increase in (micro)saccades towards the high reward location, for the group that underwent the 180° high-low reward separation contingency. That same group in Experiment 1 demonstrated a significant and robust slowing down in RT at the high reward location compared to both the low and no reward locations. Although we did not observe a significant difference in eye movements between the high and low reward location in Experiment 2, the pattern of results from the greatest

increase towards the high, low, then no reward locations indicates a prioritization of saccade activity and overt attention towards locations of higher reward value.

Taken together, and in accordance with the previously discussed literature on the anti-saccade task, we propose that the relative slowing at the high reward location in Experiment 1 was due to the requirement for greater suppression of the learned saccade to that location. The second question, whether improvements in saccade latency to (non-) rewarded locations led to similar improvements in covert attention to those locations, was also answered here: since differences in saccade latencies were seen between rewarded and non-rewarded locations during training, but no differences were observed in the covert attention task, this leads us to conclude that the learning of reward through eye movements did not transfer to covert attention in a new setting involving different stimuli, and in which a manual rather than saccade motor response was required.

4. GENERAL DISCUSSION

The current study investigated the extent to which selection history biases in the form of reward learning transfer across tasks, context and motor outputs. Specifically, we examined whether reward-based biases of overt attention to locations in space would transfer to another task and attentional domain. The underlying notion is that through reward-based learning, spatial priority maps can be shaped such that specific locations on the priority map become behaviourally more salient than other locations (Chelazzi *et al.*, 2014; Hickey, Chelazzi and Theeuwes, 2014). According to Chelazzi et al. (2014) acquired biases of spatial attention are persistent, nonstrategic in nature, and should generalize across stimuli and task contexts.

Previous studies demonstrating the transfer of acquired biases of spatial attention have generally used attentional tasks involving manual responses during both training and test phases (Chelazzi *et al.*, 2014; Hickey, Chelazzi and Theeuwes, 2014). Because covert and overt attention are assumed to be linked by a common neural architecture (Corbetta, 1998; Beauchamp *et al.*, 2001) we wished to determine whether the priority map when shaped by overt attention (i.e., eye movements) would transfer to a different task that required only covert attention. Our results show that the eye movement task was successful in shaping the priority map as saccades to rewarded locations were executed faster than to non-rewarded locations. This implies that particular locations within the priority map became behaviourally more salient. However, following this training task, during a covert attention task, observers were not faster to execute a manual response to a target presented at the behaviourally most salient location, and instead were slower to respond to a target presented there. A second experiment in which participants were not obliged

to stay fixated during the covert attention task did not reveal any benefits in manual response times to the more behaviourally salient location. On the basis of these findings, we conclude that reward-based training involving eye movements does not result in the shaping of a "general" spatial priority map which affects covert deployment of attention.

In both experiments of our study, the training of overt eye movements led to a decrease in saccade latency to all locations as the training progressed; a general learning effect seen in most experimental settings (Anderson, Laurent and Yantis, 2011a, 2011b; Le Pelley *et al.*, 2015). Throughout training, saccades to high and low reward locations became significantly faster than those to no reward locations. In addition to reward availability, another factor that may have played a role in this general improvement to rewarded locations was information sampling. Eye movements to these locations were always followed by information in the form of a reward value (10 cents or 1 cent), with no information to be obtained at the no reward locations. The opportunity to gain information has been shown to drive and influence saccade behaviour, particularly in novel settings (Gottlieb, 2012; Gottlieb *et al.*, 2014). The reward-modulated reduction in saccade latency during learning is in agreement with previous research reporting reduced response times to rewarded objects or locations, for both overt and covert attention tasks (Milstein and Dorris, 2007; Theeuwes and Belopolsky, 2012; Failing and Theeuwes, 2014; Hickey, Chelazzi and Theeuwes, 2014). It has been shown that repetition of a chosen action based on the reward received for that behaviour, leads to a more accurate and speeded response, becoming automatic over time (Berridge and Robinson, 1998; Libera and Chelazzi, 2009; Hickey, Chelazzi and Theeuwes, 2010). This conversion from voluntary actions to automatic or habitual behaviour is suggested to occur via distinct, parallel circuits in the brain (Dickinson, Nicholas and Adams, 1983; Hikosaka *et al.*, 2018) and has been linked to maladaptive behaviours, such as various forms of addiction (Everitt and Robbins, 2005, 2013). The outcome of the covert attention task is different to that found in other studies demonstrating reduced manual RTs at spatial locations associated with higher reward (Chelazzi *et al.*, 2014; Failing and Theeuwes, 2014; Stankevich and Geng, 2014). However, in all of these other studies the required response was only ever manual, in both the training and test phases. The covert discrimination task in both of our experiments revealed a decrease in manual RTs at all locations after training, also demonstrating a general practice effect. However, in Experiment 1 the covert task additionally showed a relative slowing of responses to the target when it was presented at the high reward location. Specifically, there was a significant increase in RT to the high compared to low reward location. Both no reward locations showed identical reductions in RT; since these two locations were always in opposite hemifields to each other, this

revealed a location-specific rather than general hemifield disadvantage in responses to the high reward location. The difference at the high relative to low reward location during the test phase shows that, although eye movement latencies followed a similar pattern for these locations during learning, this general reward- and/or information-driven improvement did not transfer to the covert task. Instead, biased competition between rewarded locations led to a clear separation in how motivational or behavioural biases translated across the two tasks.

Brain mechanisms

The relative slowing at the high reward location during the test phase of Experiment 1 may be explained by an increased motivational association between the high reward location and the eye movement response, resulting in the slowing down in manual responses to that location during covert attention when the learned eye movement response was suppressed. Such a reward-based link between location and eye movements has been demonstrated in previously researched in monkeys, specifying the caudate nucleus (CN) and substantia nigra pars reticulata (SNr) of the basal ganglia as the brain regions responsible for this association (Hikosaka *et al.*, 2000; Ding and Hikosaka, 2006; Nakamura and Hikosaka, 2006a; Yamamoto *et al.*, 2012; Yasuda, Yamamoto and Hikosaka, 2012; Hikosaka, 2014). The past selection history bias described by Awh et al. (2012) may be implemented in the form of reinforcement learning, whereby humans and animals voluntarily choose (or avoid) actions that have previously been reinforced in a positive (or negative) manner (Skinner, 1938; Sutton and Barto, 1998; Dayan and Balleine, 2002; Schultz, 2006). Neuronal activation in LIP and FEF representing behaviourally relevant stimuli projects to the CN. CN output is then projected, via inhibitory connections with the SNr, to the SC. Since LIP has been proposed as an integrated priority map, and forms recurrent connections with FEF (Hikosaka *et al.*, 2000), these brain areas are relevant for both the biased competition model (Desimone & Duncan, 1995) and the past selection history framework (Awh, Belopolsky and Theeuwes, 2012).

Extensive research on the anti-saccade task describes the suppression of the automatic pro-saccade to the target as the first step in performing an anti-saccade (Amador, Schlag-Rey and Schlag, 1998; Bell, Everling and Munoz, 2000; Munoz and Everling, 2004). This inhibition of saccades account provides an explanation for the greater interference and increased manual RT observed in Experiment 1 when the target was at the high reward location. This was made especially clear via our analysis on those participants who were trained with the high and low reward locations at mirror-opposite locations to each other, representing the case most similar to the anti-saccade task. Our analogy to the anti-saccade task may be limited due to the fact that our

covert attention task was not established to examine eye movements in one direction or the other, and instead was entirely constrained in this overt behaviour.

Based on our findings, we suggest that spatial-based reward priority is intimately linked to the learned motor response. In addition, since we used different task contexts across the training and test phase, this may also explain why we did not see a transfer from training to the test phase as was reported in Chelazzi et al. (2014). During the training phase of our experiment, participants were presented with circle stimuli at particular locations. During the test phase, presented stimuli were letter shapes at these locations, and were also shown in a different color to the training stimuli. In the study of Chelazzi and colleagues, all stimuli were positioned inside square place-holders that were identical across the training and test phases. Due to this design, it is likely that in their study participants could more easily assign value to the consistent square objects tied to specific spatial locations. Our attempt to make the reward associations entirely spatial-based in nature, with no consistent stimulus features across tasks, may have interfered with such a transfer mechanism.

Chapter 4

Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease

Adapted from

McCoy, B., Jahfari, S., Engels, G., Knapen, T.* & Theeuwes, J.* (2019)

Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease

Brain, 142(11): 3605-3620

* denotes equal contributions

ABSTRACT

Reduced levels of dopamine in Parkinson's disease (PD) contribute to changes in learning, resulting from the loss of midbrain neurons that transmit a dopaminergic teaching signal to the striatum. Dopamine medication used by PD patients has previously been linked to behavioural changes during learning as well as to adjustments to value-based decision-making after learning. To date, however, very little is known about the specific relationship between dopaminergic medication-driven differences during learning and subsequent changes in approach/avoidance tendencies in individual patients. 24 PD patients on and off dopaminergic medication and 24 healthy controls (HC) underwent functional magnetic resonance imaging while performing a probabilistic reinforcement learning experiment. During learning, dopaminergic medication reduced an overemphasis on negative outcomes. Medication reduced negative (but not positive) outcome learning rates, while concurrent striatal BOLD responses showed reduced prediction error sensitivity. Medication-induced shifts in negative learning rates were predictive of changes in approach/avoidance choice patterns after learning, and these changes were accompanied by systematic striatal BOLD response alterations. These findings elucidate the role of dopamine-driven learning differences in PD, and show how these changes during learning impact subsequent value-based decision-making.

INTRODUCTION

Learning from trial and error is a core adaptive mechanism in behaviour (Packard, Hirsh and White, 1989; Glimcher, 2002). This learning process is driven by reward prediction errors (RPEs) that signal the difference between expected and actual outcomes (Houk, 1995; Montague, Dayan and Sejnowski, 1996; Schultz, Dayan and Montague, 1997). Substantia nigra (SN) and ventral tegmental area (VTA) midbrain neurons use bursts and dips in dopaminergic signalling to relay positive and negative RPE to prefrontal cortex (Deniau, Thierry and Feger, 1980; Swanson, 1982) and the striatum, activating the so-called Go and NoGo pathways (Beckstead, Domesick and Nauta, 1979; Surmeier *et al.*, 2007).

PD is caused by a substantial loss of dopaminergic neurons in the SN (Edwards, Quinn and Bhatia, 2008), leading to the depletion of dopamine in the striatum (Koller and Melamed, 2007). Dopaminergic medication has been shown to alter how PD patients learn from feedback (Cools *et al.*, 2001; Bódi *et al.*, 2009) and how they use past learning to make value-based choices in novel situations (Frank, Seeberger and Reilly, 2004; Frank, 2007; Shiner *et al.*, 2012). A common finding is that, when required to make value-based decisions after learning, patients ON compared to OFF medication are better at choosing the option associated with the highest value (approach), whereas when OFF medication, they are better at avoiding the option with the lowest value (avoidance) (Frank, Seeberger and Reilly, 2004; Frank, 2007). However, it is currently unknown how dopamine-induced changes during the learning process relate to these subsequent dopamine-induced changes in approach/avoidance choice behaviour.

An influential framework of dopamine function in the basal ganglia proposes that the dynamic range of phasic dopamine modulation in the striatum, in combination with tonic baseline dopamine levels, gives rise to the medication differences observed in PD (Frank, 2005). This theory suggests that lower baseline dopamine levels in unmedicated PD are favorable for the upregulation of the NoGo pathway, leading to an emphasis on learning from negative outcomes. In contrast, higher tonic dopamine levels in medicated PD lead to continued suppression of the NoGo pathway, resulting in (erroneous) response perseveration even after negative feedback. Extremes in these medication-induced changes in brain signalling are thought to manifest behaviourally in dopamine dysregulation syndrome, in which patients exhibit compulsive tendencies, such as pathological gambling or shopping (Voon *et al.*, 2010). In support of the theory on Go/NoGo signalling, impairments in learning performance associated with higher dopamine levels have been found mainly in negative-outcome contexts; during probabilistic selection (Frank, Seeberger and Reilly, 2004), reversal learning (Cools, Altamirano and D'Esposito, 2006a), and probabilistic classification (Bódi *et al.*, 2009).

In addition to these behavioural adaptations, increased striatal activations have been reported in medicated PD patients during the processing of negative RPEs (Voon *et al.*, 2010). Similarly, a recent study on rats performing a reversal learning task revealed a distinct impairment in the processing of negative RPE with increased dopamine level (Verharen *et al.*, 2018). However, very little is known about how these medication-related changes in striatal responsivity to RPE relate to 1) later behavioural choice patterns, and 2) changes in brain activity during subsequent value-based choices.

We examined the role of dopaminergic medication in choice behaviour and associated brain mechanisms. 24 PD patients ON and OFF medication and a reference group of 24 age-matched HC performed a two-stage probabilistic selection task (Frank, Seeberger and Reilly, 2004) (Fig. 1A) while undergoing functional magnetic resonance imaging (fMRI). The experiment's first stage was a learning phase, during which participants gradually learned to make better choices for three fixed pairs of stimulus options, based on reward feedback. In the second, transfer stage, participants used their learning phase experience to guide choices when presented with novel combinations of options, without receiving any further feedback (Fig. 1A). Value-based decisions during the transfer phase were examined using an approach/avoidance framework (Fig. 1B). To better describe the underlying processes that contribute to learning, behavioural responses were fit using a hierarchical Bayesian reinforcement learning model (Jahfari *et al.*, 2018; Van Slooten *et al.*, 2018), adapted to estimate both within-patient effects of medication and across-subject effects of disease (Sharp *et al.*, 2016). This quantification of behaviour then informed our model-based fMRI analysis, in which we examined medication-related changes in blood oxygen level dependent (BOLD) brain signals in response to RPEs during learning, as well as medication-related changes in approach/avoidance behaviour and brain responses during subsequent value-based choices.

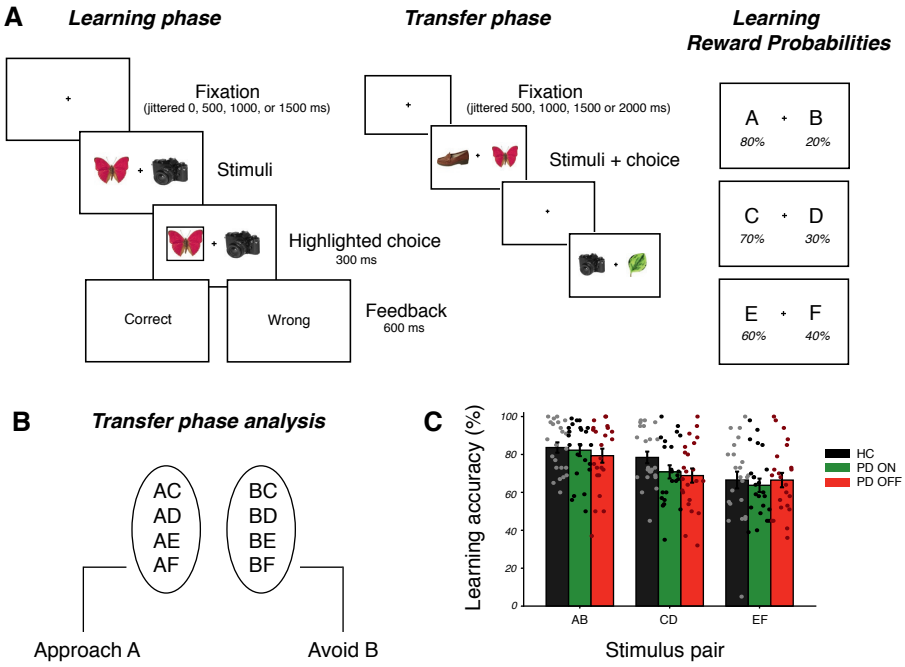


Figure 1 Experimental design and learning performance. (**A)** *Learning phase*: on each trial participants chose between two everyday objects and observed a probabilistic outcome 'correct' or 'wrong', corresponding to winning 10 cents or nothing. Each participant viewed three fixed pairs of stimuli (AB, CD, and EF) and tried to learn which was the best option of each pair, based on the feedback received. Reward probability contingency per stimulus during learning is shown on the right. *Transfer phase*: participants were presented with all possible combinations of stimuli from the learning phase and had to choose what they thought was the better option, based on what they had learned. No feedback was provided in this phase. (**B)** The transfer phase analysis was performed on correctly choosing A on trials in which A was paired with another stimulus (approach accuracy) or correctly avoiding B on trials where B was paired with another stimulus (avoidance accuracy). **(C)** Accuracy in choosing the better option of each pair across each group during learning (mean ±1 SEM). Parameter estimates of these medication and disease effects are presented in Supplementary Fig. 1.

MATERIALS AND METHODS

Participants

Twenty-four patients with PD (7 female, mean age = 63 ± 8.2 years old) were recruited via the VU medical center, Zaans medical center, and OLVG hospital in Amsterdam. All patients were diagnosed by a neurologist as having idiopathic PD according to the UK Parkinson's Disease Society Brain Bank criteria. This study was approved by the Medical Ethical Review committee (METc) of the VU Medical Center, Amsterdam. Twenty-four age-matched HCs (9 female, mean age = 60.3 ± 8.5 years old) were also recruited from the local community or via the PD patients (e.g. spouses, relatives). In total, five spouses of PD patients were included in the HC sample. At each session of the study, the severity of clinical symptoms was assessed according to the Hoehn and Yahr rating scale (Hoehn and Yahr, 1967) and the motor part of the Unified PD Rating Scale (UPDRS III; Fahn *et al.*, 1987). Demographic and clinical data of the included participants can be seen in Supplementary Table 1. Information on Parkinson-related medication per patient is available in Supplementary Table 2. We excluded one PD patient (excessive falling asleep in scanner) and one HC (could not learn the task) from both learning and transfer phase behavioural and fMRI analyses. fMRI data of one HC could not be analyzed (T1 scan was not collected; session was terminated early due to claustrophobia). Transfer phase fMRI and behavioural data were not collected for one other HC due to early termination of scanning session (technical malfunction). Overall, we included 23 PD patients ON and OFF dopaminergic medication in all behavioural and fMRI analyses. 23 HCs were included in learning phase behavioural analysis, 22 in learning phase fMRI analysis, and 21 in transfer phase behavioural and fMRI analyses. Additional participant information is provided in the Supplementary Information.

Procedure

The study was set up as a dopaminergic-manipulation, within-subject design in PD patients, to reduce the variance associated with inter-individual differences. All PD and HC participants took part in at least two sessions, the first of which was always a neuropsychological examination (lasting 2 hours; 30 minutes of which were spent practicing the reinforcement learning task with basic-shape stimuli). PD patients subsequently participated in two separate fMRI scanning sessions (once in a dopamine-medicated "ON" state and once in a lower dopamine "OFF" state), and HCs underwent one fMRI session. The patient fMRI sessions were carried out over the same weekend in all but one patient (2 weeks apart) and were counterbalanced for ON/OFF medication order. All OFF sessions had to be carried out in the morning for ethical reasons. Patients were instructed to withhold from taking their

usual dopamine medication dosage on the evening prior to and the morning of the OFF session, thereby allowing >12 hours withdrawal at the time of scanning. Patients on dopamine-agonists (pramipexole, ropinerol) took their final dopamine-agonist dose on the morning prior to the day of scanning (~ 24 hours withdrawal). One PD patient took his medication 8.5 hours before OFF day scanning to relieve symptoms but was nevertheless included in the analysis.

Neuropsychological assessment

Participants completed a battery of neuropsychological tests on their first visit. A description of these tests and self-report questionnaires, along with group results, is included in Supplementary Table 1. All patients used their dopaminergic medication as usual during this session. These assessments were not examined in the current study, but are discussed in greater detail elsewhere (G Engels *et al.*, 2018; Gwenda Engels *et al.*, 2018).

Reinforcement Learning Task

Participants completed a probabilistic selection reinforcement learning task consisting of two stages; a learning phase and transfer phase. This task has been used in several previous studies, in both PD patients (Frank, Seeberger and Reilly, 2004; Shiner *et al.*, 2012; Grogan *et al.*, 2017) and healthy participants (Jocham, Klein and Ullsperger, 2011; Jahfari *et al.*, 2018; Van Slooten *et al.*, 2018). We used pictures of everyday objects from different object categories, such as hats, cameras, and leaves (stimulus set extracted from Konkle *et al.*, 2010, *17-objects.zip*).

Learning phase

In the learning phase, three different pairs of object stimuli (denoted as AB, CD and EF) were repeatedly presented in random order. Each pair had specific reward probabilities associated with each stimulus, and participants had to learn to choose the best option of each pair based on the feedback provided (Fig. 1A). Participants were instructed to try to find the better option of a pair in order to maximize reward. Feedback was either "Goed" or "Fout" text (meaning "correct" or "wrong" in Dutch), indicating a payout of 10 cents for correct trials and nothing for incorrect trials. Different objects were used across each fMRI session of patients, so as not to induce any familiarity or reward associations with particular stimuli. In the "easiest" AB pair, the probability of receiving reward was 80% for the A stimulus and 20% for the B stimulus, with ratios of 70:30 for CD and 60:40 for EF. The EF pair was therefore the hardest to learn due to more similar reward probabilities between the two

options. All object stimuli were counterbalanced for reward probability pair and for better versus worse option of a pair across subjects (for instance, a leaf and hat as the A and B stimuli for one participant were the D and C stimuli for another participant). In total, there were 12 object stimuli and each participant viewed six of these objects in a given fMRI session, with PD patients viewing the remaining 6 stimuli in their second fMRI session. The learning phase consisted of two runs of 150 trials each (totalling 100 trials per stimulus pair). Each run was interspersed with 15 null trials to improve model fitting of this rapid event-related fMRI design. Null trials, during which only the fixation cross was presented, lasted at least 4 seconds plus an additional interval generated randomly from an exponential distribution with a mean of 2 seconds. Each task trial had a fixed duration of 5000 ms, and began with a jittered interval of 0, 500, 1000, or 1500 ms to obtain an interpolated temporal resolution of 500 ms. During the interval, a black fixation cross was presented and participants were asked to hold fixation. Two objects were then presented simultaneously left and right of the fixation cross (counterbalanced across left/right locations per pair) and remained on the screen until a response was made. If a response was given on time, a black frame surrounding the chosen object was shown (300 ms) and followed by feedback (600 ms). Omissions were followed by the text “te langzaam” (“too slow” in Dutch). The fixation cross was displayed alone after feedback was presented, until the full trial duration was reached.

Transfer phase

In the transfer phase, novel pairings of all possible combinations of the six stimuli were presented in addition to the original three stimulus pairs, thereby making up 15 possible pairings. This phase consisted of two runs of 120 trials each (8 trials per pair), and each run randomly interspersed with 12 null trials. The duration of these null trials was generated in the same way as in the learning phase. Participants were instructed to choose what they thought was the better option, given what they had learned. There was no feedback in this phase and no frame surrounded the chosen response. Each trial began with a jittered interval of 500, 1000, 1500 or 2000 ms, with a new trial starting whenever a response was made.

Learning and Transfer

Each object stimulus was presented equally often on the left or right side in both learning and transfer phases. Responses were made with the right hand, using the index or middle finger to choose the left or right stimulus, respectively. One patient was uncomfortable using two fingers of the right hand

and so responded with the left and right index finger on separate button boxes (in both ON and OFF sessions). The feedback text was made larger for one patient in both ON and OFF sessions to make it easier to read.

Computational model

The Q-learning reinforcement learning algorithm (Sutton and Barto, 1998) captures trial-by-trial updates in the expected value of options and has been used extensively to model behaviour during learning (Daw *et al.*, 2011; Jocham, Klein and Ullsperger, 2011; Schmidt *et al.*, 2014; Grogan *et al.*, 2017; Jahfari *et al.*, 2018). We used a variant of this model with three free parameters, allowing us to determine how subjects learned separately from positive and negative feedback (α_{gain} and α_{loss}) and how much they exploited differences in value between stimulus pair options (β). In hierarchical models, group and individual parameter distributions are fit simultaneously and constrain each other, leading to greater statistical power over standard non-hierarchical methods (Ahn, Krawitz and Kim, 2011; Steingroever, Wetzels and Wagenmakers, 2013; Wiecki, Sofer and Frank, 2013; Kruschke, 2015; Jahfari *et al.*, 2018). We also fit a separate model with an additional free parameter, relating to persistence of choices irrespective of feedback, and performed model comparison, allowing us to verify that the chosen model better represented the data (see Supplementary Table 3). These models were performed using R (R Development Core and Team, 2017) and RStan (Stan Development Team, 2014).

Subject-level Q-learning model

The Q-learning algorithm assumes that after receiving feedback on a given trial, subjects update their expected value of the chosen stimulus (Q_{chosen}) based on the difference between the reward received for choosing that stimulus (r = 1 or 0 for reward or no reward, respectively) and their prior expected value of that stimulus, according to the following equation:

$$Q_{chosen}(t+1) = Q_{chosen}(t) + \begin{cases} \alpha_{gain}[r(t) - Q_{chosen}(t)], & if\ r = 1 \\ \alpha_{loss}[r(t) - Q_{chosen}(t)], & if\ r = 0 \end{cases} \quad \textbf{(1)}$$

The term $r(t) - Q_{chosen}(t)$ is the reward prediction error (RPE). Accordingly, choices followed by positive feedback (r = 1) were weighted by the α_{gain} learning rate parameter and choices followed by negative feedback (r = 0) were weighted by the α_{loss} learning rate parameter ($0 < \alpha_{gain}, \alpha_{loss} < 1$). All Q-

values were initialized at 0.5 (no initial bias in value). The probability of choosing one stimulus over another is described by the softmax rule:

$$P_{chosen}(t) = \frac{exp(\beta \times Q_{chosen}(t))}{exp(\beta \times Q_{unchosen}(t)) + exp(\beta \times Q_{chosen}(t))} \quad \textbf{(2)}$$

where β is known as the inverse temperature or "explore-exploit" parameter ($0 < \beta < 100$). Effectively, β is used as a weighting on the difference in value between the two options. The free parameters α_{gain}, α_{loss} and β were fit for each individual subject, in a combination that maximizes the probability of the actual choices made by the subject.

Fig. 2A shows a graphical representation of the model. The free parameters α_{gain} and α_{loss} are labelled as α_G and α_L for viewing purposes. The quantities $r_{i,\,t-1}$(reward for participant *i* on trial *t-1*) and $ch_{i,t}$ (choice for participant *i* on trial *t*) are obtained directly from the data. The subject-level quantities α_{Gi}, α_{Li} and β_i are deterministic, and were transformed during estimation using the inverse probit (phi) transformation Z'_i (α'_{Gi}, α'_{Li}, β'_i), which is the cumulative distribution function of a unit normal distribution. An apostrophe attached to parameters indicates that a phi transformation was applied to these parameters. The transformed parameters have no apostrophe. The parameters Z'_i (i.e., α'_{Gi}, α'_{Li}, β'_i) lie on the probit scale covering the entire real line. In this way, transformed parameters were obtained by applying an inverse probit transformation to normally-distributed priors centered on zero, with a standard deviation of 1, e.g., $\mu_{\alpha' G} \sim N(0,1)$. Weakly informative priors such as these are recommended in small sample sizes to reduce the influence of the priors on posterior distributions (Gelman et al., 2013; Ahn et al., 2017). This guarantees that the converted priors will be uniformly distributed between 0 and 1 (Wetzels *et al.*, 2010; Ahn *et al.*, 2014; Ahn, Haines and Zhang, 2017). The calculation for the transformed β parameter included a multiplicative factor of 100 in the same step as the transformation to allow for a range between 0 and 100. Following recommendations from the Stan development team (2016) we used non-centered reparameterization to reduce the dependency between μ_z', δ_z' and Z'_i when for example, moving from α'_{Gi} to α_{Gi} with the phi transformation (please see below for elaboration, or Ahn et al. 2017 for more examples with non-centered reparameterization). Stan provides a fast approximation of the inverse probit transformation with the *Phi_approx* function.

Group-level Q-learning model

The subject-level model described above was nested inside a group-level model in a hierarchical manner (Ahn, Haines and Zhang, 2017). Parameters Z'_i were drawn from group-level normal distributions with mean μ_z' and standard deviation δ_z'. A normal prior was assigned to group-level means $\mu_z' \sim N(0,1)$, and a half-Cauchy prior to the group-level standard deviations $\delta_z' \sim Cauchy(0,5)$. The model was extended in two ways in accordance with Sharp *et al.*, 2016. To capture medication-related shifts (PD ON vs. OFF) in each of the three parameters, we included three additional parameters on both the subject level and on the group level (Fig. 2C, D). Similarly, we incorporated three additional parameters to capture disease-related differences (HC vs. PD) on the group level.

For the α_{gain} parameters, these were: *Med*_α_G*'p* (for the effect of medication on α_{gain} in PD patient *p*) and *Dis*_α_G*'h* (for the effect of *no* disease on α_{gain} in HC participant *h*), with the analogous terms for α_{loss} (*Med*_α_L*'p* and *Dis*_α_L*'h*) and β (*Med*_β*'p* and *Dis*_β*'h*). Symmetric boundaries for all phi transformed *Med* and *Dis* parameter distributions were used to constrain the model and assist with convergence (-5 < *Med, Dis* < 5). These boundaries were adopted from recent work with a similar hierarchical Bayesian parameter approach (Pedersen, Frank and Biele, 2016). Prior to committing to these bounds we evaluated two alternative bounds for these parameters, with either -1 < *Med, Dis* < 1 or -10 < *Med, Dis* < 10. The [-1,1] bounds were found to be too conservative, as posterior distributions were cut off at boundary values. In contrast, the [-10,10] bounds were overly liberal, since the distributions were well-contained within the [-5,5] interval. Group-level priors were the same as those on the subject-level, i.e., a normal prior was assigned to the group-level means of all the *Med* and *Dis* free parameters, e.g., *Med*_$\mu_{\alpha' G} \sim N(0,1)$, and a half-Cauchy prior was applied to all group-level standard deviations, e.g., *Med*_$\sigma_{\alpha' G} \sim Cauchy(0,5)$.

As in the mixed-effects logistic regressions, we took PD OFF as "baseline" by using two binary indicators: $I('on') = 0$, and $I('healthy') = 0$. PD ON was coded as $I('on') = 1$, $I('healthy') = 0$, and HC was coded as $I('on') = 0$, $I('healthy') = 1$. For subject s and medication condition m, the phi transformed α_{gain} parameter (denoted as αG below) of an individual subject was formulated as follows:

$$\alpha G_{s,m} = Phi_{approx}\left(\mu_{\alpha G} + \left(\sigma_{\alpha G} \times \alpha' G_{s,m}\right) + \left[Med_{\alpha G} \times I(m,' on')\right] + \left[Dis_{\alpha G} \times I(s,' healthy')\right]\right) \quad \textbf{(3)}$$

As mentioned, Phi_{approx} is an approximation of the inverse probit transformation, a function provided by Stan for efficient computation. We used a non-centered reparameterization technique to move from $\alpha' G_{s,m}$ to $\alpha G_{s,m}$; a Normal(μ, σ) distribution can be reparameterized and sampled from a unit normal distribution that is multiplied by the scale parameter σ and then shifted by the location parameter μ (Stan Development Team, 2015; Ahn, Haines and Zhang, 2017). Using the binary indicators described above, PD OFF did not contain either of the *Med_αG* or *Dis_αG* terms, PD ON included the *Med_αG* term to indicate the within-subject effect of medication, and HC included the *Dis_αG* term to denote the between-subject effect of disease. α_{loss} and β parameters were distributed in the same way with their corresponding terms. Since the medication effect was within-subject, it was itself a subject-specific random variable with its own population-level mean and variance. Once again using non-centered reparameterization, the medication effect was formulated as follows:

$$Med_\alpha G_s = Phi_{approx}\left(Med_\alpha G + \left(\sigma_{Med_\alpha G} \times Med_\alpha' G_s\right)\right) \quad \textbf{(4)}$$

Please see Supplementary Information for the model estimation procedure and Supplementary Fig. 2 for an evaluation of the model fit. Bayes factors (BFs) of group level posterior distributions for medication and disease differences were calculated as the ratio of the posterior density above zero relative to the posterior density below zero (Pedersen, Frank and Biele, 2016). This method is possible since the priors for the distributions of these parameters were symmetric (unbiased) around zero (Marsman and Wagenmakers, 2017). Categories of evidential strength of an effect are based on (Jeffreys, 1998), with BFs > 10 considered as strong evidence that the shift in the posterior distribution is different from zero. We provide all fitting code online, at: https://github.com/mccoyb4/Parkinson_RL.

Statistical evaluations of Behaviour

General

Since PD patients were tested twice and HC participants only once, we ascertained that session order effects did not affect performance during either the learning phase or transfer phase (please see Supplementary Information and Supplementary Fig. 3).

Learning phase

Bayesian mixed-effects logistic regression modelling was carried out on trial-by-trial behaviour (Wunderlich, Smittenaar and Dolan, 2012; Doll *et al.*, 2016; Sharp *et al.*, 2016). These analyses were performed in R (R Development Core and Team, 2017), using the Bayesian Linear Mixed-Effects Models (blme) package (Chung *et al.*, 2013), built on top of lme4 (Bates *et al.*, 2014). In our mixed-effects models, we coded for both fixed and random trial-by-trial effects and allowed for a varying intercept on a per subject basis. For the model on learning behaviour, the dependent variable was accuracy in choosing the better stimulus of a pair (correct=1, incorrect=0). Stimulus pair ("*Pair*") was taken as a within-subject (random-effect) explanatory variable (EV), from easiest to most difficult (AB pair=1, CD pair=0, EF pair=-1). We also included two binary covariates (as in Sharp *et al.*, 2016); the between-subject effect of disease (*Dis*, where PD=0, HC=1) and the within-subject effect of dopaminergic medication state (*Med*, where OFF=0, ON=1), as well as their interactions with the stimulus pair variable. The medication variable for HC was coded as 0 as we wanted this to capture only the within-subject effect of medication. Since disease and medication status were both included in the same model, PD OFF was considered to act as a baseline (*Dis*=0, *Med*=0). Within-subject effects of medication for PD ON (*Dis*=0, *Med*=1) were therefore captured by the medication variable only and between-subject effects of disease for HC (*Dis*=1, *Med*=0) were captured by the disease variable only (with *Dis*=1 meaning "healthy"). This is summarized in the following regression equation:

$$\text{Correct} = \text{Pair} + \text{Med} + \text{Dis} + \text{Pair*Med} + \text{Pair*Dis} + \text{Subject Intercept} \quad \textbf{(5)}$$

Positive beta estimates obtained from the model therefore indicate higher accuracy for either PD ON or HC compared to PD OFF in the *Med* and *Dis* variables, respectively, with negative estimates for those variables reflecting greater accuracy for PD OFF.

Transfer phase

The mixed-effects regression on transfer phase behaviour was carried out on trials in which either the A or B stimulus appeared, excluding those in which both appeared together (see Fig. 1B). The expectation was that participants should opt to choose A (Approach A) and avoid choosing B (Avoid B) whenever they were presented, since they were associated with the highest and lowest reward probabilities during learning, respectively. The regression was performed similarly to that in the learning phase, except that the stimulus pair variable was replaced with an Approach A / Avoid B trial variable (A=1, B=-1). The dependent variable (accuracy) was then coded as 1 for correctly choosing A in Approach A or correctly not choosing B in Avoid B trials, and as 0 for incorrectly

choosing the other option for each trial type. Medication and disease status were included as covariates, with a varying intercept per subject. To assess the role of medication and disease status on Approach A and Avoid B performance separately, we carried out a regression analysis on each subset, with the same covariates as described previously.

Learning and Transfer

The relationship between medication-induced shifts during learning and transfer was evaluated in two steps. First, we compared three multiple regression models, as shown in Supplementary Table 4, to evaluate how the learning rate medications shifts (i.e, Med_αG, Med_αL, or both) relate to the transfer phase Approach/Avoid shifts on an individual level. In these (multiple) regression models, the approach/avoid shift (defined for each subject as the OFF > ON medication difference in Avoid B > Approach A accuracies) was set as the dependent variable. Next, BIC was computed for each regression (with EVs being either only Med_αG, Med_αL, or both), to select the optimal model for the evaluation of medication relationships between the learning and transfer phase. Individual learning-rate medication differences were quantified as the modes of the within-subject medication difference parameter distributions, to capture peak probability densities (see Supplementary Figure 4).

fMRI image acquisition

fMRI scanning was carried out using a 3T GE Signa HDxT MRI scanner (General Electric, Milwaukee, WI, USA) with 8-channel head coil at the VU University Medical Center (Amsterdam, The Netherlands). Functional data for the learning and transfer phase runs were acquired using T2*-weighted echo-planar images with blood oxygenation level dependent (BOLD) contrasts, containing approximately 410 and 240 volumes for learning and transfer runs, respectively. The first two TR volumes were removed to allow for T1 equilibration. Each volume contained 42 axial slices, with 3.3 mm in-plane resolution, TR = 2,150 ms, TE = 35 ms, FA = 80 degrees, FOV = 240 mm, 64 x 64 matrix. Structural images were acquired with a 3D T1-weighted magnetization prepared rapid gradient echo (MPRAGE) sequence with the following acquisition parameters: 1 mm isotropic resolution, 176 slices, repetition time (TR) = 8.2 ms, echo time (TE) = 3.2 ms, flip angle (FA) = 12 degrees, inversion time (TI) = 450 ms, 256 x 256 matrix. The subject's head was stabilized using foam pads to reduce motion artifacts.

fMRI analysis

Preprocessing was performed using FMRIPREP version 1.0.0-rc2 (O. Esteban *et al.*, 2018; Oscar Esteban *et al.*, 2018), a Nipype-based tool (Gorgolewski *et al.*, 2011, 2017). On the learning phase data, we carried out a single-trial whole-brain analysis and deconvolution analyses on targeted striatal ROIs. For the transfer phase data, BOLD percent signal change was extracted for the relevant approach/avoidance conditions. See Supplementary Information for full details on each of these steps.

Code availability

Related analysis code is available at: https://github.com/mccoyb4/Parkinson_RL.

Data availability

For ethical reasons, we are unable to share the patient data. The raw data underpinning the findings of this study are available upon reasonable request from the corresponding author. These are in BIDS format and preprocessed with fMRIPrep to ease and encourage sharing upon request. fMRI statistics maps and associated tables of activated regions per group and per group comparison are available to view on figshare, at: https://doi.org/10.6084/m9.figshare.6989024.v2.

RESULTS

During the learning phase, participants successfully learned to choose the best option out of three fixed pairs of stimuli (Fig. 1C). Each pair was associated with its own relative reward probability among the two options, labeled as AB (with 80:20 reward probability for A:B stimuli), CD (70:30) and EF (60:40). Choice accuracy analysis showed that learning took place in PD OFF, PD ON and HC (N=23 in each group), with the probability with which participants chose the better option of each stimulus pair largely reflecting the underlying reward probabilities (PD ON: 82.3 % ± 3.1, 70.8 % ± 3.5, and 63.7 % ± 3.5; PD OFF: 76.6 % ± 3.4, 70.7 % ± 3.7, and 64.4 % ± 3.6; and HC: 83.7 % ± 2.7, 78.4 % ± 3.1, and 66.5 % ± 4.4 for AB, CD, and EF stimulus pairs, respectively).

We examined within- and between-subject differences in choice accuracy using a Bayesian mixed-effects logistic regression on the observed trial-by-trial behaviour (see Materials and Methods and Supplementary Fig. 1). This analysis assessed how choice accuracy was affected by stimulus pair, medication, disease status, and their interactions. When patients were ON medication, overall performance was more accurate in comparison to OFF, with the biggest difference for the easier AB choices and a smaller difference for the more uncertain EF pair. This was evidenced by a main effect

of stimulus pair (β (SE) = 0.35 (0.03), z = 10.19, p << .001), medication (β (SE) = 0.11 (0.04), z = 2.80, p = .005), and, specifically, an interaction between medication and stimulus pair (β (SE) = 0.17 (0.05), z = 3.47, p <.001). Importantly, this specific effect of medication was reflected in an analogous effect of disease when comparing PD OFF to HC, with a significant interaction between disease status and stimulus pair (β (SE) = 0.20 (0.05), z = 3.81, p < .001). Since learning of the AB pair plays a particularly important role in subsequent transfer phase choices during Approach A and Avoid B trials, we also carried out mixed-effects logistic regression analyses to assess how positive and negative feedback affect choice behaviour for the AB pair during learning. We found that in trials following negative, but not positive, feedback, PD ON chose the better A stimulus more often than PD OFF (β (SE) = 0.52 (0.13), z = 3.96, p <.001), indicating that PD ON are less likely to use negative outcomes to guide subsequent choices (see Supplementary Information for full details).

Overall, these first analyses show an improvement in choice accuracy when patients are ON compared to OFF medication, with performance on the easiest option pair restored to the level of HC. However, although choice accuracy provides us with a general assessment of medication effects on performance, it does not relate these effects to a mechanistic explanation of how underlying indices of learning might be affected by medication. These underlying mechanisms can be studied and defined both at the group level (HC vs. PD), and within-subject level (PD ON vs. OFF) by adopting a formal learning model of behaviour, to which we turn next.

Medication reduces learning rate for negative outcomes

Reinforcement learning theories describe how an agent learns to select the highest-value action for a given decision, based on the incorporation of received rewards (Rescorla and Wagner, 1972; Sutton and Barto, 1998). We implemented a Q-learning model, graphically represented in Fig. 2A-C, to describe both value-based decision-making and the integration of reward feedback in our experiment (Daw *et al.*, 2011; Jocham, Klein and Ullsperger, 2011; Schmidt *et al.*, 2014). Our model used separate parameters to describe, for a given agent, how strongly current value estimates are updated by positive (α_{gain}) and negative (α_{loss}) feedback, i.e. positive and negative learning rates (Grogan *et al.*, 2017; Jahfari *et al.*, 2018; Van Slooten *et al.*, 2018; Verharen *et al.*, 2018), as well as a parameter that determines the extent to which differences in value between stimuli are exploited (β). To understand how medication affects learning in PD we examined the posterior distributions of group-level parameters representing the within-subject medication shift in α_{gain}, α_{loss} and β (Fig. 2D). The large leftward shift of the α_{loss} posterior distribution indicates higher learning rates after

negative outcomes in PD OFF compared to ON (Bayes factor (*BF*) = 11.40). This is consistent with the theory that PD increases the sensitivity to negative outcomes, and that dopaminergic medication remediates specifically this disease symptom. Conversely, shifts in the distributions of the α_{gain} and β parameters were merely anecdotal (1<*BF*s<2, see Supplementary Table 5**,** and Supplementary Fig. 4 for individual within-subject effects of medication). For parameter comparisons between PD and HC based on disease status, we found strong evidence for a higher β, i.e. greater exploitation, in HC compared to PD (BF = 16.89; see Supplementary Fig. 5 and Supplementary Fig. 6), in addition to a similar effect on α_{loss}.

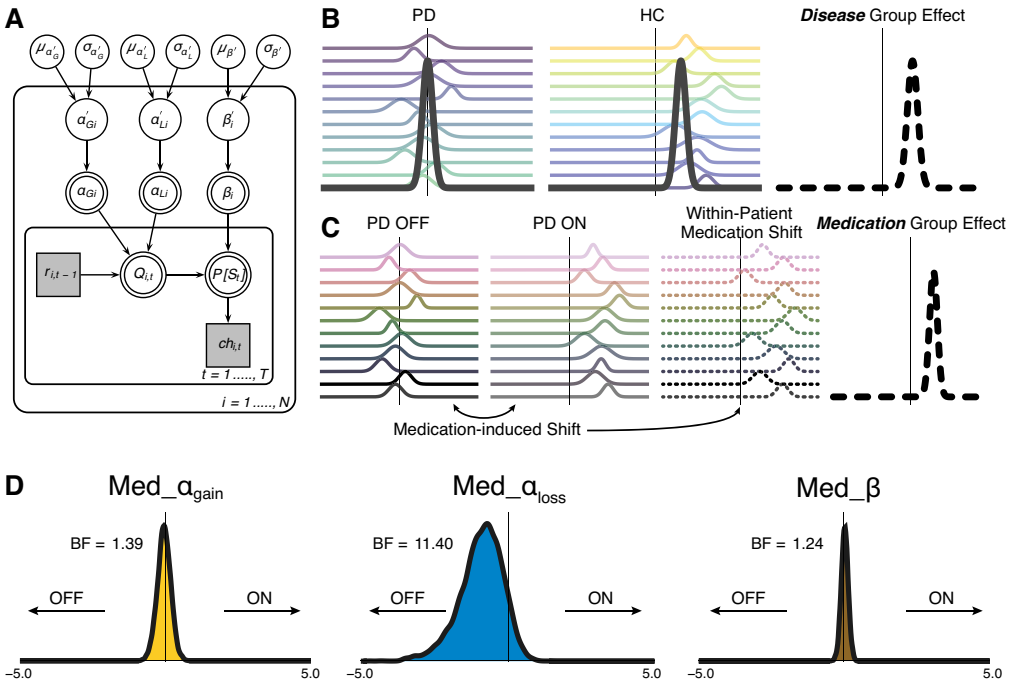


Figure 2. Modelling approach and medication-driven parameter shifts in PD. (A) Graphical outline of the Bayesian hierarchical Q-learning model with three free parameters, i.e., α_{gain} (denoted here as α_G), α_{loss} (denoted here as α_L) and β. The apostrophe attached to these parameters indicates that an inverse probit (phi) transformation was applied to the parameters (please see Materials and Methods for an elaborate description). The model consists of an outer subject (i=1,.....,N, including p=1,...,N_{PD}, and h=1...,N_{HC}), and an inner trial plane (t=1,...,T). Nodes represent variables of interest. Arrows are used to indicate dependencies between variables. Double borders indicate deterministic variables. Continues variables are denoted with circular nodes, and discrete variables with square nodes. Observed variables are shaded in grey. Per subject and session, $r_{i,t-1}$ is the reward received on the previous trial of a particular option pair, $Q_{i,t}$ is the current expected value of a particular stimulus, and $P[S_t]$ is the probability of choosing a particular stimulus in the

current trial. On top of the 3-parameter Q-learning model, dummy variables were defined in accordance with Sharp *et al*., 2016, to capture group-level disease-related differences in learning (denoted as: Dis_α_{gain}, Dis_α_{loss}, Dis_β), and within-subject medication differences (Med_α_{gain}, Med_α_{loss}, Med_β). A detailed graphical model including these group-level difference parameters is presented in Supplementary Fig. 10. **(B)** Graphical cartoon for the comparison of PD to HC in an illustrative Dis parameter. **(C)** Demonstration of the within-subject comparison of PD OFF to PD ON, resulting in both a subject-level and group-level posterior medication shift in an illustrative Med parameter. Please see Materials and Methods for a detailed description of the model with these subject/group difference parameters and definition of priors and transformations. **(D)** Group-level posteriors for medication shift in PD during the learning phase, for all parameters. A leftward shift in the Med_α_{loss} distribution indicates greater learning from negative outcomes in PD OFF compared to ON.

Medication in PD reduces the sensitivity of dorsal striatum to RPE

In the Q-learning model, the learning rate weighs the extent to which value beliefs are updated based on trial-by-trial RPE. The processing of choice outcomes is known to influence BOLD signals in the striatum, where the sensitivity to RPE is changed when dopamine levels are manipulated (Pessiglione *et al.*, 2006; Jocham, Klein and Ullsperger, 2011; Schmidt *et al.*, 2014). To establish whether RPE processing in the current study was influenced by dopaminergic state, we first examined within-subject medication-related differences in whole-brain responses to all positive and negative RPEs in the learning phase using a single-trial GLM (see Supplementary Information). This analysis provides an unbiased overview of any RPE-related (positive and/or negative) differences caused by dopaminergic medication across the entire brain. We found a significant PD OFF > ON medication difference in RPE modulation of the caudate nucleus and putamen (Fig. 3), and in several other regions including the globus pallidus interna and externa, thalamus, cerebellum, lingual gyrus and precuneus. Comparisons of HC with PD (ON and OFF) showed no RPE-related differences in the striatum, with significant RPE differences in frontal medial cortex, subcallosal cortex, and precuneus (HC > PD OFF) and in the occipital pole (HC > PD ON). The opposing contrasts, i.e. PD ON/OFF > HC, showed more extended activations, with RPE-related group differences in the paracingulate gyrus, superior frontal gyrus, frontal pole, supramarginal gyrus, cerebellum, occipital pole and lateral occipital cortex (PD OFF > HC) and in the cerebellum, brain-stem, and lateral occipital cortex (PD ON > HC). Because our model-based behavioural analysis revealed a medication-related difference specific to learning from negative outcomes (Fig. 2D), we proceeded by analyzing BOLD response time series to positive and negative outcomes separately.

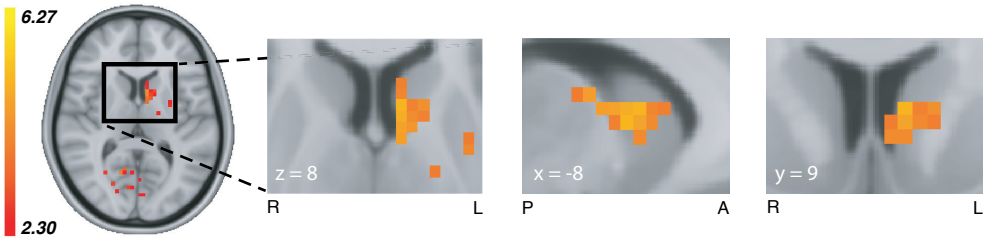


Figure 3. Whole-brain medication-related difference in RPE modulation. Whole-brain medication effects for the comparison PD OFF > ON in RPE-related modulations during the learning phase (z=2.3, p<.01, cluster-corrected), showing a dopamine-driven difference in the left dorsal striatum (see Supplementary Table 5 for a full list of brain region differences and contrast statistics). Whole-brain group-level contrasts of RPE and feedback valence are available to view on figshare, at: https://doi.org/10.6084/m9.figshare.6989024.v2.

Medication effects in dorsal striatum are specific to the processing of negative RPEs

To disentangle the separate effects of positive and negative RPE signalling, we examined feedback-triggered BOLD time courses from three independent striatal masks; the caudate nucleus, putamen, and nucleus accumbens (see Supplementary Information and Supplementary Figs. 7 and 8). We found a significant medication difference only in the caudate nucleus, in BOLD activity associated only with negative RPE (Fig. 4). RPE modulation of the BOLD response was greater in PD OFF compared to ON, during the interval 7.51s – 10.67s after the onset of negative feedback. Medication status did not alter the BOLD responses to positive RPE, indicating that changes due to dopaminergic medication are specific to negative RPE signalling in the caudate nucleus, the most dorsal part of the striatum. As well as tracking RPEs at the time of feedback, the striatum has been shown to represent the Q-value of the (to-be) chosen stimulus during the choice period (Kim *et al.*, 2009; Horga *et al.*, 2015; Jahfari, Theeuwes and Knapen, 2019). We therefore also performed a separate time-course analysis on the effect of Q-values on the BOLD signal in striatal ROIs during stimulus presentation (see Supplementary Information for details). This showed a medication-related increase in the modulation of BOLD by Q-values in the putamen (Supplementary Fig. 9).

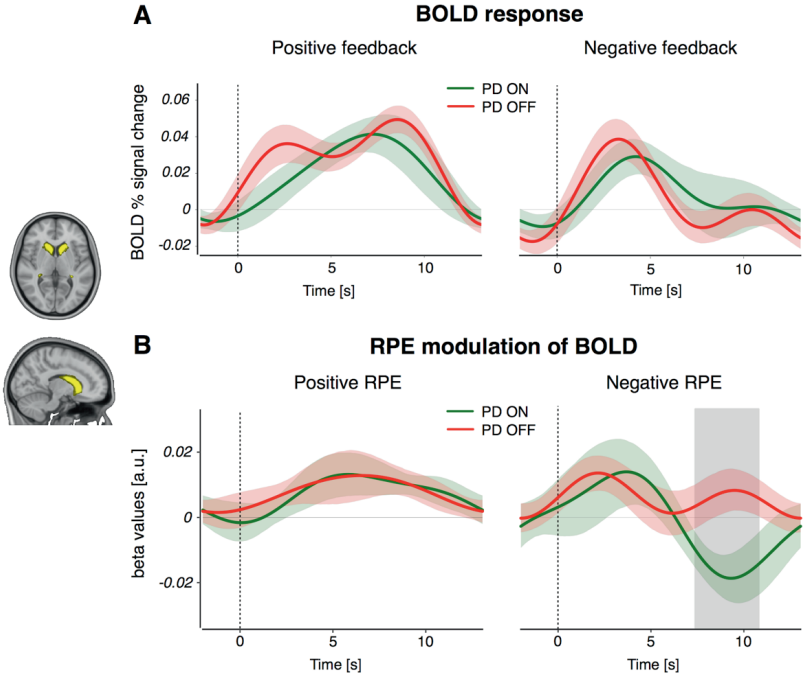


Figure 4. BOLD response and RPE modulation of the BOLD signal during feedback events. (A) BOLD percent signal change in response to positive (left panel) and negative (right panel) feedback events, in PD patients ON and OFF medication. There were no significant medication-driven differences for either event type. **(B)** BOLD RPE covariation time courses for positive (left panel) and negative (right panel) feedback events. We found a significant difference between PD OFF and ON in negative RPE responses, but not in positive RPE responses. The grey shaded area reflects a significant PD OFF > ON difference passing cluster-correction for multiple comparisons across timepoints (p<.05). Colored bands represent 68% confidence intervals (±1 SEM). A similar comparison between HC and each PD ON or OFF state showed no significant differences in the caudate nucleus (see Supplementary Fig. 7). The same analyses of putamen and nucleus accumbens ROIs revealed no medication-related RPE differences in these regions (see Supplementary Fig. 8).

Behavioural analysis of transfer phase

The previous sections reveal how medication remediates the way patients learn from negative outcomes by detailing medication-related changes in brain and behaviour. Much of the previous literature, however, has focused on how subsequent decision-making in the transfer phase is affected by dopaminergic medication (Frank, Seeberger and Reilly, 2004; Frank, 2007; Shiner *et al.*, 2012; Grogan *et al.*, 2017). We next set out to explore the relation between medication-induced changes in learning and subsequent behaviour. In the transfer phase of the experiment, participants were presented with novel pairings of the learning phase stimuli and were asked to choose the best option

based on their previous experience with the options (Fig. 1A). We examined accuracy in correctly choosing the stimulus associated with the highest value from the learning phase ("Approach A" trials) and correctly avoiding the stimulus associated with the lowest value ("Avoid B" trials) (Frank, Seeberger and Reilly, 2004; Jocham, Klein and Ullsperger, 2011), as in Fig. 1B **(**also see Materials and Methods). Replicating several previous reports (Frank, Seeberger and Reilly, 2004; Frank, 2007), results showed a strong interaction between medication (PD ON or OFF) and trial type (Approach A or Avoid B) (β (SE) = 0.34 (0.06), z = 5.75, p < .001). That is, medication in PD improved accuracy scores for Approach trials, but decreased accuracy for Avoid trials (Fig. 5A). Notably, there were no main effects of trial type, medication or disease status in addition to this pivotal approach/avoidance medication interaction. Thus, medication only influenced Approach A vs. Avoid B choice patterns, with no further differences in the overall accuracy across groups or trials. An independent analysis of Approach A and Avoid B trials separately revealed a main effect of medication on performance for both approach trials (a positive effect of medication on accuracy; β (SE) = 0.39 (0.08), z = 4.28, p < .001) and avoid trials (a negative effect of medication on accuracy; β (SE) = -0.35 (0.09), z = 4.03, p < .001). Finally, an evaluation of HC performance showed an interaction between disease status (HC vs. PD OFF) and Approach A/Avoid B trial type (β (SE) = 0.29 (0.06), z = 4.56, p < .001), with HC showing an approach/avoid asymmetry similar to PD ON (see Supplementary Fig. 10 for HC performance). There were no main effects of disease, i.e. there was no significant difference between HC and PD OFF for either trial type. Approach/avoidance asymmetries are therefore particularly evident when assessing within-patient effects of dopaminergic medication.

Medication shifts in learning rate for negative outcomes relate to behavioural and striatal changes during transfer

We have described how medication affects the updating of individual patients' beliefs after encounters with negative feedback, and replicate previous work by showing medication-induced changes in approach/avoidance choices during a follow-up transfer phase with no feedback. In this final section we explore how the shift in learning rates caused by medication during learning relates to the subsequent approach/avoidance interaction in 1) choice outcomes, and 2) the BOLD response of the dorsal striatum.

Consistent with the observation that medication only affects learning rates after negative outcomes, we found that only the medication-related shift in α_{loss} (and not α_{gain}) was predictive of the magnitude of change in approach/avoidance behaviour, as indicated by the lowest BIC in a formal

model comparison analysis (see Materials and Methods and Supplementary Table 4). In other words, the more α_{loss} was lowered by medication, the bigger the medication-induced interaction effect in future approach/avoidance choice patterns (β (SE) = 91.97 (41.26), t (22) = 2.23, p = .037, see Fig. 5B). Because the dorsal striatum was differentially responsive to RPE during learning, we additionally examined how learning rate shifts relate to the striatal BOLD response in approach/avoidance trials, while patients were ON or OFF medication. To this end, we masked the caudate and putamen using the whole-brain RPE z-statistics map shown in Fig. 3. From these masks BOLD responses were extracted for Approach A and Avoid B trials, for each of the PD ON and OFF sessions. Again, only the medication-induced shift in α_{loss} predicted the magnitude of change in the BOLD response of the caudate nucleus, but not the putamen, for approach/avoidance trials of OFF compared to ON (see Materials and Methods and Supplementary Table 4) (β (SE) = 1.54, (0.56), t (22) = 2.77, p=.012, see Fig. 5C). In summary, these findings show that within-subject medication-related shifts in learning from negative outcomes are predictive of subsequent approach/avoidance medication-related changes, both in terms of behavioural accuracy and BOLD signalling in the caudate nucleus.

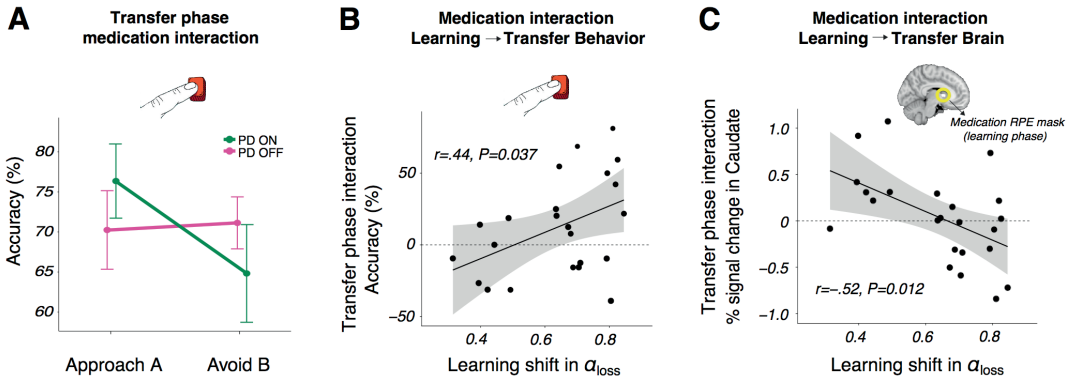


Figure 5. Medication-induced changes in learning from negative outcomes in PD predicts the magnitude of medication difference in subsequent approach/avoidance behavioural choices and striatal response. **(A)** Transfer phase behavioural accuracy in Approach A and Avoid B responses, showing a significant within-subject medication interaction in approach/avoidance behaviour (p<.001). PD ON had a higher accuracy in approach trials but a lower accuracy in avoid trials than PD OFF. HC performance is shown in Supplementary Fig. 10. **(B)** A positive relationship between the medication difference, i.e. the parameter shift for OFF > ON, in negative learning rate and the transfer phase medication accuracy difference (OFF > ON) in avoiding the lowest-value stimulus versus approaching the highest-valued stimulus, i.e. the interaction observed in **A**. **(C)** A negative

relationship between the medication difference (OFF > ON) in negative learning rate and the same transfer phase medication difference (OFF > ON) in avoid compared to approach trials, here in terms of BOLD percent signal change in the caudate nucleus.

DISCUSSION

Our findings provide a bridge between a previously disparate set of findings relating to reinforcement learning in PD. First, using a formalized learning theory, we show how dopaminergic medication remediates learning behaviour by reducing the patient's emphasis on negative outcomes. These behavioural adaptations were tied to BOLD changes in the dorsal striatum, with medication reducing the sensitivity to RPEs, specifically during the processing of negative outcomes. Second, we show a relationship between how the medication-induced change in learning and subsequent approach/avoidance choices that differ in PD when patients are ON or OFF medication. We found that the greater the degree of restoration by medication in the learning rate for negative outcomes, the greater the medication-related impact on both subsequent behaviour and associated BOLD responses of the dorsal striatum during value-based decision-making.

Our finding that medication reduces negative learning rate directly replicates studies showing a medication-driven impairment in behavioural responses relating to negative feedback, in a variety of probabilistic learning tasks (Frank, Seeberger and Reilly, 2004; Cools, Altamirano and D'Esposito, 2006a; Bódi *et al.*, 2009; Palminteri *et al.*, 2009). Furthermore, this finding corroborates a dopamine-driven reduction in model-based negative learning rate in PD patients (Voon *et al.*, 2010) and rats (Verharen *et al.*, 2018). The shift towards lower sensitivity to negative outcomes in PD ON reflects a partially restorative effect. While sensitivity to negative outcomes became more similar to that observed in healthy controls, decision-making volatility, i.e. the exploitation of higher-valued options, did not (see Supplementary Fig. 6). Although theory on dopaminergic signalling has suggested a dual influence of medication on learning from both positive and negative outcomes (Frank, 2005), conclusions in the literature have been mixed. While this dual effect has been shown in several studies (Bódi *et al.*, 2009; Palminteri *et al.*, 2009; Voon *et al.*, 2010; Maril *et al.*, 2013), much literature has indicated an effect of medication only on negative feedback learning (Frank, Seeberger and Reilly, 2004; Cools, Altamirano and D'Esposito, 2006a; Frank, 2007; Mathar *et al.*, 2017) or only on positive feedback learning (Rutledge *et al.*, 2009; Shiner *et al.*, 2012; Smittenaar *et al.*, 2012). The notion of a dual influence of medication on both positive and negative RPEs is therefore not always, and in fact frequently is not, seen in the literature.

The medication interaction in subsequent approach/avoidance behaviour we find in the transfer phase supports previous research on the transfer of learned value to new contexts (Frank, Seeberger and Reilly, 2004; Frank, 2007; Cox *et al.*, 2015). A similar interaction effect for HC compared to PD OFF suggests that medication may play a role in normalizing the balance in approach/avoidance behaviour towards healthy levels (see Supplementary Fig. 10). This reinforces the notion that dopaminergic medication shifts the balance in activation of the Go and NoGo pathways of the striatum (Frank, 2005). It has been an open question whether these Go and NoGo pathways are in competition with each other or function independently. A recent review suggests that the Go and NoGo pathways should not be viewed as separate, parallel systems (Calabresi *et al.*, 2014).The two pathways are instead described to be structurally and functionally intertwined, with "cross-talk" occurring between Go and NoGo neuronal subtypes. It is therefore possible that differences in the processing of negative feedback during learning not only affect the NoGo pathway, but also the Go pathway (in a push-pull manner). This account represents a potential means by which the dopamine-dependent alterations in learning from negative outcomes observed in the current study can lead to an integrated (interactive) effect on subsequent approach and avoidance behaviour and associated BOLD activation in the striatum.

We observed greater RPE modulation of BOLD signalling in PD OFF compared to ON, indicating a medication-related role in the modulation of caudate nucleus activity during learning. Striatal BOLD activations have previously been demonstrated to track RPE, with numerous studies implicating the caudate nucleus in RPE signalling during goal-directed behaviour (Davidson *et al.*, 2004; O'Doherty *et al.*, 2004; Delgado *et al.*, 2005; Haruno and Kawato, 2006). The whole-brain analysis used in the current study reveals greater within-subject RPE modulation in patients OFF compared to ON medication in the dorsal striatum, a region well established to suffer substantial depletion of dopamine availability in PD (Bernheimer *et al.*, 1973; Dauer and Przedborski, 2003). Patients in our study do not exhibit clear medication-related differences that signify an excessive level of dopamine in the ventral striatum, as postulated by the dopamine overdose hypothesis (Cools *et al.*, 2001; Cools, Altamirano and D'Esposito, 2006a) and presented in studies focusing on the nucleus accumbens (Cools, 2006; Schmidt *et al.*, 2014). In our data, there does appear to be a quantitative medication-induced increase in the modulation of nucleus accumbens activity by positive RPE, however, this effect is not significant (see Supplementary Fig. 8). One recent study describing the mechanisms underlying "optimism bias" (a higher rate of learning from positive than negative outcomes) revealed greater RPE signalling in the ventral striatum for individuals who had a higher optimism bias (Lefebvre *et al.*, 2017). Given that we find reduced sensitivity to negative

outcomes in PD ON than OFF, with no difference in learning from positive outcomes, we deem it likely that there is a relationship between optimism bias and (quantitative) medication-related differences in the ventral striatum in PD.

Activation of the dorsal striatum has been reported for instrumental but not Pavlovian learning, suggesting its role in establishing stimulus-response-outcome associations (O'Doherty *et al.*, 2004). A prominent theory of dopamine functioning, the actor-critic model, highlights distinct roles for reward prediction and action-planning in reinforcement learning (Houk, 1995; Suri and Schultz, 1999; Joel, Niv and Ruppin, 2002), with the ventral striatum (critic) implicated in the prediction of future rewards(Cardinal *et al.*, 2002), and the dorsal striatum (actor) proposed to maintain information about rewarding outcomes of current actions to help inform future actions (Packard and Knowlton, 2002; Atallah *et al.*, 2007). Connectivity between the midbrain SN and dorsal striatum has also been found to predict the impact of differing reinforcements on future behaviour (Kahnt *et al.*, 2009). Overall, the caudate nucleus has been put forward as a hub that integrates information from reward and cognitive cortical areas in the development of strategic action planning(Haber and Knutson, 2010). The dopamine-dependent differences in RPE modulation of BOLD activity in the caudate nucleus presented here therefore suggest that PD's dopamine-related effects are specific to the processing of feedback to guide future actions. The dopamine-related interaction in approach/avoidance behaviour found in the transfer phase, in which actions were guided by previously learned values, provides further support for this interpretation.

A separate evaluation of medication-related differences during the choice period revealed that modulation of BOLD activation by Q-values was higher in the putamen when patients were on compared to off medication (see Supplementary Fig. 9). Interestingly, the putamen has been demonstrated to track action-specific (Q-) value signals (Jahfari, Theeuwes and Knapen, 2019) and the covariation of this tracking was found to be higher in good compared to bad learners (Horga *et al.* 2015). Our behavioural analysis on choice accuracy during learning demonstrated greater overall learning in PD ON compared to OFF, which fits well with this PD ON > OFF group level difference of Q-value signalling in the putamen. Medication-related differences in the putamen for choice valuation during learning is thus an interesting avenue for future PD research.

We established a link between medication-dependent changes in learning from negative outcomes to subsequent changes in approach/avoidance striatal activity by specifically focusing on the region that showed a robust medication-dependent difference in phasic RPE modulation during learning. This suggests that the caudate nucleus' processing of negative RPE in PD ON plays an

important role in the subsequent medication-induced shift in balance between approach and avoidance behaviour. Although focusing on the ventral striatum, a recent study on rats showed that increased activation in the VTA-NAc (nucleus accumbens) pathway associated with a higher dopaminergic state was reflected in behaviour by a reduced sensitivity to negative outcomes (Verharen *et al.*, 2018). Our findings suggest that the caudate nucleus may play a similar role in the processing of negative outcomes in PD. Future research could address whether this is modulated by SN-caudate nucleus connectivity and/or the interplay between instrumental and Pavlovian learning.

In several previous studies, dopamine level was manipulated pharmacologically in healthy adults, via levodopa medication (Pessiglione *et al.*, 2006) or NoGo (D2) receptor antagonists(Jocham, Klein and Ullsperger, 2011; Van Der Schaaf *et al.*, 2014). Here, we examined separable disease-related and dopaminergic medication-related effects in PD. Patients in the current study used a combination of dopaminergic medications, including those acting on both Go and NoGo receptors (levodopa), inhibitors that slow the effect of levodopa to give a more stable release, and dopamine agonists, which have a particular affinity for NoGo receptors. Accordingly, a limitation of our study is that we cannot pin down the relationship between specific dopaminergic medications and changes in learning. Dissociation between the different types of dopaminergic medication could therefore be a potential avenue for future research.

Although there is moderate evidence for a higher sensitivity to negative feedback in PD OFF compared to HC, we find that the greatest disease-related difference lies in the explore/exploit parameter of the model (see Supplementary Fig. 5). Higher choice accuracy during easier decisions in HC is likely strongly influenced by greater exploitation of value differences between options; indeed, a positive correlation has recently been shown between choice accuracy and exploitation in a similar reinforcement learning task (Jahfari *et al.*, 2018). In the current study, this difference in exploitation was observed regardless of PD medication state (Supplementary Fig. 6), showing that dopamine medication in PD does not reinstate healthy exploitative behaviour. This selectivity of dopaminergic medication's effects on learning may indicate certain mechanisms underlying PD-related psychiatric disorders (Voon *et al.*, 2010). Recent evidence from a perceptual decision-making study in PD showed an impaired use of prior information in patients in making perceptual decisions (Perugini, Ditterich and Basso, 2016), a deficiency that also was not alleviated by dopaminergic medication (Perugini *et al.*, 2018). Thus, regardless of medication status, PD patients show impairment in the integration of memory with the current sensory input. Since the explore/exploit parameter of the task used in our experiments is dependent upon the retrieval of the expected value

of chosen options, a similar memory-guided decision-making impairment may have also played a role in the current reinforcement learning task.

We included several spouses of PD patients in our HC sample. Spouses of patients may be under more stress or anxiety than usual, which may impact how they learn from reinforcements. Since HCs as a group performed significantly better than PD patients during the learning phase and similar to HCs during the transfer phase in a similar previous study (Frank, Seeberger and Reilly, 2004), it seems likely that our control sample was sufficiently representative of healthy older adults to allow us to examine disease-related differences in learning.

Computational psychiatry is a burgeoning field of research with the aim of translating advances in computational methods to practical benefits for patient diagnosis and intervention (Huys, Maia and Frank, 2016; Maia and Conceição, 2017). The surge in the application of reinforcement learning models to patient data warrants extensive examination of the model fitting procedures, parameter recovery, and model identifiability, i.e. if parameters are highly correlated, then one parameter may falsely absorb an effect that is not actually true (see Maia and Conceição, 2017). With this in mind, we used a hierarchical Bayesian modelling approach where individual and group parameters are estimated simultaneously in a mutually constraining manner (Wetzels *et al.*, 2010; Steingroever, Wetzels and Wagenmakers, 2013; Wiecki, Sofer and Frank, 2013; Ahn, Haines and Zhang, 2017). The performance of this model was subsequently extensively evaluated with a focus on reliability. Overall, we show: 1) that our model's parameters are only weakly related (Supplementary Fig. 11), 2) accurate parameter recovery for each participant in our study, and 3) accurate data recovery (Supplementary Fig. 2), which shows that the model can well reproduce the observed data for both patients and healthy controls. Moreover, we note that the parameter estimates in this study are comparable to our other work using this task and 2-alpha Q-learning model (Jahfari and Theeuwes, 2017; Jahfari *et al.*, 2018; Van Slooten *et al.*, 2018; Van Slooten, Jahfari and Theeuwes, 2019).

In conclusion, we comprehensively illustrate how dopaminergic medication used in PD can help remediate sensitivity to negative outcomes, indicated by both changes in negative learning rate and the dorsal striatum's response to negative RPE. Furthermore, we show how, when using experience garnered during learning to guide subsequent value-based decisions, these effects shift the balance of approach/avoidance behaviour and associated striatal activation. Aside from explicating dopamine's role in reinforcement learning and value-based decision-making, our findings open new avenues of treatment in PD and its associated psychiatric symptoms.

Acknowledgements

We would like to thank Annemarie Vlaar, Henk Berendse, and Odile van den Heuvel from the neurology and psychiatric departments of the Zaans MC and VUmc for help with patient recruitment, Ton Schweigmann and Joost Kuijer for practical and technical assistance at the MRI scanner, and Charlotte Koning, Daan Beverdam, and Rosa Broeders for help with data collection. We also thank Helen Steingroever and Ruud Wetzels for valuable insights on Bayesian modelling, Gilles de Hollander for inspiring discussions about fMRI modelling and Eduard Ort for exciting conversations about fMRI analyses and helpful comments on a version of the manuscript. Finally, we would also like to thank the anonymous reviewers of the manuscript for their insightful comments.

Chapter 4

Supplementary Materials

(Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease)

SUPPLEMENTARY INFORMATION

Additional participant information

Prior to participation, all subjects were screened according to the following inclusion criteria: age range 40-75 years old, normal/corrected-to-normal vision, and a prior diagnosis of PD in patients. Exclusion criteria were: current psychotropic medication usage (other than dopamine-/Parkinson-related medication in patients), major somatic disorder or psychosis, dementia diagnosis, or a history of head injury, stroke or any other neurological diseases. Patients were furthermore not included if they took selective serotonin reuptake inhibitors (SSRIs), in order to primarily examine the effects of dopamine, as serotonin has also been implicated in learning mechanisms (Daw, Kakade and Dayan, 2002; den Ouden *et al.*, 2013).

All participants provided written informed consent in accordance with the Declaration of Helsinki. Payment for participation was a minimum of €100 (PD patients, three sessions) or €70 (HCs, two sessions) for participation. A reward bonus was additionally paid out based on performance during the reinforcement learning task (PD ON mean= €8.86 ± 0.99, PD OFF mean= €8.85 ± 1.00, HC mean= €9.34 ± 1.42, per learning run).

The study was set up to capture both within-subject effects of dopaminergic medication in Parkinson's disease as well as across-subject effects of disease (patients compared to healthy controls), across multiple behavioural and fMRI measurements. Since within-subject comparisons require a smaller sample size than across-subject, we aimed to collect sample sizes based on the lower-powered across-subjects (HC vs. PD) comparisons. We expected medium effect sizes of disease status, so to achieve 80% power, with a type 1 error of .05 (two-tailed), this required a sample size of 24 participants in each group.

Computational model estimation

The model was estimated using Markov Chain Monte Carlo (MCMC) inference. Models were implemented using the Stan programming language (Stan Development Team, 2015; Hoffman and Gelman, 2016). We ran three chains of 5000 samples each (discarding the first 2500 of each chain for burn-in), and ensured convergence using manual examination of the trace plots (hairy caterpillars, easily moving around the parameter space) and evaluation of $\hat{r}$ statistics, which were all <1.1 (Gelman and Rubin, 1992). We additionally generated a number of quantities of interest

from the model, including individual subjects' trial-by-trial RPE for inclusion in fMRI whole-brain and deconvolution analyses. Simulations displayed in Supplementary Fig. 2 show adequate data recovery, along with the fitted mean group-level posterior distributions. Further assessment of model suitability demonstrates that the model is identified, as indicated by only weakly correlated parameter estimates (Supplementary Fig. 11) and satisfactory parameter recovery of the modes of the fitted parameter distributions for each participant (Supplementary Fig. 12).

Behavioural analysis of AB trials

We carried out an additional analysis on AB trials to examine medication- or disease-related differences in AB choice behaviour associated with receiving either positive or negative feedback. Data were separated into trials in which positive feedback was received on the previous trial of that pair (for regression 1) and trials in which negative feedback was received on the previous trial (for regression 2). These were divided into separate subsets since including a regressor for each of them in the same GLM would lead to rank deficiency (i.e. one is (necessarily) a linear combination of the other). We carried out a mixed-effects logistic regression analysis on each subset. In regression 1, the dependent variable (*DV*) was a binary indicator for choosing A (*DV*=1) or choosing B (*DV*=0) on the current trial. The first independent variable (*IV1*) coded for whether the A (optimal) or B (suboptimal) stimulus was chosen on the previous trial (*IV1* = 1 or -1, respectively). In this way, a large positive beta estimate from the regression represented choosing A on the previous trial, receiving positive feedback, and choosing A again on the next trial. Medication and disease status were included as covariates, as in the other mixed-effects regressions, along with a varying subject intercept to account for individual differences. In regression 2 (trials following negative feedback), we used the same DV and covariates, but with *IV1* coding the opposite sign to that in regression 1, i.e., choosing optimal A (*IV1*= -1) and choosing suboptimal B (*IV1*=1) on the previous trial. This inversion of sign means that a large positive beta estimate from the regression captured choosing B on the previous trial, receiving negative feedback, and switching to the optimal A stimulus on the next trial. For trials following positive feedback (regression 1), we found a main effect of *IV1* (β (SE) = 0.84 (0.05), z = 17.33, p <.001), i.e. in general, participants stuck with the optimal A stimulus if rewarded for that choice on the previous trial. For trials following negative feedback, we found a large effect of medication (β (SE) = 0.52 (0.13), z = 3.96, p <.001), with a trend effect of disease (β (SE) = 0.64 (0.33), z = 1.95, p = .051) and *IV_1* (β (SE) = −0.17 (0.09), z = 1.92, p = .055). The main effect of medication shows that in trials following negative feedback, PD ON chose the optimal A stimulus more often than PD OFF, regardless of what they chose in the previous trial. This suggests

that, for the AB pair at least, PD ON were less inclined to use negative feedback to influence subsequent choice behaviour.

Session order effects

Experimental sessions were performed twice on PD patients and once on HCs. PD ON/OFF medication status was counterbalanced across first/second session in patients, so any session order effects on performance would be expected only in relation to HCs. We therefore carried out additional analyses on the behavioural data of the learning and transfer phases, to include session ("*Sess*") as a separate binary covariate in the mixed effects regression analyses reported in the Methods and materials section. We coded this as *Sess*=0 for HC since they only had one session, with *Sess*=0 for a patient's first session and *Sess*=1 for a patient's second session. The interaction between this variable and stimulus pair was also included, with the rest of the regression set up as in Equation 5 in manuscript.

For the learning phase, the effects already reported in the manuscript were also significant after carrying out this analysis; there was a main effect of stimulus pair (β (SE) = 0.37 (0.04), z = 8.75, $p << .001$), medication (β (SE) = 0.14 (0.04), z = 3.41, $p < .001$), an interaction between medication and stimulus pair (β (SE) = 0.21 (0.05), z = 4.27, $p <.001$), and an interaction between disease and stimulus pair (β (SE) = 0.18 (0.06), z = 3.11, p=.002) (see Supplementary Fig. 3A). This analysis revealed an additional interaction between session and stimulus pair (β (SE) = −0.11 (0.05), z = 2.30, p=.02), which suggests that when patients were in their second session, the difference in accuracy between the AB and EF pair was smaller than when participants were in their first session. There were no interactions between session and medication or disease.

For the transfer phase, the previously reported significant effects were again present; there was an interaction between medication (PD ON or OFF) and Approach A/Avoid B trial type (β (SE) = 0.35 (0.06), z = 5.78, $p < .001$) and an interaction between disease (HC or PD OFF) and Approach A/Avoid B trial type (β (SE) = 0.30 (0.07), z = 4.25, $p < .001$) (see Supplementary Fig. 3B). There was also a main effect of session (β (SE) = -0.13 (0.06), z = 2.15, p = .03), with patients in their second session performing worse in general than participants in their first session. There were no interactions between session and medication or disease.

fMRI preprocessing

The following information was generated from FMRIPREP based on the preprocessing pipeline used in this study. Each T1w (T1-weighted) volume was corrected for INU (intensity non-uniformity) using N4BiasFieldCorrection v2.1.0 (Tustison *et al.*, 2010) and skull-stripped using antsBrainExtraction.sh v2.1.0 (using the OASIS template). Brain surfaces were reconstructed using recon-all from FreeSurfer v6.0.1 (Dale, Fischl and Sereno, 1999), and the brain mask estimated previously was refined with a custom variation of the method to reconcile ANTs-derived and FreeSurfer-derived segmentations of the cortical gray-matter of Mindboggle (Klein *et al.*, 2017). Spatial normalization to the ICBM 152 Nonlinear Asymmetrical template version 2009c (Fonov *et al.*, 2009) was performed through nonlinear registration with the antsRegistration tool of ANTs v2.1.0 (Avants *et al.*, 2008), using brain-extracted versions of both T1w volume and template. Brain tissue segmentation of cerebrospinal fluid (CSF), white-matter (WM) and gray-matter (GM) was performed on the brain-extracted T1w using fast (Zhang, Brady and Smith, 2001) (FSL version 5.0.9). Functional data was motion corrected using *mcflirt* (FSL version 5.0.9) (Jenkinson *et al.*, 2002)."Fieldmap-less" distortion correction was performed by co-registering the functional image to the same-subject T1w image with intensity inverted (Huntenburg *et al.*, 2012; Wang *et al.*, 2017) constrained with an average fieldmap template (Treiber *et al.*, 2016), implemented with *antsRegistration* (ANTs). This was followed by co-registration to the corresponding T1w using boundary-based registration (Greve and Fischl, 2009) with 9 degrees of freedom, using *bbregister* (FreeSurfer version 6.0.1). Slice timing correction was not performed on the data. Motion correcting transformations, field distortion correcting warp, BOLD-to-T1w transformation and T1w-to-template (MNI) warp were concatenated and applied in a single step using *antsApplyTransforms* (ANTs version 2.1.0) using Lanczos interpolation. Physiological noise regressors were extracted applying CompCor (Behzadi *et al.*, 2007). Principal components were estimated for the two CompCor variants: temporal (tCompCor) and anatomical (aCompCor). A mask to exclude signal with cortical origin was obtained by eroding the brain mask, ensuring it only contained subcortical structures. Six tCompCor components were then calculated including only the top 5% variable voxels within that subcortical mask. For aCompCor, six components were calculated within the intersection of the subcortical mask and the union of CSF and WM masks calculated in T1w space, after their projection to the native space of each functional run. Frame-wise displacement (Power *et al.*, 2014) was calculated for each functional run using the implementation of Nipype. Many internal operations of FMRIPREP use Nilearn (Abraham *et al.*,

2014), principally within the BOLD-processing workflow. For more pipeline details, please refer to: https://fmriprep.readthedocs.io/en/latest/workflows.html.

Single-trial whole-brain analysis

We were interested in estimating trial-by-trial representations of learning (e.g. RPEs) in the brain. For this, we carried out a single-trial whole-brain analysis to capitalize on the variability in BOLD signal across trials, using Nipype's FSL interface (Gorgolewski *et al.*, 2011, 2017). A Least Squares All (LSA) GLM was fit to each subject's brain data, per learning run (see Mumford *et al.*, 2012). The feedback onset of each trial was included as a separate regressor, all in one model. We included 13 confound regressors to remove nuisance effects that might have contributed to the brain signal: Framewise Displacement (FD), 6 rigid-body transform motion parameters (3 translational, 3 rotational), and 6 aCompCor physiological noise regressors (to help exclude physiological noise in the CSF and WM). Spatial smoothing was performed using a Gaussian kernel with a full width at half maximum of 4 mm. A Savitsky-Golay filter was used for high-pass filtering, with a window length of 120 seconds and polynomial order of 3. The first-level design was set up using the Nipype interface to FMRI Expert Analysis Tool (FEAT) from FSL (FMRIB's Software Library, www.fmrib.ox.ac.uk/fsl). Delta functions of all regressors in the model were convolved with the canonical hemodynamic response function (HRF) and regressed against each subject's fMRI data, using Nipype's FSL FILMGLS interface. From this, a contrast of parameter estimates (COPE) was obtained for each trial of every subject. Next, we performed a second-stage analysis on the single-trial copes, to model per trial feedback valence and RPE. The feedback regressor was coded as +1 or -1 for positive and negative feedback trials respectively, to model brain activity that co-varied with valence. A single RPE covariate regressor was constructed by taking the per-trial RPE (retaining its positive or negative sign), de-meaning this distribution across trials, and convolving these events with the HRF. Since feedback valence and RPE are correlated, i.e. positive feedback is accompanied by a positive RPE, this allowed us to assign brain activity co-varying specifically with valence or RPE. This was run as a fixed effects multiple regression model using FLAMEO on a per subject basis. Fixed effects multiple regression models for collapsing across runs and deriving within-patient medication difference COPEs were carried out in a similar way. Medication difference COPEs of feedback and RPE were then brought to the group level in a random effects model, using FSL's FLAME 1+2 and outlier detection procedures. All group level Z (Gaussianized T) statistic images were thresholded using clusters of $z > 2.3$ and a cluster-corrected significance threshold of $p < 0.01$. Group-level analyses were carried out in this way for each regressor (feedback valence and RPE) on each separate group (HC, PD ON, PD OFF), on the within-subject medication differences ON > OFF and

OFF > ON, and on across-subject disease differences HC vs PD (OFF or ON). All group-level z-statistic contrasts can be viewed at: https://doi.org/10.6084/m9.figshare.6989024.v2. The MNI152_T1_1mm_brain standard brain was converted to functional space using FSL FLIRT, eroded by 1 voxel, and used as a brain mask for all of the analyses described above.

ROIs

We obtained high-resolution probabilistic atlas masks from a recent open-source dataset (Pauli, Nili and Tyszka, 2018). These sub-cortical ROIs have been well-established in playing an important role in reinforcement learning (Schultz, Dayan and Montague, 1997; Brown, Bullock and Grossberg, 1999; Hazy, Frank and O'Reilly, 2010; O'Doherty, Cockburn and Pauli, 2017). We focused on striatal ROIs (caudate nucleus, putamen, and nucleus accumbens) for the learning phase deconvolution analysis, as these have been most extensively studied in the past (Frank, 2005; Cools *et al.*, 2007; Cox *et al.*, 2015; Jahfari *et al.*, 2018). FSL FLIRT was used to register the masks to FMRIPREP output space. BOLD percent signal change during the transfer phase was extracted from ROIs informed by the learning phase of the task. We took the cluster-corrected RPE medication difference (OFF-ON) z-statistic COPE from the learning phase and multiplied it by independent striatal ROIs from the Pauli et al. (2018) dataset. These masks were thresholded to exclude the lowest 25% of voxels and then binarized. The masks are available on figshare, at: https://doi.org/10.6084/m9.figshare.6989024.v2.

Deconvolution analysis of feedback interval and RPEs

Deconvolution analyses were carried out on BOLD timeseries from striatal ROIs (see *ROIs* for mask information). We extracted BOLD time courses and teased apart the covariation with BOLD signal of positive and negative RPEs separately, using the fMRI timeseries data as preprocessed by FMRIPREP. A Savitsky-Golay filter was used for high-pass filtering, with a window length of 120 seconds and polynomial order of 3. These timeseries were then converted to percent signal change (PSC). PSC was calculated by dividing the timeseries by the mean of the entire timeseries, multiplying by 100, and then subtracting 100, to get a mean-centered output timeseries. Data from each subject were weighted per voxel according to the probability of belonging to a particular striatal ROI, and then averaged across voxels of that ROI. We set up a model with three regressors: stimulus onsets (with RT duration), positive feedback onsets and negative feedback onsets. Positive and negative RPEs were z-scored separately and included as covariates of their respective positive

or negative feedback event type. The deconvolution was implemented using the Python-based *nideconv* package (de Hollander and Knapen, 2017). Events and covariates were deconvolved with a Fourier basis set, which uses a combination of sines and cosines to model the data. This was implemented instead of the standard finite impulse response (FIR) function as it substantially reduces the number of regressors, thereby improving the robustness of parameter estimates. Five Fourier regressors (1 intercept, 2 sine waves and 2 cosine waves) were used for each of the positive and negative feedback events and positive and negative RPE covariates. We also included several confound regressors in the model: FD, 6 rigid-body transform motion parameters (3 translational, 3 rotational), WM, stdDVARS (standardized derivative of RMS variance over voxels), and 6 aCompCor physiological noise regressors. Time courses were then estimated simultaneously using a least-squares fit, for the time window -2 to 13.05 seconds (7 TRs) around feedback onsets. Up-sampling during the fitting procedure was implemented 20-fold as part of *nideconv* functionality. The resulting time courses were then brought to the group level, and within-patient medication differences were calculated per up-sampled timepoint of each fit. Clusters of significant intervals were identified using permutation-based one-sample t-tests (t-threshold set at $p<.05$, n=5000 permutations) as implemented in *mne-python* (version 0.15.2; (Gramfort *et al.*, 2013, 2014). Shaded regions in Fig. 4 and Supplementary Fig. 8 represent 68% confidence intervals (±1 SEM; bootstrapped using n=5000 permutations). Shaded regions in Supplementary Fig. 7 represent 95% confidence intervals (±1 SEM; bootstrapped using n=5000 permutations), to make group differences clearly visible.

Deconvolution analysis of choice interval and Q-values

We established a separate deconvolution analysis to assess modulations of the BOLD signal by Q-values during the choice period, i.e. while the stimulus options were presented. We set up this model with the same three regressors as in the previous analysis: stimulus onsets (with RT duration), positive feedback onsets and negative feedback onsets, all with a Fourier basis set. However, here the Q-values of the chosen stimulus were z-scored and entered as a covariate on the stimulus onset events. RPEs were not included in the model. All other aspects of the analysis were the same as those described previously.

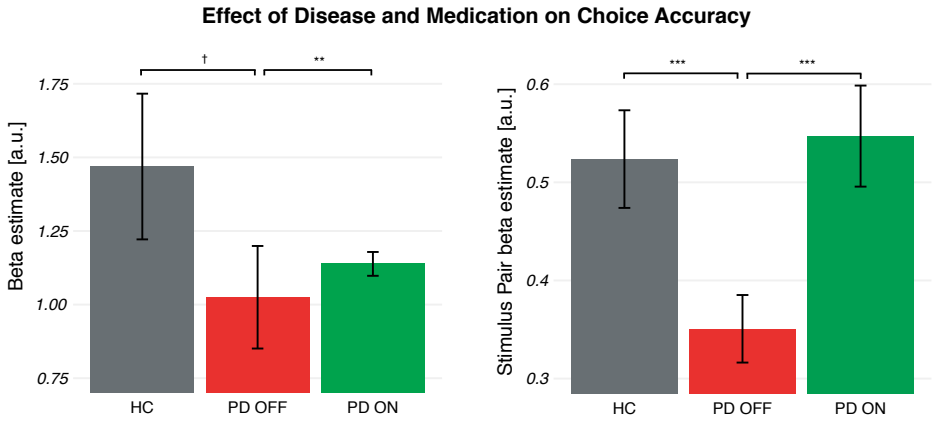
Applying the analysis to striatal ROIs, we did not find a significant medication-related difference in the caudate nucleus or nucleus accumbens. However, the analysis did reveal a

significant PD ON > OFF difference in modulation of the BOLD signal by Q-values in the putamen (see Supplementary Fig. 9).

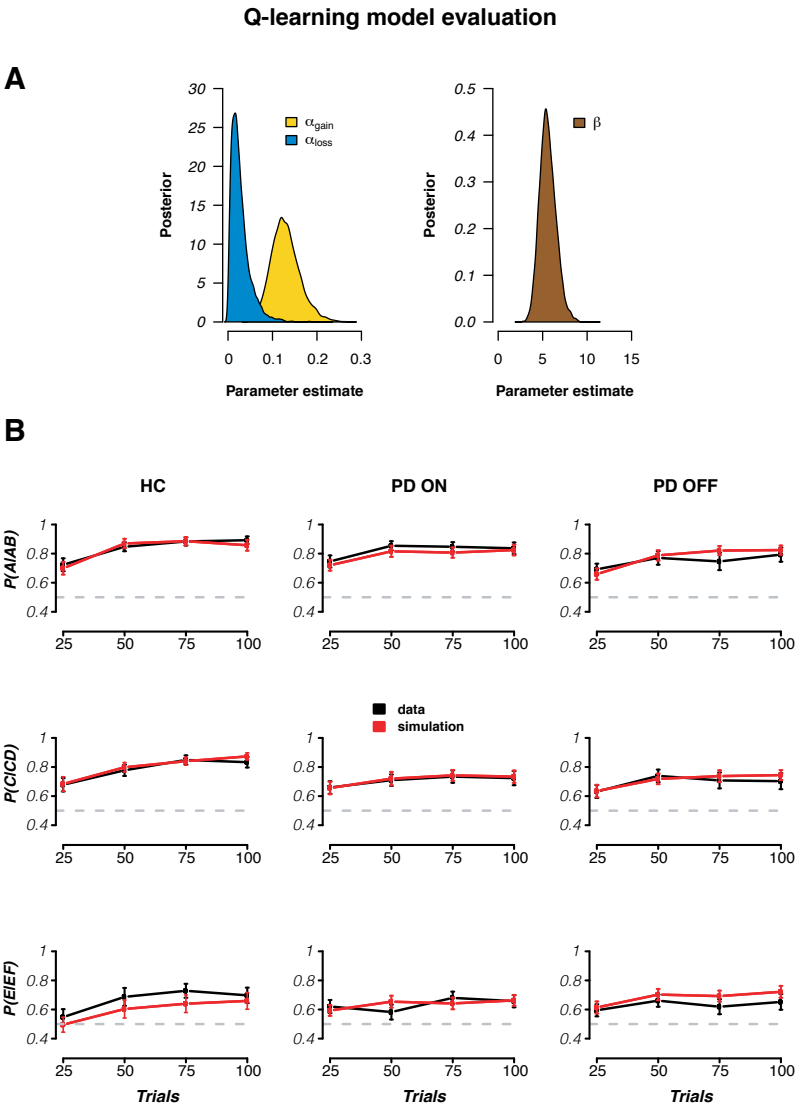
Transfer phase BOLD percent signal change

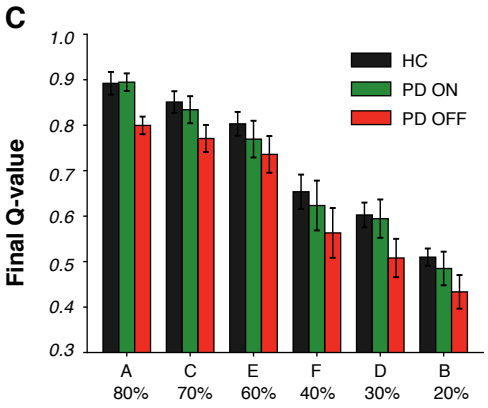
A standard GLM was set up to model BOLD responses to events in the transfer phase. Stimulus onsets and durations for three regressors were included: Approach A trials, Approach B trials, and all other trials (of no interest). Similar to the steps carried out in the learning phase whole-brain analysis, we performed 4mm smoothing, Savitsky-Golay high-pass filtering with a window length of 120 seconds and polynomial order of 3, and included the same 13 confound regressors in the design and convolved with a canonical HRF. Fixed effects analyses were performed across runs and for medication differences within patients. We then took the resulting two COPES for Approach A and Avoid B trials per subject and used FSL's *featquery* to calculate the mean percent signal change in the striatal ROIs that showed a significant learning phase medication difference in RPE (described in *ROIs*). In a similar way to the behavioural correlation analysis between learning rate and transfer accuracy (see Materials and Methods and Fig. 5B), we included learning rate EVs to explain the PSC (OFF>ON) medication difference in Avoid B>Approach A trials in striatal ROIs using robust (multiple) regression, with EVs as either: both positive and negative learning rate medication differences ($k\alpha_{gain}$ and $k\alpha_{loss}$), $k\alpha_{gain}$ only, or $k\alpha_{loss}$ only. Models were compared based on calculated BIC values (see Supplementary Table 4), and the learning to transfer PSC correlation *p*-value was obtained from the winning α_{loss}-only model. Individual medication differences were quantified as the modes of the within-subject medication difference parameter distributions.

SUPPLEMENTARY FIGURES

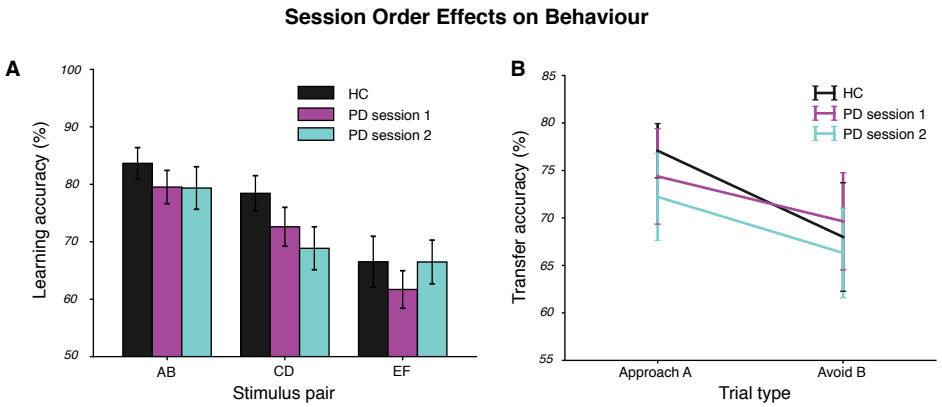


Supplementary Figure 1 | Beta parameter estimates in learning phase mixed-effects logistic regression model, with choice accuracy as the dependent variable. PD OFF was considered as 'baseline', with any relative increase in beta parameters for PD ON or HC representing the effect of medication and disease status, respectively. Here, the main effects of disease and medication on choice accuracy are presented (left), as well as interaction effects of stimulus pair and disease, and stimulus pair and medication, on choice accuracy (right).

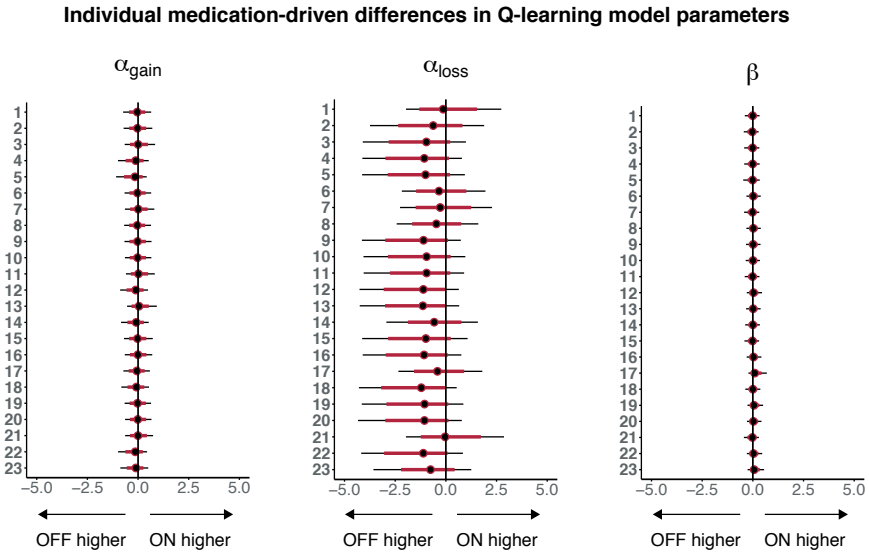

Q-learning model evaluation
A
Posterior
α_{gain}
α_{loss}
Parameter estimate
Posterior
β
Parameter estimate
B
HC
PD ON
PD OFF
P(A|AB)
P(C|CD)
P(E|EF)
data
simulation
Trials
Trials
Trials



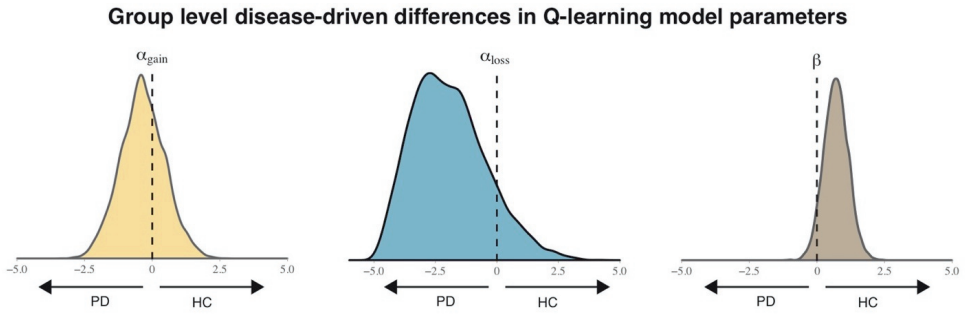
Supplementary Figure 2 | Bayesian hierarchical computational model assessment and simulation. A) Group-level parameter estimate distributions. Similar to other studies using separate learning rates for positive and negative events, we found higher learning rates for positive compared to negative feedback, termed "optimism bias" (Lefebvre *et al.*, 2017; Jahfari *et al.*, 2018; Van Slooten *et al.*, 2018). **B)** Simulation of the fitted model. To check whether the model sufficiently captured actual choice behaviour of participants, we simulated the probability of choosing the best option using the posterior distributions of the fitted free parameters of each participant. Plots of the modeled against empirical data across each group and stimulus pair show that the model is a good representation of overall learning. **C**) Final Q-values across each stimulus and group from the fitted model, showing that the model did a good job in capturing declining Q-value according to decreasing reward contingency.



Supplementary Figure 3 | Behavioural performance based on experimental session order in patients (not dopamine medication manipulation). See accompanying Supplementary Information: *Session order effects*. PD patients performed the task twice (ON or OFF medication, counterbalanced across first/second session). HC performed the task once; HC results here are the same as in manuscript Fig. 1 and Fig. 5A. **A)** Learning phase accuracy per group in choosing the better option of each stimulus pair. **B)** Transfer phase accuracy per group in correctly choosing the best A stimulus when it is presented and correctly avoiding the worst B stimulus when it is presented.

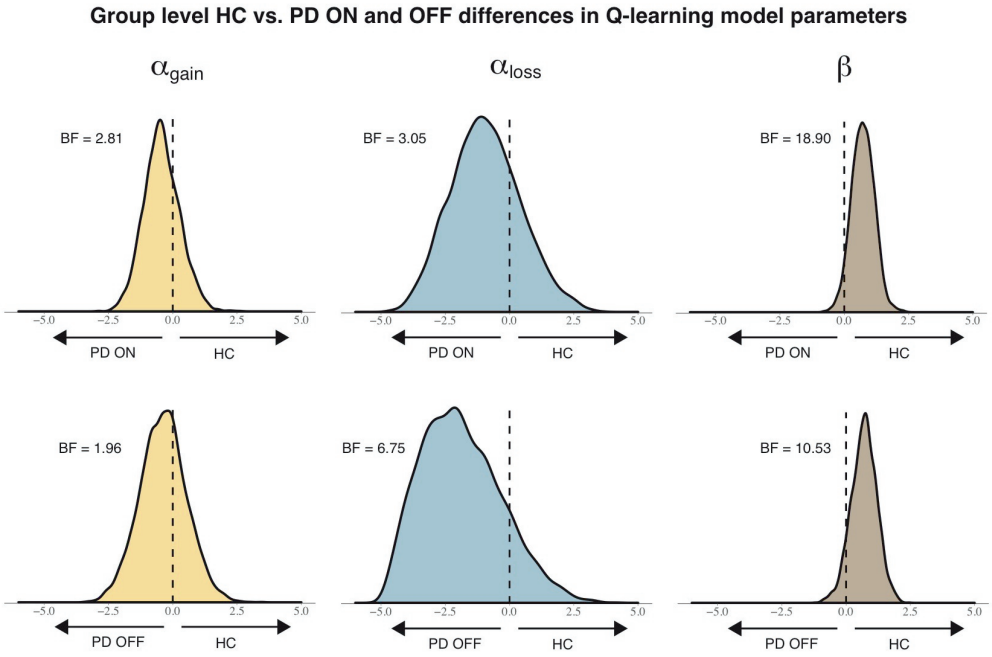


Supplementary Figure 4 | Individual medication differences in Parkinson's patients in computational model parameters. Medians are displayed, with red and black horizontal bars denoting 85% and 95% highest density intervals (HDI), respectively. Note that wherever individual medication difference parameters are used, we take the mode rather than median of the distributions, as this point has the highest posterior density of all parameter values.

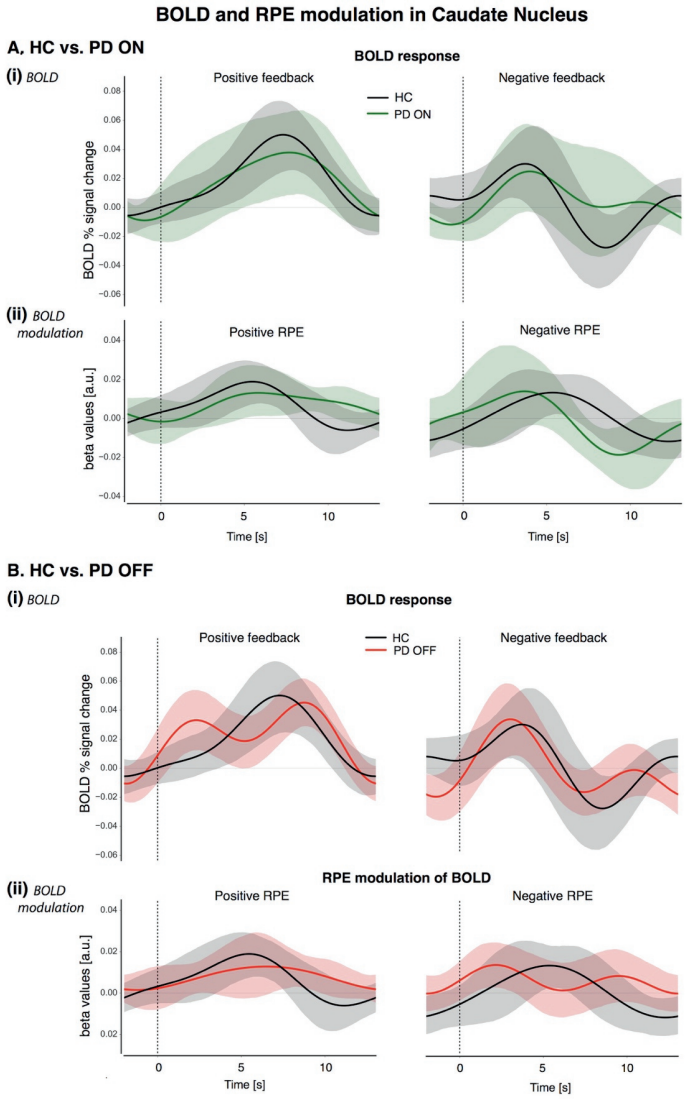


Parameter	Mean	SEM	95% HDI	BF_{10}
β	0.707	0.014	[-0.185, 1.565]	16.59
α_{gain}	-0.354	0.021	[-1.944, 1.298]	2.05
α_{loss}	-1.928	0.038	[-4.500, 1.065]	7.89

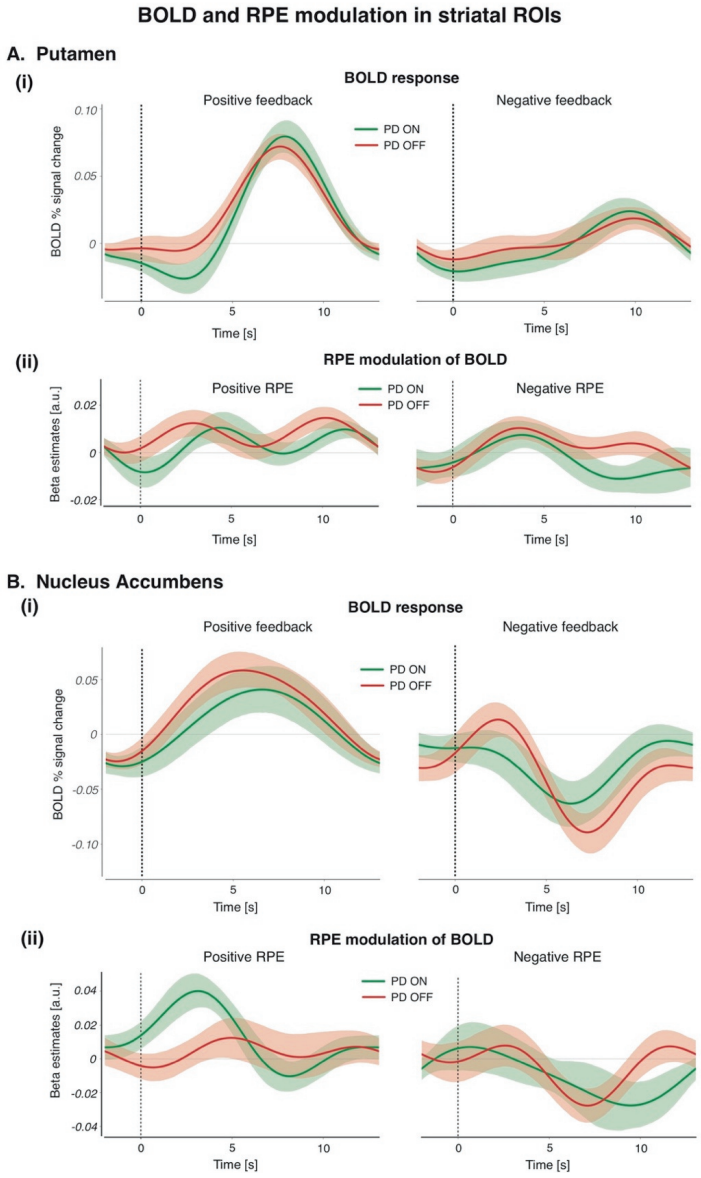
Supplementary Figure 5 | Disease-related differences in computational model parameters. Parameter distributions are shown for HC v PD (N=46). We found moderate evidence for greater α_{loss} negative learning in PD compared to HC (BF = 7.89), and strong evidence for greater exploitation in HC compared to PD patients (BF = 16.89).



Supplementary Figure 6 | Disease-related differences in computational model parameters, separately per PD medication status. To assess whether both PD ON and OFF contributed in a similar or differential way to the disease-related difference distributions shown in Supplementary Figure 4, two additional smaller models were run using a subset of data entered into the main model: HC vs. PD ON and HC vs. PD OFF. We ran 3000 samples each (discarding the first 1500 of each chain for burn-in). Checks for model convergence via visual inspection and $\hat{r}$ statistics was carried out in the same way as in the main model. The effect of disease on group-level parameters are displayed above for HC vs. PD ON (top row) and HC vs. PD OFF (bottom row). There is strong evidence for greater exploitation in HC compared to both PD ON and OFF, as indicated by a rightward shift in the β parameter distribution and BFs > 10.

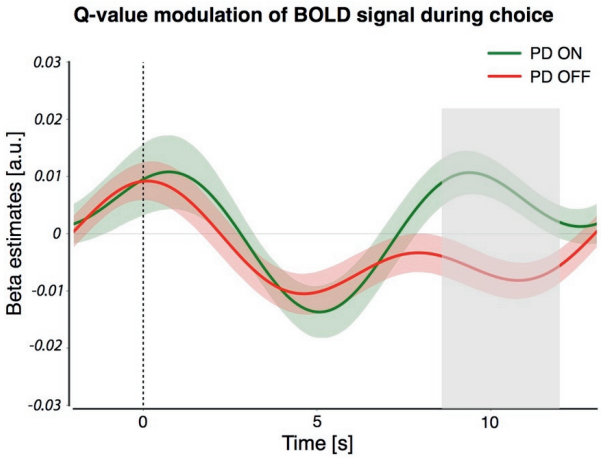


Supplementary Figure 7 | HC vs. PD ON or OFF comparisons of caudate nucleus feedback-related BOLD response and modulation of the BOLD signal by positive and negative RPEs. A) BOLD percent signal change (i) and RPE modulation of BOLD (ii) for HC and PD ON. **(B)** BOLD percent signal change (i) and RPE modulation of BOLD (ii) for HC and PD OFF. The shaded regions represent 95% confidence intervals, so visual inspection alone is enough to detect significant differences between groups, i.e. non-overlapping bands. The PD ON/OFF plots are the same as those shown in Fig. 4 of the manuscript (caudate nucleus ROI). Similar group comparison analyses of the putamen and nucleus accumbens ROIs (not shown) also show no group differences.

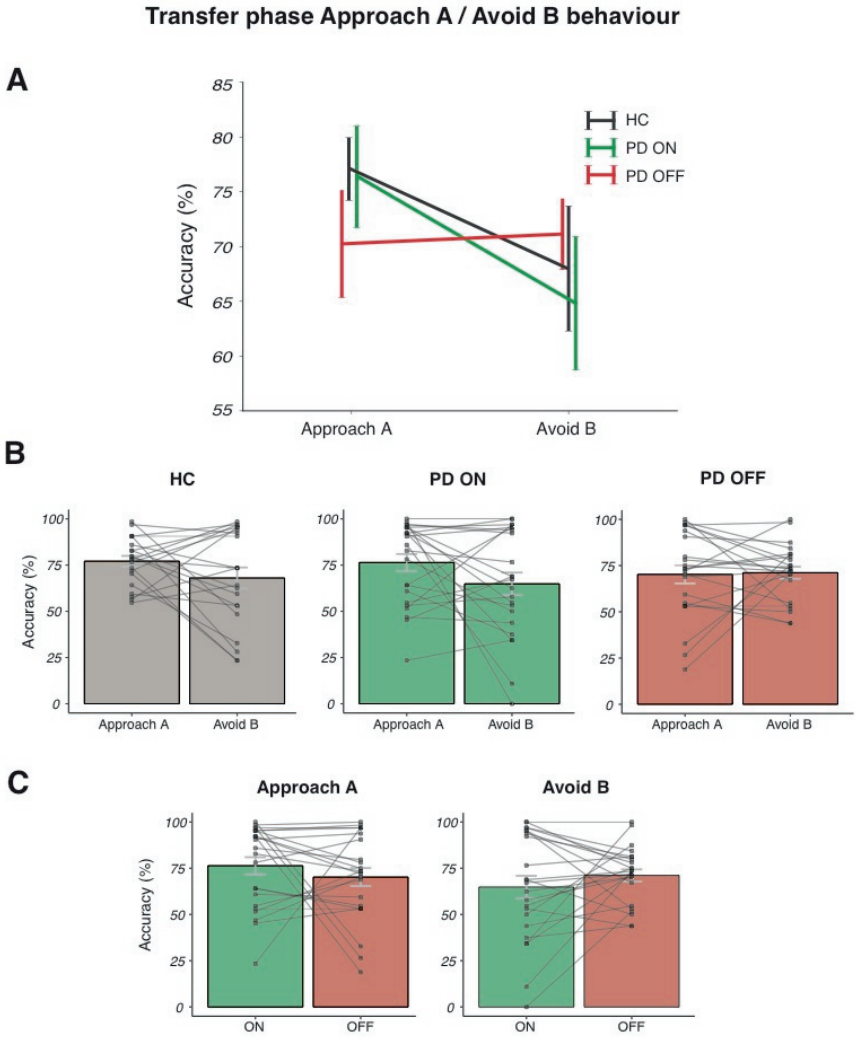


Supplementary Figure 8 | BOLD percent signal change and RPE modulation of BOLD for positive and outcomes in striatal ROIs of PD patients (related to Fig. 4). (A) BOLD percent signal change (i) and RPE modulation of BOLD (ii) in the putamen. **(B)** BOLD percent signal change (i) and RPE modulation of BOLD (ii) in the nucleus accumbens (NAc). Although no medication difference in RPE was found for ventral striatal NAc activity, it was included here for informational purposes since it has been implicated in several previous

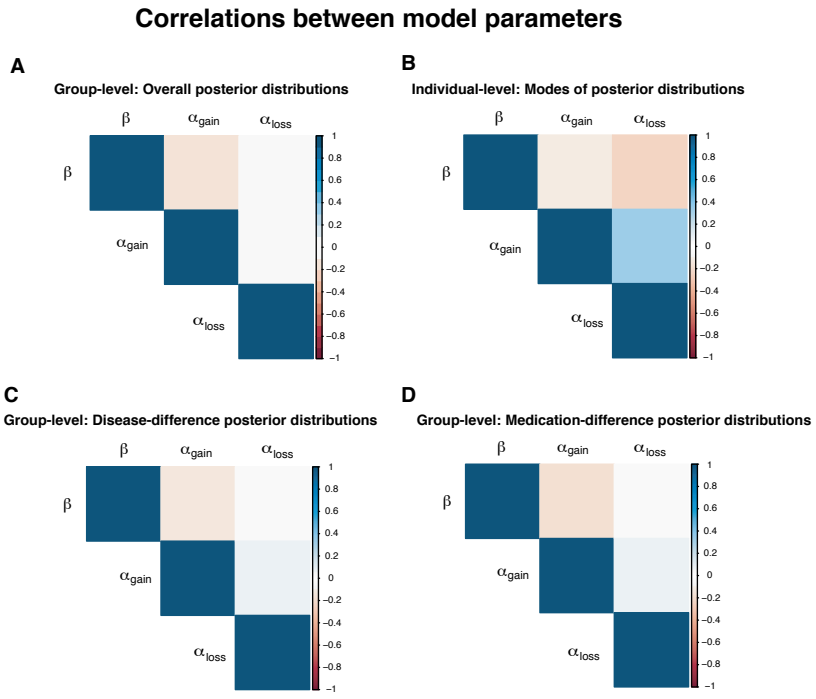
studies on the effects of dopamine on learning (Breiter *et al.*, 2001; McClure, Berns and Montague, 2003; J O'Doherty *et al.*, 2003; Cools *et al.*, 2007). In NAc, there appears to be a quantitative PD ON > OFF group level difference in positive RPE, however this difference is not statistically significant when within-subject differences were cluster-corrected across multiple timepoints. Positive and negative RPEs in each ROI were z-scored around that ROI's BOLD response for positive and negative events, respectively. Colored bands represent 68% confidence intervals (±1 SEM).



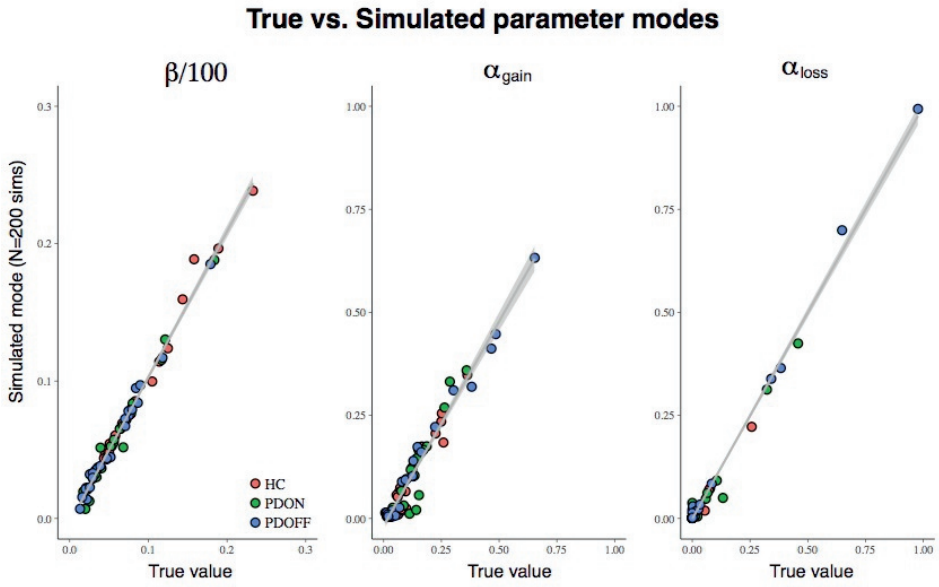
Supplementary Figure 9 | Modulation of BOLD signal in the putamen by Q-value of the chosen stimulus, time-locked to the onset of the stimulus options. Q-values were z-scored around stimulus onset events. A significant difference between PD ON and OFF is represented by the grey shaded region, which passed cluster-correction for multiple comparisons across timepoints ($p<.05$). Colored bands represent 68% confidence intervals (±1 SEM). See Supplementary section "*Deconvolution analysis of choice interval and Q-values"* for analysis description.



Supplementary Figure 10 | Accuracy in transfer phase approach and avoid behaviour for each group and subject. (A) Fig. 4 from manuscript with HC group included. **(B)** Averages per condition and within-subject plots, separately for HC, PD ON and PD OFF. **(B)** Within-subject plots across PD medication session, separately per Approach/Avoid condition.



Supplementary Figure 11 | Correlation between fitted model parameters. Correlations between full posterior distributions (group-level) or the modes of the posterior distributions (individual-level) of the three α_{gain}, α_{loss}, β model parameters, at each level of the hierarchical model. **(A)** Correlations at the overall group level, i.e. using the full posterior distributions of each parameter displayed in Supplementary Fig. 2A. **(B)** Individual- (session-) level correlations, using the modes of individual parameter distributions. **(C, D)** Parameter correlations between full posteriors of disease- and medication-difference distributions, respectively. These evaluations show that the parameters used in the model are only weakly related (Pearson $r < .4$) and indicates that these parameters can capture different aspects of the observed behaviour during learning.



Supplementary Figure 12 | Parameter recovery of fitted model parameters for each participant. The mode of each participant's posterior parameter distributions was taken and used to simulate 200 new datasets per participant. The actual parameter value (β, $\alpha_{gain,}$ and α_{loss}) used for 200 simulated sets on the x-axis ("True value") is plotted against the parameter modes of the 200 simulation fits on the y-axis ("Simulated mode") for each participant (denoted by a dot). Posterior modes obtained from the hierarchical Bayesian model taken for each participant (HC=red, PD ON=green, PD OFF=blue) represent the true value with which data was simulated per participant and used for recovery. As can be seen, the true and simulated modes were highly correlated across all three parameters (all p<.001), showing that this model is well able to recover the actual (original) parameters used for data simulations.

SUPPLEMENTARY TABLES

Characteristic	Healthy Controls (HC) N=23	Parkinson patients (PD) N=23	HC vs. PD (difference)
Age (years)	60.35 (8.72)	63.30 (8.24)	$t(44) = 1.182$, $p=.244$
Gender	8 females	6 females	$\chi^2(1) = .411$, $p = .522$
Education level (Verhage)	6.09 (0.85)	5.26 (1.14)	$t(44) = 2.793$, $p=.008$**
MoCA score	27.91 (1.88)	26.96 (1.92)	$t(44) = 1.708$, $p=.095$
BDI	4.09 (3.03)	6.07 (4.31)	$t(44) = 1.802$, $p=.078$
BAI	23.17 (2.27)	31.48 (6.15)	$t(44) = 6.077$, $p<.001$***
Digit Span Backwards	7.00 (2.09)	6.35 (2.39)	$t(44) = .986$, $p=.329$
NLV reading score	90.96 (8.47)	89.57 (7.36)	$t(43) = .588$, $p=.560$
Verbal fluency	43.91 (11.26)	37.91 (11.01)	$t(43) = 1.808$, $p=.078$
Disease duration (years)	-	4.00 (3.18)	-
UPDRS III on	-	17.87 (7.77)	-
UPDRS III off	-	20.65 (8.22)	-
Hoehn & Yahr	-	2.11 (0.58)	-
LEDD (mg)	-	790 (629)	-

Supplementary Table 1 | Demographic and clinical characteristics of participants. Executive functioning was assessed using the following tests: the Montreal Cognitive Assessment (MoCA), the Dutch version (NLV) of the National Adult Reading Test (NART) as a measure of pre-morbid IQ, the Stroop color-word task to assess effects of interference, verbal (category) fluency tests, and the rule-shift cards test of the Behavioural Assessment of the Dysexecutive Syndrome (BADS), to assess mental flexibility (Wilson *et al.*, 1997). The Complex Figure of Rey (CFR) was used as a measure of visuospatial memory. Verbal memory was assessed using the Dutch version of the Rey Auditory Verbal Learning Test (AVLT), testing both short- and long-term verbal memory (Saan and Deelman, 1986). Digit span forwards and backwards in short form (WAIS) was used to assess working memory. Participants also completed several self-report questionnaires: Beck Depression Inventory (BDI), Beck Anxiety Inventory (BAI), and Monetary Choice Questionnaire (MCQ). PD patients additionally completed the Wearing-Off Questionnaire (WOQ-Q10; related to the wearing off of DA medication), and the Questionnaire for Impulsive-Compulsive Disorders in Parkinson's Disease-Rating Scale (QUIP-RS). The motor part of the Unified Parkinson's Disease Rating Scale (UPDRS III) was carried out before each fMRI session. An overview of several test scores is provided in the table below. These assessments were not examined in the current study but are discussed in greater detail elsewhere (Engels *et al.*, 2018a; Engels *et al.*, 2018b). Quantities are presented as the mean across the sample, with brackets denoting 1 standard deviation.

Patient (Nr.)	Age (years)	Disease duration (years)	LEDD (mg)	Medication information			Time to scan since medication (hours)	
				Levodopa	DA-agonist	Other	OFF	ON
1	55	3.5	564	Yes			15.0	1.0
2	73	2.0	752	Yes			17.0	1.0
3	67	10.0	564	Yes			20.5	1.0
4	72	3.0	375	Yes			16.5	1.5
5	68	1.0	828	Yes			16.0	2.5
6	56	2.0	375	Yes			26.5	2.0
7	68	2.0	378	Yes			19.5	5.5
8	65	8.0	850	Yes		MAO-B inhibitor (Rasagiline) COMT inhibitor (Entacapone)	12.5	1.5
9	62	4.0	2780	Yes		COMT inhibitor (Entacapone)	14.5	3.0
10	64	5.5	982	Yes	Pramipexol		14.5	2.0
11	68	2.0	125	Yes			15.5	2.5
12	69	1.0	500	Yes			15.0	8.0
13	73	5.0	375	Yes			13.5	3.5
14	70	3.0	1548	Yes		COMT inhibitor (Entacapone)	15.5	1.5
15	71	6.0	1038	Yes	Pramipexol		8.5	5.5
16	47	6.0	1428	Yes	Ropinirol		14.0	1.0
17	48	0.5	1000	Yes			15.0	1.5
18	56	6.0	935	Yes	Pramipexol	MAO-B inhibitor (Rasagiline)	13.0	1.5
19	66	1.0	90	Yes			16.5	2.0
20	53	5.0	615	Yes	Ropinirol		19.5	1.5

21	72	13.0	1150	Yes	Pramipexol	MAO-B inhibitor (Rasagiline)	16.0	2.0
22	57	1.0	106	No	Pramipexol		18.5	4.5
23	51	6.0	1645	Yes	Ropinirol	Amantadine	14.0	2.5
24	61	1.5	108	No	Pramipexol		27.0	12.0

Supplementary Table 2 | Medication information for PD patients. Patient information regarding levodopa, dopamine agonists and any other dopaminergic medication. Levodopa Equivalent Daily Dosage (LEDD) was calculated according to Tomlinson *et al.*, 2010**.**

Model	Parameters	LOOIC (estimate± SE)
1	α, β	12432 ± 778
2	α_{gain}, α_{loss}, β	12167 ± 789
3	α_{gain}, α_{loss}, β, π	26325 ± 1476

Supplementary Table 3 | Model comparison. In order to compare the validity of the reported model with three free parameters α_{gain}, α_{loss}, and β, we evaluated two additional hierarchical Bayesian models and used the leave-one-out cross-validation information criterion (LOOIC) procedure for model comparisons. LOOIC is highly recommended for model comparisons of hierarchical Bayesian structures that use MCMC sampling (Vehtari *et al.,* 2017). All models were set up in the same fully hierarchical way as the model shown in the manuscript (Model 2), either in a reduced form (Model 1) to include only two free base parameters (α and β), i.e. with a learning rate that is updated on every trial regardless of a positive or negative outcome, or in extended form (Model 3) to include one additional perseverance ("stickiness") parameter (π), to account for any bias in choosing the same stimulus of a pair regardless of the reward outcome (Kable and Glimcher, 2007; Schönberg *et al.*, 2007). π was included in the softmax equation and was bounded as [-5, 5] in accordance with previous research using this parameter (Wunderlich, Smittenaar and Dolan, 2012). A lower LOOIC score indicates a better-fitting model. LOOIC estimates pointwise out-of-sample prediction accuracy from a fitted Bayesian model using the log-likelihood evaluated at the posterior simulations of the parameter values and may be extracted using the "loo" package in R (Gelman et al. 2013, Vehtari et al. 2017, Yao et al., 2017, Ahn et al. 2017). The log-likelihoods were calculated per subject in the Stan model using the log categorical probability mass function. This was updated on a trial-by-trial basis to reflect whether a trial was correct given the probability of choosing that option. These log-likelihoods were then extracted from the fitted model on a per subject basis and used to calculate the LOOIC. Model 2 with three free parameters α_{gain}, α_{loss}, and β (as presented in the manuscript) was found to be the best fitting model.

Explanatory variable	BIC (behaviour)	BIC (brain)
$k\alpha_{loss}$ only	231.38	33.36
$k\alpha_{gain}$ + $k\alpha_{loss}$	234.34	36.17
$k\alpha_{gain}$ only	234.48	37.92

Supplementary Table 4 | Summary of BIC values for the role of medication-related shifts in learning rate parameters in subsequent medication-related changes in transfer phase behaviour and BOLD activity. BIC values relating to transfer phase behaviour (second column) describe the explanatory power of within-patient medication-related shifts in learning rate parameters ($k\alpha_{gain}$, $k\alpha_{loss}$) in the transfer phase medication-related interaction in approach/avoidance behavioural accuracy. BIC values relating to brain activity (third column) describe the explanatory power of the same within-patient medication-related shifts in learning rate parameters in the transfer phase medication-related interaction in caudate nucleus BOLD activity during approach versus avoid trials. Overall, in both brain and behaviour, the medication-related shift in only the negative learning rate, α_{loss}, parameter best explained subsequent medication-related changes in approach/avoidance trials.

Parameter	Mean	SEM	95% HDI	BF_{10}
β	0.015	0.003	[-0.213, 0.268]	1.24
α_{gain}	-0.048	0.005	[-0.505, 0.407]	1.39
α_{loss}	-0.960	0.021	[-2.566, 0.338]	11.40

Supplementary Table 5 | Summary of learning phase PD medication differences in posterior distributions of the Bayesian model group parameters. Bayes factors (BF) for medication differences in learning parameter distributions represent the BF according to direction of the visible shift in the posterior difference, i.e., the α_{loss} parameter in Fig. 2B is shifted to the left (higher OFF medication), so the BF represents the probability of OFF > ON being greater than zero. HDI = highest density interval.

# voxels	log10(p)	MAX X (mm)	MAX Y (mm)	MAX Z (mm)	COPE-MAX X (mm)	COPE-MAX Y (mm)	COPE-MAX Z (mm)	Areas included
529	25.6	12.3	-82.8	47.4	35.2	-72.9	54	lateral occipital cortex, precuneous (R)
233	13.7	5.71	-89.3	-15.3	2.43	-79.5	-8.7	V1, occipital fusiform gyrus, occ. pole (R)
101	6.75	-33.7	-43.4	-45	-30.4	-43.4	-45	cerebellum (L)
79	5.35	-17.3	-13.9	11.1	-7.41	9.08	4.5	caudate nucleus, putamen, glob. pallidus, thalamus (L)
75	5.08	-43.5	-30.3	44.1	-27.1	-10.6	67.2	postcent. gyrus, supramarg. gyrus, sup. parietal lobe (L)
64	4.31	32	-69.7	-45	-4.13	-63.1	-48.3	cerebellum (R)
44	2.78	51.6	35.3	30.9	38.5	45.2	37.5	middle front. gyrus, frontal pole (R)
42	2.62	61.5	-43.4	-8.7	61.5	-46.7	-8.7	middle temp. gyrus, inf. temp. gyrus (R)
42	2.62	28.7	-7.32	50.7	32	-4.04	50.7	precent. gyrus, middle front. gyrus, sup. front. gyr (R)
40	2.45	2.43	25.5	54	5.71	18.9	47.4	sup. front. gyr. (R)
39	2.37	38.5	-4.04	44.1	38.5	2.52	37.5	precent. gyrus, middle front. gyrus. sup. front. gyr (R)

Supplementary Table 6 | Medication difference in whole brain RPE signal. Clusters of group-level PD OFF > ON medication difference (p <. 01, z = 2.3, cluster-corrected).

Chapter 5

Distractor inhibition and reinforcement learning in Parkinson's disease: behavioural and brain commonalities

Adapted from

McCoy, B. & Theeuwes, J.

Distractor inhibition and reinforcement learning in Parkinson's disease:

behavioural and brain commonalities

In preparation

ABSTRACT

Dopamine is known to be involved in several important cognitive processes, most notably in learning from rewards and in the ability to attend to task-relevant aspects of the environment. Both of these features of dopaminergic transmission have been studied separately in research involving Parkinson's disease (PD) patients, who exhibit diminished levels of dopamine. Here, we tie together some of the commonalities in the effects of dopamine on these aspects of cognition by having PD patients (ON and OFF dopaminergic medication) and healthy controls (HCs) perform two tasks that probe these processes. Within-patient behavioural measures of distractibility, from an attentional capture task, and learning performance, from a probabilistic classification reinforcement learning task, were included in one model to assess the role of distractibility during learning. Dopamine medication state and distractibility level were found to have an interactive effect on learning performance; less distractibility in PD ON was associated with higher accuracy during learning, and this was altered in PD OFF. Functional magnetic resonance imaging (fMRI) data acquired during the learning task furthermore allowed us to assess multivariate patterns of positive and negative outcomes in fronto-striatal and visual brain regions involved in both learning processes and the executive control of attention. Here, we demonstrate that while PD ON show a clearer distinction between outcomes than OFF in dorsolateral prefrontal cortex (DLPFC) and putamen, PD OFF show better distinction of activation patterns in visual regions that respond to the objects presented during the task. Finally, an analysis of grey matter volume (GMV) of brain anatomy revealed that GMV in the putamen in PD OFF is associated with both distractibility and learning from negative outcomes. Taken together, these findings highlight the impact of dopamine on the interplay between attentional processes and learning, and the common brain regions implicated in this relationship.

INTRODUCTION

Dopaminergic effects on cognition in Parkinson's disease (PD) have typically been investigated in two separate domains that probe different aspects of cognition: 1) changes in learning from feedback, e.g. in probabilistic classification (Frank, Seeberger and Reilly, 2004; Bódi *et al.*, 2009) or reversal learning tasks (Cools, Altamirano and D'Esposito, 2006a), and 2) changes in the ability to remain goal-focused and resist task-irrelevant information, e.g. in attentional switching (Cools *et al.*, 2001; Cools, Clark and Robbins, 2004; Bloemendaal *et al.*, 2015) or interference tasks (Wylie *et al.*, 2009). These cognitive domains are known to have shared neural circuitry; most notably the dorsolateral prefrontal cortex (DLPFC) which is central to both maintaining focused attention (Chao and Knight, 1995; Miller, Erickson and Desimone, 1996; Bloemendaal *et al.*, 2015) and supporting learning, by holding associative relationships among events and recent reinforcements in a working memory-like state (O'Reilly and Frank, 2006; Frank *et al.*, 2007; Blumenfeld *et al.*, 2010). Here, we wish to develop a better understanding of the commonalities across these two domains in terms of dopaminergic processing by administering both *attentional* and *learning* tasks to healthy controls and PD patients (on and off dopaminergic medication).

Dopaminergic nuclei in the midbrain fire differentially depending on the positive or negative valence of an outcome, with firing magnitude representing the size of a *prediction error*, i.e. the difference between the expected value of a chosen stimulus and the outcome received for that action (Schultz, 2007; Bromberg-Martin, Matsumoto and Hikosaka, 2010). These signals are transmitted to widespread brain regions via several dopaminergic pathways, including a nigrostriatal projection from the midbrain's substantia nigra pars compacta (SNc) to the striatum, involved in reward processing (Balleine, Delgado and Hikosaka, 2007), motivation (Robbins and Everitt, 1996), and movement (Draganski *et al.*, 2008; Helmich *et al.*, 2010), and a mesocortical projection from the midbrain's ventral tegmental area (VTA) to prefrontal cortical regions involved in high-level executive functioning. The SNc also projects to several parts of the frontal cortex (Ott and Nieder, 2019), particularly the DLPFC. DLPFC is strongly implicated in cognitive control, including the control of attention (Hornak *et al.*, 2004), inhibition of actions or thoughts (Pierrot-Deseilligny *et al.*, 1995; Anderson, Bunce and Barbas, 2016), rule learning (Cools, Clark and Robbins, 2004), and working memory processes (Curtis and D'Esposito, 2003).

The central role of DLPFC in cognitive control underlies our ability to stay focused and resist distracting stimuli or events (see Ott and Nieder, 2019 for a review). Dopaminergic modulation of the DLPFC is proposed to improve distractor resistance by stabilizing task-relevant,

working-memory representations in the DLPFC and making them less vulnerable to new inputs (Seamans and Yang, 2004; Durstewitz and Seamans, 2008; Bloemendaal *et al.*, 2015). Along similar lines, recruitment of DLPFC has been shown during the suppression of undesirable actions and events, such as the inhibition of responses in stop-signal and anti-saccade tasks (Brown, Bullock and Grossberg, 2004; van Belle *et al.*, 2014), and the suppression of unwanted memories in think/no think memory tasks (Anderson, Bunce and Barbas, 2016). DLPFC lesions in humans, as well as ablation of analogous PFC regions in rhesus monkeys, were found to be accompanied by reduced distractor resistance (Chao and Knight, 1995, 1998). DLPFC damage has furthermore been associated with problems in suppressing unwanted saccades (Pierrot-Deseilligny *et al.*, 1995)

Evidence from studies on visual selective attention points towards the role of top-down attention, exerted by PFC regions, in sharpening feature- or object-based representations in visual regions (Noudoost *et al.*, 2010; Noudoost and Moore, 2011). For example, impaired dopamine-related distractor resistance has been associated with reduced connectivity between DLPFC and the visual regions important for encoding task-relevant stimuli (Bloemendaal *et al.*, 2015). Bloemendaal and colleagues found that the dopamine D2-receptor agonist bromocriptine led to decreased face (compared to scene)-related activation, and this decrease was associated with greater distraction by face stimuli. In an earlier study, DLPFC activation, as well as the connectivity between DLPFC and fusiform face area, was shown to be perturbed after face compared to scene distractors (Yoon, Curtis and D'Esposito, 2006). This, the fronto-visual processing that is integral to maintaining visual attention, at the expense of distractibility, therefore presumably depends on dopaminergic mechanisms.

Many studies have shown that associating rewards with specific visual stimulus features can differentially boost the representation of rewarded compared to unrewarded visual stimuli (Serences, 2008; Serences and Saproo, 2010). The effect of positive or negative reinforcements on these visual representations is suggested to act in a similar manner, although the unique influences of attention and reward and their interactions on visual processing have proven difficult to tease apart (Maunsell, 2004). Since dopamine signalling is a primary a modulator during reinforcement learning, researchers have thus begun to investigate the influence of dopamine on visual processing. One such study found that dopamine decreases functional magnetic resonance imaging (fMRI) blood oxygen level dependency (BOLD) activity in visual cortex (Zaldivar *et al.*, 2014). Another study that examined the effect of dopamine on visual representations, i.e. neuronal patterns, of rewarded features, demonstrated the ability to classify a highly-rewarded cue using

neuronal activation patterns associated with a subsequent un-cued reward (Arsenault *et al.*, 2013). The authors also found that univariate BOLD activity selectively decreased within the representation of a high compared to low reward-predicting cue in visual cortex. This decrease in univariate activity suggests that dopamine-driven prediction error signals may tag representations of highly rewarded features in visual cortex. It therefore seems that, similar to the dopaminergic enhancement of visual attention, the representation of rewarded visual features is also enhanced by dopamine. In the current study, we test this explicitly.

The attentional capture (AC) task is a useful paradigm to measure the extent of distractor-resistance across individuals, i.e. "distractability" (Theeuwes, 1991; Theeuwes *et al.*, 1998). The PFC may be crucial in suppressing the response of posterior regions, such as visual cortex, to an irrelevant but salient stimulus, suggesting a means by which top-down mechanisms can modulate AC effects (Nobre *et al.*, 1997; de Fockert *et al.*, 2004; Talsma *et al.*, 2010; Bisley, 2011). A recent study in Parkinson's disease (PD) patients showed that when off dopaminergic medication, patients showed greater AC than healthy controls, indexed by an increased error rate for distractor compared to no distractor trials (Tommasi *et al.*, 2015). This suggests that distractor-resistance during the AC task is compromised by altered dopamine levels in PD.

PD patients suffer from a depletion of dopaminergic neurons, leading not only to the motor deficits characteristic of the disease, but also to changes in cognitive functioning, such as learning from feedback (Cools, Altamirano and D'Esposito, 2006b; Cools *et al.*, 2007; McCoy *et al.*, 2019). Replacement dopamine medication in patients has proven beneficial for certain cognitive processes, such as attentional switching (Cools *et al.*, 2001; Aarts *et al.*, 2014), but dopamine-related impairments have been observed in learning contexts (Cools, Altamirano and D'Esposito, 2006b). Much research suggests that these impairments are driven by reduced learning from negative outcomes when patients are on compared to off medication (Frank, Seeberger and Reilly, 2004; Cools, Altamirano and D'Esposito, 2006b; Bódi *et al.*, 2009; Mathar *et al.*, 2017; McCoy *et al.*, 2019), although other studies also report medication-related differences in learning from positive outcomes (Rutledge *et al.*, 2009; Shiner *et al.*, 2012; Smittenaar *et al.*, 2012). Overall, there is general consensus that increased dopamine by medication in PD leads to a positivity bias (Sharot and Garrett, 2016; Lefebvre *et al.*, 2017; McCoy *et al.*, 2019).

Here we consider how dopamine levels in PD influence the inhibition of task-irrelevant distractors by measuring the extent of distractor-resistance in a behavioural AC task (Theeuwes, 1992), as well as dopaminergic effects on brain representations of reinforcements during learning,

by assessing outcome-related activation patterns while participants performed a reinforcement learning task (Frank, Seeberger and Reilly, 2004) in the MRI scanner (see also McCoy *et al.*, 2019). Participants were 24 PD patients who completed the tasks both ON and OFF dopamine medication, and 24 healthy, age-matched controls. Our aim was to uncover commonalities, in terms of behaviour and brain mechanisms, in dopamine-related indices of distractibility and learning. Firstly, we assessed whether levels of distractibility play a role in reinforcement learning performance and if this is modulated by dopamine. Next, we used MVPA to address whether fronto-striatal and occipital regions known to be involved in both learning and distractor inhibition contain patterns that differentially code for positive and negative outcomes. Finally, using voxel-based morphometry (VBM), we assessed whether grey matter volume (GMV) differences in PD relate to measures of learning and distractibility.

MATERIALS AND METHODS

Subjects

The participant sample in this study was recruited as part of a larger project on reinforcement learning in PD, with original results published elsewhere (see McCoy *et al.*, 2019). 24 patients with Parkinson's disease were tested twice, once while ON their standard schedule of dopaminergic medication and once in a clinically-defined OFF state (>12 hours withdrawal). 24 age-matched healthy controls (HCs) were tested once. All patients were diagnosed by a neurologist as having idiopathic PD according to the UK Parkinson's Disease Society Brain Bank criteria. This study was approved by the Medical Ethical Review committee (METc) of the VU Medical Center, Amsterdam. All participants provided written informed consent in accordance with the Declaration of Helsinki. Data from one PD patient and one HC were excluded from analyses, due to falling asleep in the scanner and insufficient learning of the reinforcement learning task (<60% accuracy across all stimulus pairs), respectively. The anatomical scan of one HC was furthermore not collected due to requested early termination of scanning. Full participant details, inclusion and exclusion criteria, and medication information is included elsewhere (see McCoy *et al.*, 2019).

Tasks

Participants completed two experimental tasks: a behavioural visual attentional capture task (based on Theeuwes, 1992) outside the MRI scanner, and a probabilistic reinforcement learning task (based on Frank, Seeberger and Reilly, 2004; Frank, 2007) while undergoing scanning.

Attentional capture task

Participants performed this task on a laptop outside the scanning room. The standard stimulus display contained several green circles and one green diamond, placed on an imaginary circle equidistant from central fixation (see **Fig. 1a**). The total number of stimuli ("set size") was either five (small set size), seven (medium set size) or nine (large set size). Each shape contained a white line element that could be oriented either vertically or horizontally with equal probability across trials and conditions. Participants were instructed to respond as fast as possible to the orientation of the line inside the diamond, by pressing the "z" key when the line was vertical and the "/" key when the line was horizontal. A warning beep sounded to inform participants when they made an error. The session consisted of one practice block of 72 trials and 5 experimental blocks of 72 trials. Each set size contained 24 trials per block, totalling 120 trials per set size across all experimental blocks. Participants pressed the space bar when they were ready to begin a new block. In 50% of the trials, one of the circle stimuli was a red distractor. A trial started with the presentation of a central fixation cross for 1000ms, after which the stimulus display was shown. There was no time limit placed on responding to ensure that slowed motor-related responses of PD patients did not result in lost trials. Data were later analyzed with a threshold of 3000ms to ensure that responses due to inattention to the task were not included. This is a liberal threshold compared to other studies using this task, e.g. de Fockert *et al.*, 2004; Lavie and de Fockert, 2006, which used a 1600ms threshold. However, these studies were in young, healthy participants. We established this threshold to allow for age- and disease-related slowing.

Reinforcement learning task

In the learning phase, three different pairs of everyday object stimuli, e.g. shoes, balls, leaves (denoted as AB, CD and EF) were presented in random order. For each pair specific reward probabilities were associated with the individual stimuli, and participants had to learn to choose the best option of each pair based on the feedback provided (see **Fig. 1b**). This contingency was 80:20 for the AB pair, indicating that the A stimulus would be rewarded 80% of the time it was

chosen, whereas the B stimulus would be rewarded only 20% of the time. The reward probabilities were 70:30 and 60:40 for the CD and EF pairs, respectively. Participants were instructed to try to find the better option of a pair in order to maximize reward. Outcomes were either "Goed" or "Fout" text (meaning "correct" or "wrong" in Dutch), leading to a payout of 10 cents for correct trials and nothing for incorrect trials. Participants completed two learning runs, with 50 trials per stimulus pair per run. Object stimuli were obtained from an existing stimulus database (Konkle *et al.*, 2010).

Behavioural Analyses

Attentional capture (AC) task

Distractability RTs (RT distractor absent - distractor present conditions) and absolute RTs for correct trials from the AC task were analyzed using repeated-measures ANOVA in JASP (JASP Team 2019) and linear mixed-effects regressions with the *lme4* package in R (Bates *et al.*, 2014, R Development Core and Team 2017). Error rates were extremely low since participants could take as long as required to choose the correct response and were not analysed (see task description). Repeated-measures ANOVA were carried out separately per group comparison, either as within-subject comparisons for PD ON vs. OFF or between-group comparisons for PD ON/OFF vs. HC. The linear mixed-effects model incorporated both fixed and random trial-by-trial effects. RTs were log transformed to overcome positive skewing of the raw distribution and included as the dependent variable of the model. The fixed effects variables were distractor (absent/present), set size (small, medium, large) and their interaction. Two binary covariates were also included (as in *McCoy et al., 2019; Sharp et al., 2016*); the between-subject effect of disease (*Dis*, where PD = 0, HC = 1) and the within-subject effect of dopaminergic medication state (*Med*, where OFF = 0, ON = 1), along with their interactions with the fixed effects variables. Participant was entered as a random effect (Baayen, Davidson and Bates, 2008). Random slopes were included for distractor and set size to capture additional variability at the subject level (Barr *et al.*, 2013). The full model was therefore set up according to **Eq. 1**.

(Eq.1)

*log(RT) = Distractor + Setsize + Dis + Med + Distractor*Setsize + Distractor*Med + Distractor*Dis + Setsize*Med + Setsize*Dis + (1 | Participant) + (0 + Distractor | Participant) + (0 + Setsize | Participant)*

Reinforcement learning and distractibility

To quantify the distinct contribution of distractibility on learning, we carried out a mixed effects logistic regression on all trial-by-trial learning data. A similar analysis has previously been carried out on this data (McCoy *et al.*, 2019), however here we also include distractibility, along with stimulus pair (AB, CD, or EF), medication and disease status, to address our specific research questions (see **Eq. 2**). This type of analysis is robust for assessing within- and between-subject individual differences and has been previously been used to examine dopaminergic effects on learning (Doll *et al.*, 2016; Sharp *et al.*, 2016). The dependent variable encoded whether the better option of the stimulus pair was chosen on each trial (correct = 1, incorrect = 0). The within-subject (random-effect) explanatory variable was stimulus pair (AB = 1, CD = 0, EF = -1). We also included two binary covariates, each interacting with stimulus pair: the between-subject effect of disease (*Dis*: Parkinson's disease (PD) = 0, control = 1), and the within-subject effect of dopaminergic medication (*Med*: OFF = 0, ON =1, with HCs considered as being in an OFF state). Distractibility RTs from the AC task, i.e., the mean difference between distractor present vs. distractor absent trials, were normalized across all participants and included in the model as a fully interacting covariate. PD OFF therefore acted as the baseline group (*Med*=0, *Dis*=0), which allowed comparison to PD ON (where the Distractibility**Med* interaction reflects the interacting effect of medication and distractibility on learning) and to HC (where the Distractibility**Dis* term reflects the interacting effect of disease (HC or PD OFF) and distractibility on learning). In addition to this model, similar separate models were carried out per HC, PD ON, and PD OFF groups, without the *Med* and *Dis* terms. Statistical analyses were performed in R (R Development Core and Team, 2015), using the lme4 package (Bates *et al.*, 2014).

(Eq.2)

*Correct = Stim_pair + Distractibility_RT + Med + Dis + Stim_pair*Distractibility_RT + Stim_pair*Med + Stim_pair*Dis + Distractibility_RT*Med + Distractibility_RT*Dis + Distractibility_RT*Med*Stim_pair + (1 | Participant)*

fMRI Data Acquisition

BOLD fMRI data were acquired using a 3T GE Signa HDxT MRI scanner (General Electric, Milwaukee, WI, USA) with 8-channel head coil at the VU University Medical Center (Amsterdam, The Netherlands). Functional data for the reinforcement learning and localizer tasks were acquired using T2*-weighted echo-planar images with BOLD contrasts. Each brain volume contained 42 axial

slices, with 3.3 mm in-plane resolution, TR = 2,150 ms, TE = 35 ms, FA = 80 degrees, FOV = 240 mm, 64 x 64 matrix. The first two TR volumes were removed from each run to allow for T1 equilibration. Structural images were acquired with a 3D T1-weighted magnetization prepared rapid gradient echo (MPRAGE) sequence with the following acquisition parameters: 1 mm isotropic resolution, 176 slices, repetition time (TR) = 8.2 ms, echo time (TE) = 3.2 ms, flip angle (FA) = 12 degrees, inversion time (TI) = 450 lms, 256 x 256 matrix. The subject's head was stabilized using foam pads to reduce motion artifacts. Preprocessing of fMRI data was carried out using FMRIPREP version 1.0.0-rc2 (O. Esteban *et al.*, 2018), a Nipype-based tool (Gorgolewski *et al.*, 2011, 2017). Full details of the scanning protocol and preprocessing are included elsewhere (see McCoy *et al.*, 2019). Below, we include details relevant to the specific analyses that targeted the current research questions.

fMRI Localizer Analysis

Each fMRI scanning session began with a localizer run to extract a visual region of interest (ROI) for the MVPA analysis (see below). Stimuli in this run were objects of the same categories that were presented in the subsequent reinforcement learning task (six object categories per session), obtained from the same object database as the task stimuli (Konkle *et al.*, 2010). The actual stimuli used in the learning task were not included. Stimuli were centrally-presented within a black square frame kept constant in size across the run. Pink noise within these square patches were also included as stimuli. An object-selective cortex (OSC) ROI was taken as those voxels that responded more strongly to objects vs. noise, within anatomical masks of the superior and inferior lateral occipital cortex (LOC) (from the Harvard-Oxford Cortical Structural Atlas of the FSL package). The localizer run consisted of four similarly-structured blocks. Each block was made up of eight mini-blocks, containing either stimuli from each of the six object categories that were used in the subsequent reinforcement learning task, pink noise, or a *null* mini-block during which only the central fixation cross was presented. In all but the null mini-block, stimuli were flashed briefly for 500 ms, with 300 ms fixation-cross only intervals. There were ten distinct object stimuli per mini-block, presented twice, making up 20 stimulus presentations per mini-block. Each mini-block lasted 16 seconds. This localizer run lasted ~ 9 min in total. Participants had to push a button when two consecutive object images were identical, to ensure they were paying attention to the stimuli. Data from the localizer run were corrupted in three participants (2 PD, 1 HC) and were excluded when creating functional OSC masks. Preprocessing was carried out as described in McCoy *et al.*, 2019.

fMRI Single-trial Analysis

Single-trial whole-brain GLM analyses were performed on each participant's reinforcement learning fMRI data to obtain trial-by-trial brain activation patterns. The single-trial regressors were locked to the onset of positive/negative feedback presented at the end of each trial. The analyses were carried out using Nipype's FSL interface (Gorgolewski *et al.*, 2011, 2017). Full details are provided elsewhere (see Supplementary Materials in McCoy *et al.*, 2019). One notable difference between the analysis described here and that described in McCoy *et al.*, 2019 is that data were unsmoothed for the current MVPA procedure, as compared to spatial smoothing with a 4mm Gaussian kernel carried out in McCoy *et al.*, 2019. Unsmoothed data were used because MVPA analyses require as fine-grained information as possible to more accurately classify brain representations of interest (Poldrack, 2011). These GLM analyses resulted in one contrast of parameter estimates (COPE) for each trial of every subject and session.

fMRI Multivariate Pattern Analysis

A supervised machine learning technique using multivariate pattern analysis (MVPA) was carried out with the Python-based *scikit-learn* package (Abraham *et al.*, 2014). Classification of events was done on a per run basis, i.e. across each learning run separately, and classification results were averaged across runs per participant. The unsmoothed fMRI data from the single-trial analysis was used as input and each trial was labelled as either 'good' or 'bad' feedback. The MVPA analysis allowed us to identify patterns in the brain that can distinguish between positive and negative feedback. Activity patterns ("features") were standardized per run by removing the mean across that run and scaling to unit variance.

Since positive and negative feedback was provided probabilistically, depending on the presented stimulus pair and the stimulus chosen, it was not possible to balance the number of positive and negative feedback events, neither across run nor across participant. Overall, participants received more positive than negative feedback since their task was to learn which was the better stimulus of each pair. Such an imbalance in the number of positive vs. negative samples provided to the classifier can lead to a bias in the classification procedure and should be addressed. To account for this, we used a combination of oversampling and undersampling with the *SMOTETomek* function provided in the *imbalanced-learn* Python package (Batista, Bazzan and

Monard, 2003). Classification accuracy score was estimated per learning run and overall accuracy was the average across runs.

In addition to the OSC ROI described above (created using functional localiser data), we created anatomical masks for four other relevant frontal and striatal ROIs using the Harvard-Oxford Cortical Structural Atlas in FSL; DLPFC (middle frontal gyrus), caudate nucleus, putamen and nucleus accumbens.

To obtain group results per ROI, classification accuracies and number of samples per participant were entered into a univariate normal-binomial model with a variational Bayes implementation using the MICP package in the TAPAS Matlab toolkit (Brodersen *et al.*, 2012, 2013). This takes both fixed (within-subject) and random (between-subject) effects into account in one model. A hierarchical structure is employed by this model for mixed effects analysis, using a variational Bayes approach. Variational Bayes is more efficient (albeit, less exact) than more standard Markov chain Monte Carlo (MCMC) sampling procedures used for inference in Bayesian modelling.

This model was set up separately per HC, PD ON and PD OFF group. Statistical comparisons between groups were performed using either paired-samples t-tests (PD ON vs. OFF) or independent-samples t-tests (HC vs. PD ON/OFF) on the original subject-level classification accuracies extracted using *scikit-learn*. To compare classification accuracies between fronto-striatal ROIs and the visual OSC ROI in PD patients ON and OFF medication, i.e. the relative fronto-striatal vs. visual involvement in classifying outcomes, we first found the within-ROI difference between ON/OFF sessions and compared these differences across ROIs. Due to exclusions in creating the OSC masks (see *Participants*), MVPA results from two participants were here dropped from the non-OSC ROIs to make the appropriate statistical comparisons across ROIs.

Voxel-based Morphometry Analysis

Using SPM12 and VBM tools (Ashburner and Friston, 2005), anatomical images were first bias-corrected to obtain more uniform intensities across different tissue types. Light regularisation was applied and images were initially smoothed using a 60mm FWHM Gaussian kernel. The resulting images were then segmented into grey matter (GM), white matter (WM), and cerebral spinal fluid (CSF) in subjects' native space. Both bias-corrected and segmented gray matter images were aligned and warped to the ICBM space template of European brains, resampled down to 1.5 mm

isotropic voxels, and registered to a participant-specific template by the DARTEL Toolbox (Ashburner, 2007). GM images were normalized to MNI space using a Gaussian kernel with 8mm FWHM. This spatial smoothing kernel size is common in VBM studies (Nickl-Jockschat *et al.*, 2012) as it has been shown to improve VBM accuracy (Pepe et al., 2014) without increasing false-positive rates (Scarpazza *et al.*, 2015). We used a frontal-striatal ROI as described in O'Callaghan et al. (2013), and included the following regions from the Harvard-Oxford cortical and subcortical atlases: frontal pole, superior frontal gyrus, frontal medial cortex, subcallosal cortex, paracingulate gyrus, cingulate gyrus (anterior division), frontal orbital cortex, caudate, putamen, and nucleus accumbens. We additionally included the LOC, as this area was relevant to our research questions. We searched for brain regions within this ROI where GMVs correlated with i) learning rates (positive and negative separately) from the reinforcement learning task (estimated elsewhere using a reinforcement learning model, see McCoy *et al.*, 2019), for both PD ON and PD OFF groups, and ii) mean (across set size) attentional distractibility RT measures from the AC task for both PD ON and PD OFF groups. It should be noted that VBM anatomy was the same across the ON/OFF state (since T1 was collected only once, during the ON state), however the learning rates and distractibility measures differed depending on the medication ON/OFF state. For HC vs. PD group contrasts, images were family-wise error (FWE) corrected for multiple comparisons at a significance level of $p<.05$. For VBM correlations with learning rates and distractibility measures, images were tested for significance at $p<.01$ uncorrected, with a cluster threshold of 40 contiguous voxels (as per O'Callaghan *et al.*, 2013).

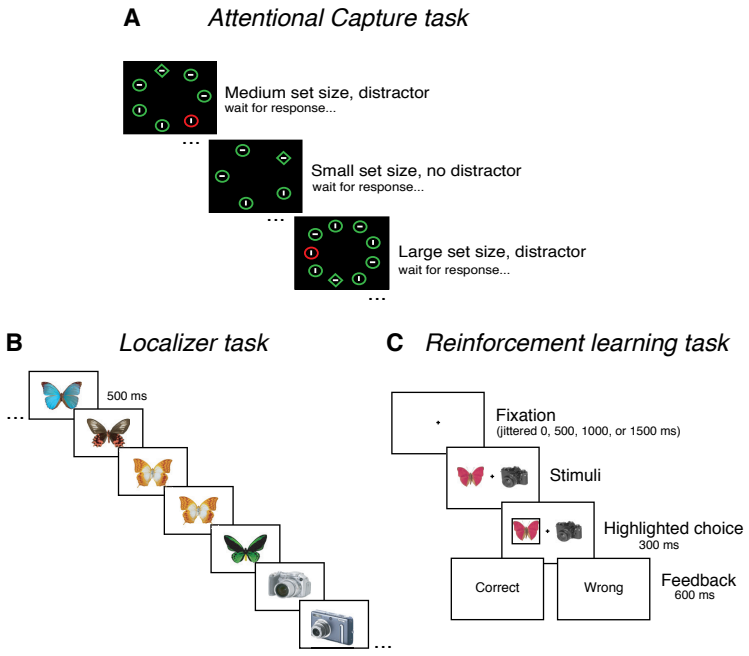


Figure 1. Experimental tasks. (A) A behavioural attentional capture task was performed outside the MRI scanner, to obtain a measure of distractibility, i.e. the amount of slowing down in no-distractor compared to distractor trials. **(B)** In the MRI scanner, participants first performed a localizer task, during which they simply pressed a button if two of exactly the same stimuli were presented sequentially. Stimuli consisted of many pictures from the same object categories as were used in the subsequent reinforcement learning task, as well as pink noise stimuli presented within the same central region as the object stimuli. A statistical contrast between object and noise trials provided voxels more involved in object-processing. Individual-level subject visual ROI masks were created using the overlap of these voxels with the lateral occipital complex (LOC), and was labelled object-selective cortex (OSC). **(C)** Reinforcement learning task, also performed in MRI scanner (picture adapted from McCoy et al. 2019). Participants learned to choose the better option of each of three fixed stimulus pairings. Outcomes were probabilistic with a higher chance of receiving reward for one stimulus compared to the other; reward contingency was 80:20 for the easiest "AB" pair, 70:30 for the "CD" pair, 60:40 for the hardest "EF" pair.

RESULTS

Attentional capture: distractibility by task-irrelevant stimuli

We first analysed RTs from the AC task, for correct trials only (see *Methods*). Absolute RTs per set size, group and distractor condition are provided in **Table 1**. Statistical comparisons were made on a derivative of these quantities; distractability RT, which was calculated as the within-participant difference (within the same ON/OFF medication session in PD patients) between the distractor and no distractor conditions, i.e. a higher value means a participant took longer to respond to the target when a distractor was present (see **Fig. 2A).** For the smallest set size, mean distractability RTs for the HC, PD ON, and PD OFF groups were: 6.10 ± 12.42 ms, 10.44 ± 16.19 ms, and 25.80 ± 13.76 ms, respectively. For the medium set size, these distractability RTs were: 36.38 ± 13.44 ms, 32.58 ± 16.42 ms, and 51.15 ± 21.37 ms. For the large set size, these measures were: 37.83 ± 13.12 ms, 7.50 ± 18.12 ms, and 66.96 ± 17.07 ms.

We carried out separate repeated-measures ANOVA on these distractibility RTs, per group comparison, e.g. PD OFF vs. ON, PD OFF vs. HC, and PD ON vs. HC. For the within-subject PD OFF vs. ON comparison, medication status, set size, and their interaction were included as independent variables. We found a main effect of medication on distractability ($F(1,22)=8.53$, $p=.008$). This suggests that PD OFF medication were more distractible than PD ON. There was no evidence for a main effect of set size, or an interaction effect (both $p>.1$). For the PD OFF/ON vs. HC comparisons, there was tentative evidence for a main effect of set size for PD OFF vs. HC only ($F(2,88) = 2.96$, $p=.057$). There was no effect of group and no set size * group interaction for either comparison (all $p>.09$).

Next, a linear mixed-effects regression analysis was carried out on correct trial-by-trial log-transformed RTs (to manage positive skewing of the raw RT distribution) across all participants, to address both fixed experimental effects and random within-subject effects within one model (see **Eq.1**). We found main effects of distractor ($\beta = .028$, SE = .009, $t = 3.26$, $p = 0.001$) and set size ($\beta = .004$, SE = .005, $t = 6.86$, $p < .001$). There was a group effect of disease, i.e., HC vs. PD OFF ($\beta = -.012$, SE = .006, $t = -2.21$, $p=0.032$). Finally, there was also a significant interaction between medication status and distractor condition ($\beta = -.022$, SE = .009, , $t = -2.48$, $p = 0.013$). This indicates that PD ON had less AC than OFF when distracting stimuli were present. We found no other significant effects.

Set size / Group	HC	PD ON	PD OFF
Small			
Distractor absent (mean ± std)	1005.5 ± 179.7 ms	1150.0 ± 254.8 ms	1166.4 ±294.8 ms
Distractor present (mean ± std)	1011.6 ± 165.7 ms	1160.4 ± 278.3 ms	1192.2 ±288.9 ms
Medium			
Distractor absent (mean ± std)	1033.9 ±159.0 ms	1182.8 ±272.7 ms	1186.8 ±299.5 ms
Distractor present (mean ± std)	1070.3 ±185.1 ms	1215.4 ±272.8 ms	1238.0 ±289.8 ms
Large			
Distractor absent (mean ± std)	1064.4 ±176.7 ms	1236.5 ±282.6 ms	1241.8 ±317.8 ms
Distractor present (mean ± std)	1102.2 ±185.8 ms	1244.0 ±263.8 ms	1308.8 ±318.6 ms

Table 1. RTs in the attentional capture task. Absolute RTs are shown for correct trials only. Mean and standard deviation (std) are provided across stimulus set size, group, and distractor conditions.

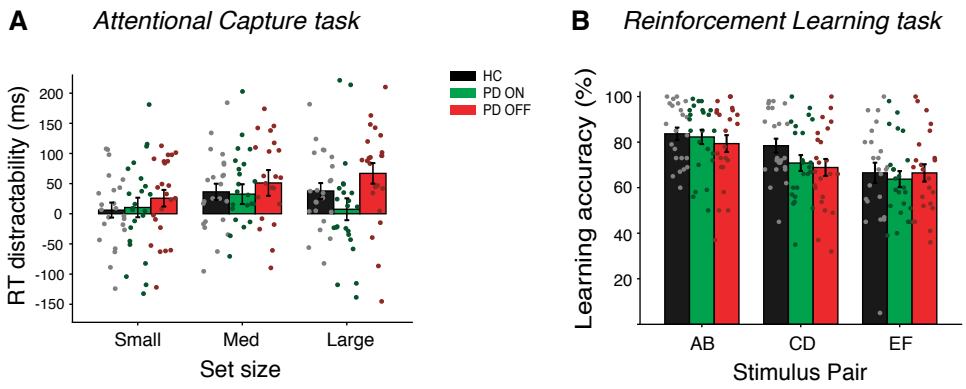


Figure 2. Behavioural results. A) *Attentional capture*: RT as a measure of attentional capture ("resistance" to distraction), per HC, PD ON and PD OFF group. RT distractibility is the amount of slowing caused when the distractor is present. **B)** *Reinforcement learning*: learning accuracy, i.e. the percentage of trials for each stimulus pair in which participants chose the 'better' option of the pair, per group.

Role of distractibility in reinforcement learning

Here, we attempt to describe how distractibility, or the inability to suppress attentional capture towards task-irrelevant aspects of the environment, affects reinforcement learning performance. We ran a linear mixed-effects regression on trial-by-trial data from the reinforcement learning task, with accuracy in choosing the better option of each pair as the dependent variable, and mean distractibility across all set sizes per participant included as an independent variable. See *Methods* for a full description of the model, which includes disease and medication status as binary covariates as previously described (see **Eq. 2** and McCoy *et al.*, 2019). Overall accuracies in choosing the best option per stimulus pair and group can be seen in **Fig. 2B**. As detailed elsewhere (see McCoy *et al.*, 2019), we found a main effect of stimulus pair (β = 0.217, SE = 0.078, z = 2.78, Pr(>|z|) = 0.005), and medication status (β = 0.489, SE =0.113, z = 4.31, Pr(>|z|)<.001), and both a stimulus pair * medication interaction (β =0.605, SE = 0.120 , z = 5.06 Pr(>|z|)<001), and a stimulus pair * disease interaction (β =.577, SE =.140, z=4.13, Pr(>|z|)<.001). Importantly for the current study, we found an interaction effect of medication status and mean distractibility RT on choice accuracy during learning (β = -0.606, SE = 0.187, z=-3.24, Pr(>|z|) = 0.001). The negative beta estimate here indicates that PD ON patients who showed less distractibility during the AC task were more accurate during the reinforcement learning task. Furthermore, results showed a 3-way interaction effect of medication status, mean distractibility RT and stimulus pair (β = -0.700, SE = 0.190, z=-3.68, Pr(>|z|) < .001), and a 3-way interaction effect of disease, mean distractibility RT and stimulus pair (β = -0.547, stderr=0.206, z=-2.66, Pr(>|z|) = 0.008). To follow up on these 3-way interactions, we carried out separate mixed-effects regression models per group to assess whether the distractibility RT*stimulus pair interaction was present for all groups. We found a significant distractibility RT*stimulus pair interaction for the HC group (β = -0.366, SE = 0.162, z = -2.26, p = 0.024) and PD ON group (β = -0.50299, SE = 0.14161, z = -3.55, p > .001) only. Analysis of the PD OFF data did not show any evidence for this interaction (p>.1). This indicates that the lower the degree of distractibility in HC and PD ON participants, the better they performed at the easiest relative to most difficult stimulus pair choice during learning. Lower dopamine levels in PD OFF appear to alter this relationship.

fMRI multivariate pattern analysis of positive vs. negative outcomes during learning

MVPA analyses of each participant's trial-by-trial BOLD data were carried out in predefined ROIs (see Methods). These ROIs were: visual OSC (defined using functional localizer data), and the

fronto-striatal regions DLPFC, caudate nucleus, putamen and nucleus accumbens (using anatomically-defined masks). In all ROIs and groups, classification accuracies were significant for positive vs. negative outcomes (all p<.001).

In OSC, classification accuracy in HC was significantly higher than in PD ON (t(40) = 2.14, p = .038; HC: μ = 72.31%, ci = [68.73%, 75.70%], p<.001; PD ON: μ = 68.46%, ci = [64.78%, 71.98%], p<<.001). The within-PD subject comparison showed greater accuracies in PD OFF than ON, which was not statistically significant (t(20) = 1.93, p=.068; PD OFF: μ = 70.78%, ci = [67.35 74.05], p<.001). There were no significant differences in HC vs. PD OFF (p>.1).

In DLPFC, we found significantly higher decoding accuracy in HC compared to PD OFF (t(42)=2.50, p=.017; HC: μ = 74.44%, ci = [71.15% , 77.53%], p <.001; PD OFF: μ = 70.20, ci = [66.61, 73.62], p<.001). There were no significant differences between HC and PD ON (p>.1; PD ON: μ = 71.92%, ci = [68.49%, 75.17], p<.001) or in PD patients ON vs. OFF medication (p>.1).

In the putamen, we also found significantly higher classification accuracies in HC compared to PD OFF (t(43)=2.94, p=.005; HC: μ = 73.32%, ci = [70.21%, 76.26%], p<<.001; PD OFF: μ = 69.29%, ci = [66.04%, 72.40%], p<.001). There were no differences between HC and PD ON (p>.1; PD ON: μ = 71.30%, ci = [68.01%, 74.44%], p<.001), nor between PD ON and OFF (p>.1). Statistical comparisons in the caudate nucleus showed no pair-wise significant differences between any of the HC (μ = 71.45%, ci = [67.98%, 74.75%], p<<.001), PD ON (μ = 68.89%, ci = [65.37%, 72.25%], p<.001), and PD OFF groups (μ = 68.72%, ci = [65.05%, 72.23%], p<.001). The same was true for decoding results in the nucleus accumbens (HC: μ = 62.23%, ci = [58.88%, 65.50%], p<.001; PD ON: μ = 60.93%, ci = [57.51%, 64.27%], p<.001; PD OFF: μ = 60.22%, ci = [56.77, 63.60], p<.001).

We next calculated within-PD patient differences in classification accuracy in each of these ROIs, and carried out paired t-tests between fronto-striatal ROIs and the visual OSC ROI. Previous research has indicated a dopaminergic role in top-down control of visual regions (Noudoost and Moore, 2011). Here, we sought to explore how relative information, i.e. patterns of activation, in these regions may differ according to dopamine levels (see **Fig. 3**). We found a significantly higher PD ON-OFF difference in classification accuracy both in DLPFC compared to OSC (t(20) = 2.99, p = .007; DLPFC: 1.72%± 1.77% SEM; OSC: -2.31%±1.71% SEM) and in the putamen compared to OSC (t(20) = 3.01, p = .006; Putamen: 1.76%±1.67 SEM). There were no significant dopamine-related differences between the caudate nucleus or nucleus accumbens and OSC.

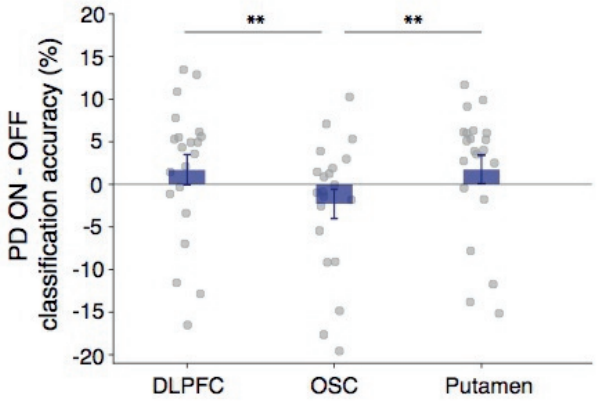


Figure 3. Within-PD patient ON vs. OFF differences in classification accuracies, in the DLPFC, OSC, and putamen. Error bars are ± 1 SEM. ** represents significant differences between regions at $p<.01$.

Voxel-based morphometry analysis of grey matter volume

GMV was estimated per participant using VBM on anatomical brain scans (see *Methods*). We first sought to replicate a previous finding showing greater atrophy in PD patients compared to HC participants in several fronto-striatal regions, including frontal orbital cortex, subcallosal cortex (extending to the nucleus accumbens) and inferior frontal gyri (O'Callaghan *et al.*, 2013). Using the same fronto-striatal mask as in that study (see *Methods* for included regions) we could not replicate GMV differences in these regions, nor were there any regions surviving threshold correction (FWE-corrected for multiple comparisons at $p<.05$).

Next, we examined how behavioural/computational measures from the two tasks correlate with GMV in PD patients, entering each measure separately as a covariate in the design matrix of the VBM. These measures were positive learning rate, negative learning rate and mean distractibility RT. Learning rates in reinforcement learning models play an important role in updating the value of stimuli by putting more or less emphasis on the most recent feedback received. Learning rates for the current patient sample were estimated elsewhere (see McCoy *et al.*, 2019). Although underlying brain anatomy is the same regardless of ON/OFF medication session, we estimated these correlations separately for each dopamine state as it possible that dopaminergic medication leads to changes in how different brain regions are recruited, thereby

relying on brain structure in an altered manner. In the following VBM analyses, only the positive correlations were extracted, e.g. higher GMV correlates with a higher value in the behavioural measure. We found a significant correlation between GMV and the positive learning rate parameter in multiple regions in PD ON, including the frontal pole and lateral occipital cortex (see **Supplementary Fig. 1** for all coordinates and statistics). For the same analysis in PD OFF, a relationship was found mainly in superior and inferior lateral occipital regions (see **Supplementary Fig. 2**). For GMV correlates of negative learning rate, inferior and superior lateral occipital regions were significant in PD ON (see **Supplementary Fig. 3**). In PD OFF, negative learning rate was found to covary with GMV in more widespread brain regions, including the left putamen and inferior lateral occipital complex (see **Fig. 4A** and **Supplementary Fig. 4**). For GMV correlations with distractibility RT, we found no associations in PD ON, and several clusters in PD OFF, including the left putamen and left nucleus accumbens (see **Fig. 4B**). In sum, across both learning and distractibility measures, we found that greater GMV in the left putamen is correlated with both greater learning from negative outcomes (higher negative learning rate) and increased distractibility in PD OFF (see **Fig. 4**).

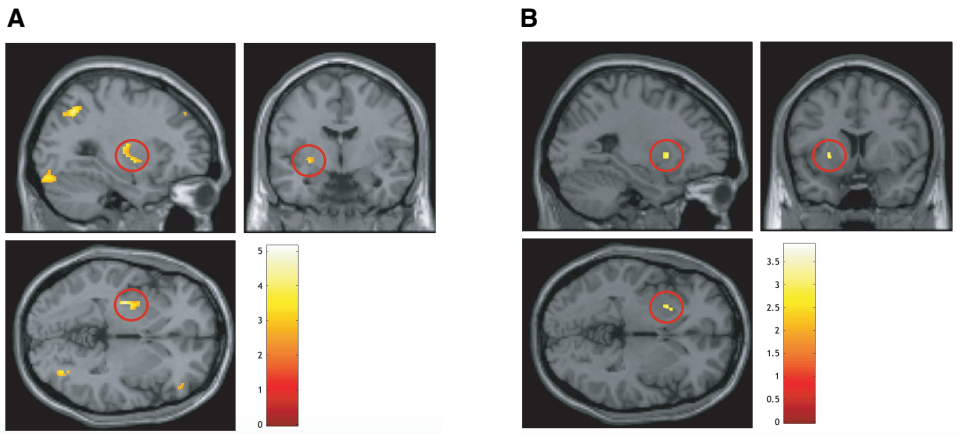


Figure 4. A) GMV that correlates with alpha loss learning rate parameter in PD OFF. **B)** GMV that correlates with attentional distractibility reaction time in PD OFF. The right putamen is significantly correlated for both measures. All statistics are based on a fronto-striatal and OSC mask, thesholded at p <.01 uncorrected, with a cluster size of greater than 40 contiguous voxels.

DISCUSSION

In this study we report on several behavioural and brain mechanisms claimed to be altered in PD, by means of the disease itself or by dopaminergic medication used to treat associated motor symptoms. PD patients and HCs performed an attentional capture task, to assess levels of distractibility by task-irrelevant items in the environment, and a reinforcement learning task, to describe how these groups learn from the outcomes of actions. Participants underwent fMRI while completing the reinforcement learning task, and the acquired BOLD data were used to assess patterns of activation associated with positive and negative outcomes in fronto-striatal and visual brain regions recruited during learning. VBM analyses on anatomical scans were carried out to highlight any differences in GMV among groups, and to evaluate relationships between GMV and behavioural/computational measures of distractibility and learning.

Our behavioural results show that PD ON demonstrate a significantly greater ability to ignore distracting stimuli than PD OFF, as indicated by reduced RT interference in the distractor vs. no distractor condition in the AC task. We furthermore found that medication status and mean distractibility RT interact to affect choice accuracy during learning, specifically; PD ON patients who showed less distractibility in the AC task performed better in the reinforcement learning task.

MVPA analysis of functional neuroimaging data from the reinforcement learning task show a medication-related effect in representations that distinguish positive from negative feedback across fronto-striatal and visual brain regions. Specifically, we found a medication-related interaction between DLPFC and OSC, with greater classification accuracy in PD ON (vs. OFF) in DLPFC, but greater accuracy in PD OFF (vs. ON) in OSC (see **Fig. 3**). A similar interaction was found between the putamen and OSC. This suggests that brain activations in these fronto-striatal regions are more separable for positive compared to negative outcomes when PD patients are ON medication, whereas visual brain regions show more separable responses for positive compared to negative outcomes when patients are OFF medication. Higher classification accuracies were furthermore found in HC participants compared to PD OFF in DLPFC and the putamen, and compared to PD ON in OSC. These results compliment the within-patient findings shown in **Fig. 3**. Finally, a VBM analysis of anatomical neuroimaging data show that positive and negative learning rate parameters and distractibility RT measures correlate with GMV differences in several fronto-striatal and occipital brain regions. An interesting result is that in PD OFF, greater GMV in the right putamen is correlated with both greater learning from negative outcomes (higher alpha loss) and greater distractibility (see **Fig. 4**).

Our finding that dopaminergic medication improves the ability to resist task-irrelevant (salient) stimuli in the environment dovetails with theory and results from previous research describing the role of dopamine in distractor inhibition (Sawaguchi and Goldman-Rakic, 1991; Seamans and Yang, 2004; Durstewitz and Seamans, 2008; Bloemendaal *et al.*, 2015; Tommasi *et al.*, 2015). These studies suggest that dopaminergic effects on working memory, via the excitatory D1 and inhibitory D2 pathways in the basal ganglia and PFC, relate to improvements and impairments in the ability to resist distracting input, respectively (Bloemendaal *et al.*, 2015). Prior research of ours shows that reward magnitude plays a role in distractor inhibition, with an increase in erroneous saccades to distractors that signal increasing levels of reward in the environment (McCoy and Theeuwes, 2016). This highlights the interplay between (irrelevant) reward signals and the ability to remain goal-directed. Dopamine in fronto-striatal regions is strongly implicated in rewarding and motivational responses (Robbins and Everitt, 1996). Although the distractor stimulus in the current AC task was not associated with reward, the dopamine-related improvements in PD patients show a shift in this balance of AC towards task-irrelevant stimuli vs. task-relevant goals, that is likely mediated by these fronto-striatal regions. The relationship between AC and learning from rewards or reinforcements is further suggested by our findings regarding the combined influence of distractibility (as indexed by overall RT differences between distractor conditions) and dopaminergic medication on performance accuracy during learning. Investigating how the balance between distractibility and goal-directness in PD patients ON vs. OFF medication may shift in accordance with varying distractor reward magnitude in the AC task is an interesting avenue for future research.

Findings from our MVPA analysis align with the suggestion of medication-related changes in the balance between distractibility and goal-directedness. We show that positive and negative outcomes can be more easily distinguished in PD ON than OFF in two fronto-striatal regions, the DLPFC and putamen, whereas these events are more easily decoded in PD OFF than ON in the lower, visual OSC. This provides some evidence for PD ON relying more on information from higher-level cognitive control and reward-related regions during learning, with PD OFF being more visually guided by lower-level, salience-related regions. This notion of competition between top-down vs. bottom-up processing has been long discussed in the literature (Egeth and Yantis, 1997; Itti and Koch, 2000; Corbetta and Shulman, 2002). OSC functional masks in this study were created on an individual basis, using a functional localizer to create an objects - noise contrast. Although this hones in on visual object-related regions, rather than feedback-related regions, positive and negative outcomes can still be easily classified. Several previous studies have shown reward-related

signals in visual cortex (Serences, 2008; Serences and Saproo, 2010). According to reinforcement learning theory, at the time of the outcome, positive or negative feedback is used to update the value of the chosen object (Rescorla and Wagner, 1972; Sutton and Barto, 1998). This presumably requires the integration of the object representation itself with the outcome and/or new object value. On the basis of our results we speculate that reward binding to objects may occur more at a stimulus-driven, visual level in PD OFF but more at a conceptual, reward-driven, and working-memory level in fronto-striatal regions in PD ON. It has been shown that disrupted dopamine-related distractor resistance is associated with reduced connectivity between DLPFC and the visual regions recruited for encoding task-relevant stimuli (Bloemendaal *et al.*, 2015). Interestingly, dopamine has previously been associated with a reduction of univariate activation in visual regions (Zaldivar *et al.*, 2014). Similarly, reward delivery without visual stimulation leads to decreased univariate activity within reward-associated cue representations in visual cortex (Arsenault *et al.*, 2013). Although it is not possible to make direct comparisons between these studies and our MVPA results, the current findings extend the notion that dopamine and reward lead to differential activations in visual regions. It is possible that higher dopamine levels in PD ON reduce univariate activity within the representation of positive outcomes compared to PD OFF, thereby interfering with multivariate patterns that may more easily distinguish between positive and negative outcomes.

VBM analyses of GMV correlations with learning rates and distractibility measures revealed associations with numerous fronto-striatal and occipital regions in PD patients. Of particular interest was the relationship between GMV changes in the left putamen and both negative learning rate and level of distractibility in PD OFF. This indicates that more GMV in this region correlates with both greater sensitivity to negative outcomes, i.e. quicker down-weighting of the value of stimuli after receiving negative feedback, and greater interference from distracting or task-irrelevant stimuli in the environment. Since these relationships are not present in the analysis of PD ON (for the same underlying anatomy), this suggests that dopamine may work in a way so as to disrupt or compensate for this relationship. The relationship between GMV of the putamen and distractibility, i.e. attentional capture, in PD OFF may be explained by the putative pre-motor association of this region, which is known to be connected to motor cortex and heavily implicated in motor control (Draganski *et al.*, 2008; Helmich *et al.*, 2010) and habitual action selection (Wunderlich, Smittenaar and Dolan, 2012).

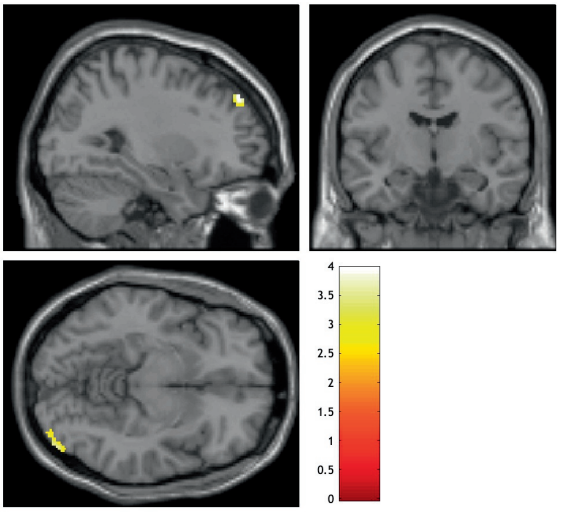
As with several analysis techniques in neuroscience, VBM is limited by power constraints; it is difficult to capture anatomical GMV differences with the sample sizes that are common in most fMRI studies. As such, the lack of group differences in GMV found here is not definitive. A further limitation of the current study related to the flexibility of the MVPA analyses. Due to the simultaneous presentation of two stimuli on each trial, as well as extremely large differences in choosing the better vs. worse option of pairs, e.g. some participants may only have chosen the worst "B" option several times across the whole experiment, it was not possible to appropriately tease out individual stimulus representations for pattern classification. An examination of dopamine-related differences in DLPFC representations of individual objects and how these might relate to distractibility, e.g. by presenting the objects as distractor stimuli in a separate task, is a compelling avenue for future research.

In summary, the current study examines the role of dopamine in the interplay between attention and learning. We provide evidence for dopamine-related effects of attentional distractibility on reinforcement learning, as well as a dopamine-related dissociation of multivariate representations in executive control, reward-associated and stimulus-driven brain regions. Relationships between both distractibility levels and learning rates with differences in GMV in several fronto-striatal and visual regions further suggest that commonalities between attention and learning mechanisms have both a functional and anatomical basis in the brain.

Chapter 5

Supplementary Materials

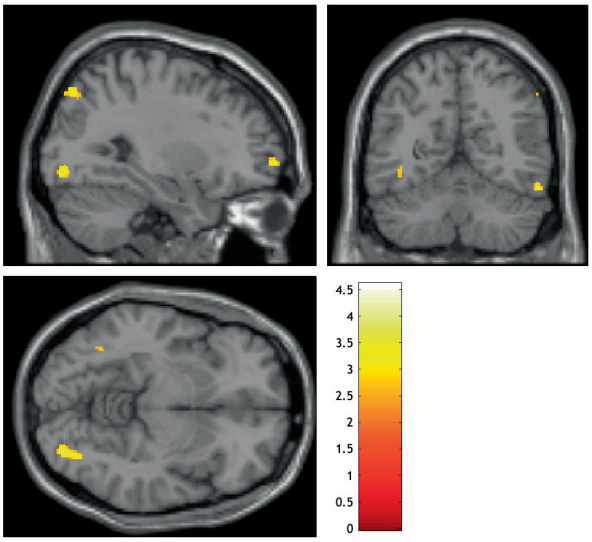
(Distractor inhibition and reinforcement learning in Parkinson's disease: behavioural and brain commonalities)



Statistics: ***p-values adjusted for search volume***

set-level		cluster-level				peak-level					mm mm mm
p	*c*	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	*T*	($Z_{=}$)	p_{uncorr}	
1.000	4	1.000	0.951	65	0.442	1.000	0.936	3.97	3.37	0.000	30 46 39
		1.000	0.951	165	0.219	1.000	0.936	3.86	3.30	0.000	56 -76 12
						1.000	0.947	2.81	2.55	0.005	52 -81 22
		1.000	0.951	248	0.136	1.000	0.936	3.52	3.07	0.001	46 -87 -9
						1.000	0.936	3.16	2.81	0.002	42 -93 -3
		1.000	0.951	45	0.528	1.000	0.936	3.14	2.80	0.003	-27 -57 58

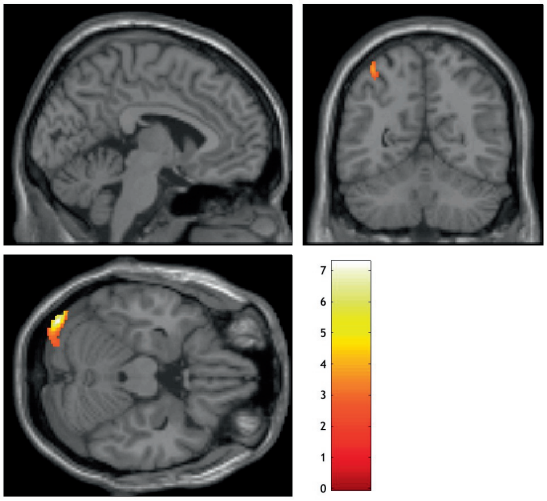
Supplementary Fig.1. GMV that correlates with alpha gain learning rate parameter in PD ON. Statistics based on a fronto-striatal and LOC mask, thesholded at p <.01 uncorrected, with a cluster size of greater than 40 contiguous voxels.



Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm
p	c	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	T	$(Z_=)$	p_{uncorr}	
1.000	11	1.000	0.951	212	0.167	0.968	0.680	4.61	3.76	0.000	−14 −76 60
		1.000	0.951	184	0.196	0.997	0.695	4.29	3.57	0.000	33 −84 −9
		1.000	0.951	85	0.377	1.000	0.746	3.98	3.38	0.000	18 −64 70
		1.000	0.951	52	0.495	1.000	0.799	3.75	3.22	0.001	28 60 −2
		1.000	0.951	59	0.466	1.000	0.799	3.73	3.21	0.001	16 −76 63
		1.000	0.951	79	0.395	1.000	0.836	3.49	3.05	0.001	−45 −68 26
		1.000	0.951	93	0.355	1.000	0.836	3.48	3.04	0.001	−40 −81 45
		1.000	0.951	58	0.470	1.000	0.867	3.28	2.90	0.002	56 −56 −20
		1.000	0.951	147	0.246	1.000	0.867	3.19	2.83	0.002	30 −80 51
						1.000	0.975	2.97	2.67	0.004	36 −82 45
		1.000	0.951	47	0.519	1.000	0.883	3.14	2.80	0.003	57 −56 42
		1.000	0.951	40	0.555	1.000	0.975	2.74	2.49	0.006	−39 −60 −4
						1.000	0.975	2.65	2.43	0.008	−40 −68 0

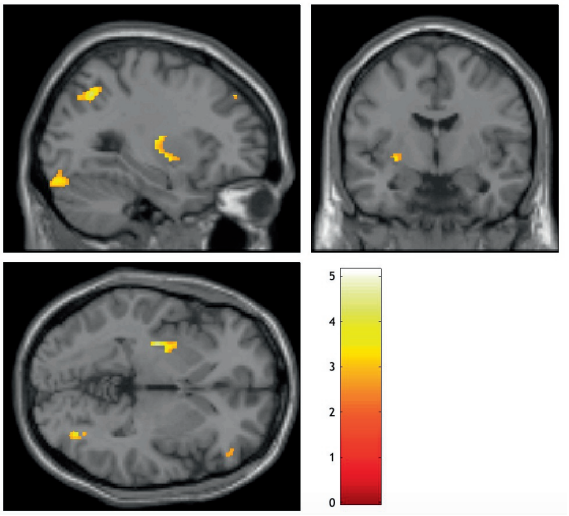
Supplementary Fig.2. GMV that correlates with alpha gain learning rate parameter in PD OFF. Statistics based on a fronto-striatal and LOC mask, thesholded at p <.01 uncorrected, with a cluster size of greater than 40 contiguous voxels.



Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm
p	*c*	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	*T*	($Z_{=}$)	p_{uncorr}	
1.000	5	0.999	0.951	326	0.093	0.036	0.731	7.28	5.03	0.000	−40 −87 −24
		1.000	0.951	79	0.397	1.000	0.961	3.87	3.30	0.000	42 −93 6
		1.000	0.951	42	0.546	1.000	0.961	3.78	3.25	0.001	−10 −86 50
		1.000	0.951	81	0.391	1.000	0.961	3.72	3.21	0.001	−38 −56 58
		1.000	0.951	132	0.273	1.000	0.961	3.69	3.18	0.001	33 −92 26

Supplementary Fig.3. GMV that correlates with alpha loss learning rate parameter in PD ON. Statistics based on a fronto-striatal and LOC mask, thesholded at p <.01 uncorrected, with a cluster size of greater than 40 contiguous voxels.



Statistics: *p-values adjusted for search volume*

set-level		cluster-level				peak-level					mm mm mm
p	c	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	k_E	p_{uncorr}	$p_{FWE\text{-}corr}$	$q_{FDR\text{-}corr}$	T	$(Z_{=})$	p_{uncorr}	
1.000	19	0.117	0.102	1373	0.002	0.761	0.529	5.14	4.06	0.000	39 -63 24
						0.987	0.529	4.49	3.69	0.000	45 -57 21
						1.000	0.561	4.13	3.47	0.000	45 -75 12
		0.139	0.102	1315	0.002	0.796	0.529	5.08	4.02	0.000	-21 -64 52
						0.986	0.529	4.50	3.70	0.000	-36 -64 42
						0.998	0.529	4.28	3.56	0.000	-40 -54 45
		0.831	0.361	613	0.024	0.814	0.529	5.05	4.01	0.000	-36 -84 -21
						1.000	0.761	3.29	2.91	0.002	-22 -86 -16
		1.000	0.621	255	0.125	0.944	0.529	4.73	3.83	0.000	-30 -14 -4
						1.000	0.706	3.59	3.12	0.001	-28 -12 12
						1.000	0.783	3.21	2.85	0.002	-30 -2 -3
		1.000	0.882	66	0.429	0.958	0.529	4.68	3.80	0.000	51 -56 -10
		1.000	0.785	84	0.370	0.977	0.529	4.57	3.74	0.000	46 -68 54
		1.000	0.744	119	0.286	0.993	0.529	4.42	3.65	0.000	-26 44 42
		1.000	0.708	170	0.204	0.999	0.561	4.16	3.49	0.000	40 -84 -22
		1.000	0.882	69	0.419	1.000	0.561	4.14	3.48	0.000	40 9 58
		0.995	0.595	361	0.072	1.000	0.578	4.03	3.41	0.000	-50 -60 14
						1.000	0.758	3.45	3.02	0.001	-44 -63 24
						1.000	0.758	3.30	2.91	0.002	-40 -74 27
		1.000	0.785	91	0.351	1.000	0.656	3.67	3.17	0.001	46 -52 42
		1.000	0.806	71	0.412	1.000	0.656	3.66	3.16	0.001	36 -74 -4
		1.000	0.806	74	0.401	1.000	0.677	3.65	3.16	0.001	-33 -69 12
		1.000	0.728	142	0.244	1.000	0.677	3.63	3.15	0.001	14 -66 57
						1.000	0.904	2.91	2.63	0.004	18 -70 51
		1.000	0.785	96	0.338	1.000	0.758	3.41	2.99	0.001	63 -57 39
		1.000	0.950	43	0.530	1.000	0.783	3.23	2.87	0.002	-9 54 36
		1.000	0.949	45	0.519	1.000	0.809	3.14	2.80	0.003	48 40 2
		1.000	0.950	40	0.546	1.000	0.858	3.04	2.72	0.003	-8 14 -8
		1.000	0.949	44	0.524	1.000	0.904	2.91	2.63	0.004	-36 -84 9

table shows 3 local maxima more than 8.0mm apart

Height threshold: T = 2.53, p = 0.010 (1.000)
Extent threshold: k = 40 voxels, p = 0.546 (1.000)
Expected voxels per cluster, <k> = 112.829
Expected number of clusters, <c> = 40.70
FWEp: 7.103, FDRp: Inf, FWEc: 1835, FDRc: 1499

Degrees of freedom = [1.0, 20.0]
FWHM = 11.2 12.3 11.7 mm mm mm; 7.5 8.2 7.8 {voxels}
Volume: 2290447 = 678651 voxels = 1344.5 resels
Voxel size: 1.5 1.5 1.5 mm mm mm; (resel = 480.29 voxels)

Supplementary Fig.4. GMV that correlates with alpha loss learning rate parameter in PD OFF. Statistics based on a fronto-striatal and LOC mask, thesholded at p <.01 uncorrected, with a cluster size of greater than 40 contiguous voxels.

Chapter 6

Summary

The research presented in this thesis explores the behavioural and brain mechanisms underlying the roles of visual attention and dopamine in value-based learning and decision-making. This is accomplished using a variety of behavioural, eye-tracking and fMRI methods. Based on a series of experiments, we show that:

1) rewards in the environment are distracting, lead to changes in eye movement behaviour, thereby interfering with current goals.
2) the learned association between rewarding objects and certain motor behaviours, e.g. an eye movement to the reward, does not translate to other motor output to obtain the reward, e.g. a manual response.
3) lower dopamine levels in Parkinson's disease (PD) patients OFF compared to ON mediation lead to an increased sensitivity to negative outcomes, i.e. higher negative learning rates, during learning, which affects subsequent decision-making processes based on the value of items accrued during learning.
4) less dopamine in PD OFF than ON leads to greater tracking of negative reward prediction errors (RPEs) in the striatum.
5) less dopamine in PD OFF than ON results in better classification of good and bad outcomes in visual object-selective cortex (OSC), but worse classification in higher-level dorsolateral prefrontal cortex (DLPFC) and putamen.
6) changes in grey matter volume (GMV) in PD OFF are associated with the extent of distractibility and negative learning rates in the putamen.

In **chapter 2**, participants performed a fast-paced eye-tracking task, in which they had to follow a black target circle around the screen. A distractor circle was presented in each trial, which signified the level of reward that could be obtained in that trial: high, low or no reward. Crucially, rewards could only be earned if the eyes correctly moved to the target, and not to the distractor. We found increased oculomotor capture to distractors in accordance with their level of reward association, i.e. more saccades were made to the distractor signifying high compared to low and no reward, even though these erroneous saccades resulted losing the associated reward. In spite of the detrimental effect of looking to distractors, participants were unable to control it, providing strong evidence that distractors suggesting potential reward in the environment capture the eyes against top-down, goal-oriented intentions of the observer. This effect furthermore got stronger over time. Since this reward effect got stronger over time, i.e. the difference between the percentage of

erroneous saccades towards the high compared to no reward-signaling distractor became significantly bigger over time, it suggests that top-down attempts to suppress maladaptive saccades are not enough to overcome the repeated capture of the eyes by high reward-signaling distractors, and are likely weakened by fatigue as the experiment progresses. The simple action of repeated approaches to reward signals appears to further induce a similar motor command, even if the action is counterproductive. It is suggestive of a sort of avalanche, habitual effect of detrimental reward-driven behaviours as motivation, effort and/or cognitive control dwindles with fatigue.

We revealed reward-driven effects on several other oculomotor properties: *i)* there was evidence for an interaction between reward-driven changes in saccades and the remote distractor effect; specifically, high reward-signaling distractors at 180 degrees from the target led to slower saccades to the target than low reward-signaling distractors at that position. *ii)* saccades fell short of the target (reduced *amplitude*) by a greater magnitude when distractors signaling higher reward were present, and this was more pronounced when distractors were closer to the target (global effect) *iii)* the eyes were drawn closer to the distractor (from the midpoint of the target) when it signaled higher reward.

We followed up these analyses with a hierarchical drift diffusion modelling in a Bayesian framework. This model included all trials from all participants, including both correct and incorrect trials, and is therefore a powerful method that use both group and individual-level information to constrain the model parameters. We found evidence for lower drift rates, i.e. slower accumulation of evidence for the target, for high compared to low and no reward-signaling distractor trials.

In **chapter 3**, we investigated whether reward-based spatial priority maps translate from overt to covert attention. In the domain of reward-based learning, covert spatial priorities have mainly been observed in a competitive context, rather than during learning itself (see Chelazzi *et al.*, 2014). We showed that, during learning, the cueing of rewarded compared to unrewarded locations led to shorter saccade latencies over time. Here, there was no difference between high and low rewarded locations – the mere potential for reward was enough to reduce latencies. In a subsequent test phrase, the eyes were to remain fixed, and stimuli were presented at multiple spatial locations thereby competing for covert attentional priority. We found that RT improvement (compared to a baseline, pre-learning block) for targets presented at the high reward location was actually significantly less than that for the low reward location. This effect was stronger for participants who experienced the high and low reward locations in opposite hemifields. The reduced RT

improvement for the high reward location was also significant compared to the no reward location when they were presented in opposite hemifields.

In a second experiment, a separate participant sample repeated the same experiment, with the exception that they were not required to remain fixated during the test phase. In this experiment, there were no differences in RT improvement at any of the locations. A further analysis of (micro)saccades during this phase showed the greatest increase in number of saccades (relative to baseline) to the high reward location, which was significantly higher than for the no reward location in the same hemifield.

The dopamine neurotransmitter is crucial for motivation, reward-learning, and movement control (Robbins and Everitt, 1996; Balleine, Delgado and Hikosaka, 2007; Draganski *et al.*, 2008; Helmich *et al.*, 2010). Parkinson's disease (PD) patients suffer from substantial depletion of dopaminergic neurons in the midbrain, and exhibit abnormalities in learning from rewards and punishments in addition to their more commonly addressed motor symptoms. In **chapter 4**, we carried out a probabilistic classification reinforcement learning task on PD patients (based on (Frank, Seeberger and Reilly, 2004)), along with a decision-making task to test how well they could translate their learning about the best and worst options to appropriate approach/avoid behaviours during decision-making. Using a Bayesian hierarchical reinforcement learning model, we found that when patients were off dopaminergic medication they were more sensitive to negative outcomes than when on medication. PD ON were better at learning which was the better choice out of two options than OFF when there was a large difference in value between the options, but were less accurate in choosing the better option when the choices were closer in value. This finding may be interpreted in terms of how people adapt to *uncertainty* and *volatility* in the environment. It has been shown that people increase their learning rate in volatile environments, by placing more emphasis on current, incoming information to aid decision-making, rather than relying too heavily on past outcomes that may no longer be reliable (Behrens *et al.*, 2007). Research has shown that people adapt their positive and negative learning rates independently according to the statistics of the environment (Pulcu and Browning, 2017), i.e. if the likelihood of a loss outcome for a particular action changes from stable to volatile, people increase their negative learning rate, and can do this while concurrently reducing their learning rate for positive outcomes that change from volatile to stable. In our reinforcement learning task, two learning rate parameters (separately for positive and negative events) are fit across responses to *all* stimulus pairs. Although each stimulus pair has a stable likelihood of receiving reward, there is substantial uncertainty since, not only does each

pair have a different reward association, but there is also very little value difference between options in the most uncertain pair (60:40 reward contingency). It is therefore possible that PD patients OFF are more sensitive to this overall uncertainty (at least in the context of negative feedback). The interaction between medication status and stimulus pair suggests that this adaptation to a higher negative learning rate is beneficial for the most uncertain pair, for which they are more accurate than PD ON, but is maladaptive for the least uncertain pair, for which PD ON perform better. It would therefore be interesting to test whether PD patients OFF perform worse if there were only stable pairs in this task, i.e. is it the presence of the more uncertain pairs that drives this increase in negative learning rate, or is this learning rate adopted, e.g. due to disease status, even when there is very little uncertainty? Another experiment to further probe these findings could have PD patients perform tasks that include volatility manipulations, as has been carried out in anxious individuals (Browning *et al.*, 2015) and in people with autism spectrum disorder (Lawson, Mathys and Rees, 2017).

As well as these changes in learning rate, as estimated from behavioural responses, we also found medication-related brain activation changes that covary with reward prediction errors (RPEs), i.e. the trial-by-trial difference between the expected value of an option and the feedback received for having chosen that option, in several brain regions, most notably in the dorsal striatum. Using targeted-ROI deconvolution analyses on these striatal regions, we assessed medication-related activation differences separately for positive and negative RPEs, and found that PD OFF showed greater negative RPE-related activation than ON in the caudate nucleus.

In the subsequent test phase, we replicated a previous finding showing that while PD ON are better than OFF at choosing the overall best option when it is presented, they are worse than OFF at avoiding the overall worst option, i.e. an approach/avoid medication-driven interaction (also see Frank, Seeberger and Reilly, 2004). Finally, we investigated whether individual-level medication-related changes during learning were predictive of the extent to which patients exhibited this interactive effect, i.e. the balance between approaching the best stimulus and avoiding the worst. We found that medication-related shifts in negative learning rate correlated with this approach/avoid difference. Specifically, PD OFF patients who had a greater negative learning rate were subsequently better at avoiding the worst option vs. approaching the best option, compared to when ON medication. We furthermore showed that medication-related differences in negative learning rate predicted the approach/avoid difference in percent signal change in the caudate nucleus. This was a negative relationship, i.e. those patients who had a

greater negative learning rate while OFF medication showed greater signal reduction during avoid compared to approach trials, in contrast to ON. This finding fits with theory describing how phasic decreases (or dips) in dopaminergic activation in response to negative events (e.g. negative RPEs) are better detected when baseline dopamine levels are lower, as exhibited by PD patients OFF medication (Frank, 2005).

In **chapter 5** we probed the brain and behavioural similarities in PD between learning from reinforcements (introduced in **chapter 4**) and attentional capture by task-irrelevant stimuli. Dopamine-driven changes in both of these domains have previously been observed in PD (Cools *et al.*, 2001; Frank, Seeberger and Reilly, 2004; Cools, Altamirano and D'Esposito, 2006b; Bloemendaal *et al.*, 2015). As well as the reinforcement learning task previously described, the same PD patients and HCs performed a behavioural attentional capture task, to assess levels of distractibility by task-irrelevant items in the environment. We found that PD ON were better at ignoring distracting stimuli than PD OFF, as illustrated by less RT interference in distractor compared to no distractor conditions, i.e. distractibility, in the attentional capture task. To assess how individual levels of distractibility affected learning performance, we included this distractibility measure in a trial-by-trial analysis of the reinforcement learning task, and included medication and disease status as covariates. We found that dopamine medication and distractibility RT interact to affect choice accuracy during learning, e.g. PD OFF patients who showed greater distractibility in the AC task performed worse in the reinforcement learning task.

fMRI data from the reinforcement learning task were used to analyze multivariate patterns of activation associated with positive and negative outcomes in fronto-striatal and visual brain regions employed during learning. We identified significant medication-driven interactions between fronto-striatal regions and visual object-selective cortex (OSC); specifically, we found greater classification accuracy for PD ON compared to OFF in both dorsolateral prefrontal cortex (DLPFC) and the putamen, but lower classification accuracy in ON compared to OFF in visual OSC. Based on this finding, we suggest that valence processing of reinforcements may occur more at a bottom-up, visual level in PD OFF but more at a top-down, reward-driven, and working-memory level in fronto-striatal regions in PD ON.

Using voxel-based morphometry (VBM) analyses on anatomical brain scans, we furthermore assessed changes in grey matter volume (GMV) associated with individual levels of distractibility, as well as with positive and negative learning rates, separately in PD ON and OFF. We

found GMV differences relating to distractibility and learning parameters in several fronto-striatal regions. Most notably, in PD OFF, we identified the right putamen as a region that shows GMV changes according to both distractibility and negative learning rate measures. This result suggests that more GMV in this region correlates with both greater interference from task-irrelevant stimuli and a greater reduction in internal estimates of the value of stimuli after receiving negative reinforcement.

Taken together, findings from the research presented in this dissertation highlight the combined functions of visual attention, eye movements, and dopamine in learning from reinforcements. Not only do rewarding objects in the environment capture our attention, but this attentional deployment plays a role in how well we learn from feedback. Cognitive control exerted by frontal brain regions, such as the DLPFC, and dopamine-driven reward signals in the striatum, interact with lower-level visual regions to maintain a balance (appropriate or otherwise) in the processing of rewards to produce desirable goal-directed behaviours according to the current context.

References

Aarts, E. *et al.* (2014) 'Dopamine and the Cognitive Downside of a Promised Bonus', *Psychological Science*, 25(4), pp. 1003–1009. doi: 10.1177/0956797613517240.

Abraham, A. *et al.* (2014) 'Machine Learning for Neuroimaging with Scikit-Learn', *Frontiers in Neuroinformatics*, 8(February), pp. 1–10. doi: 10.3389/fninf.2014.00014.

Adcock, R. A. *et al.* (2006) 'Reward-Motivated Learning: Mesolimbic Activation Precedes Memory Formation', *Neuron*, 50(3), pp. 507–517. doi: 10.1016/j.neuron.2006.03.036.

Ahn, W.-Y., Haines, N. and Zhang, L. (2017) 'Revealing neuro-computational mechanisms of reinforcement learning and decision-making with the hBayesDM package', *Computational Psychiatry*, 1, pp. 24–57. doi: 10.1101/064287.

Ahn, W., Krawitz, A. and Kim, W. (2011) 'A model-based fMRI analysis with hierarchical Bayesian parameter estimation', *Journal of Neuroscience, Psychology, and Economics*, 4(2), pp. 95–110. doi: 10.1037/a0020684.A.

Ahn, W. Y. *et al.* (2014) 'Decision-making in stimulant and opiate addicts in protracted abstinence: Evidence from computational modeling with pure users', *Frontiers in Psychology*, 5(AUG), pp. 1–15. doi: 10.3389/fpsyg.2014.00849.

Albers, A. M. *et al.* (2013) 'Shared representations for working memory and mental imagery in early visual cortex.', *Current biology : CB*. Elsevier Ltd, 23(15), pp. 1427–31. doi: 10.1016/j.cub.2013.05.065.

Alexander, G. E. and Crutcher, M. D. (1990) 'Functional architecture of basal ganglia cicuits: neural substrates of parallel processing', *Trends in Neuroscience*, 13, pp. 266–271.

Alvarez, G. a and Cavanagh, P. (2005) 'Independent resources for attentional tracking in the left and right visual hemifields.', *Psychological science*, 16(8), pp. 637–43. doi: 10.1111/j.1467-9280.2005.01587.x.

Amador, N., Schlag-Rey, M. and Schlag, J. (1998) 'Primate antisaccades. I. Behavioral characteristics', *Journal of neurophysiology*, 80(4), pp. 1775–1786.

Anderson, B. A. (2019) 'Neurobiology of value-driven attention', *Current Opinion in Psychology*. Elsevier Ltd, 29, pp. 27–33. doi: 10.1016/j.copsyc.2018.11.004.

Anderson, B. A., Laurent, P. A. and Yantis, S. (2011a) 'Learned value magnifies salience-based attentional capture.', *PloS one*, 6(11), p. e27926. doi: 10.1371/journal.pone.0027926.

Anderson, B. A., Laurent, P. A. and Yantis, S. (2011b) 'Value-driven attentional capture.', *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), pp. 10367–71. doi: 10.1073/pnas.1104047108.

Anderson, D. J. and Adolphs, R. (2014) 'A framework for studying emotions across species', *Cell*. Elsevier Inc., 157(1), pp. 187–200. doi: 10.1016/j.cell.2014.03.003.

Anderson, M. C., Bunce, J. G. and Barbas, H. (2016) 'Neurobiology of Learning and Memory

Prefrontal – hippocampal pathways underlying inhibitory control over memory', *Neurobiology of Learning and Memory*, 134(December), pp. 145–161. doi: 10.1016/j.nlm.2015.11.008.

Arsenault, J. T. *et al.* (2013) 'Dopaminergic Reward Signals Selectively Decrease fMRI Activity in Primate Visual Cortex', *Neuron*. Elsevier Inc., 77(6), pp. 1174–1186. doi: 10.1016/j.neuron.2013.01.008.

Ashburner, J. (2007) 'A fast diffeomorphic image registration algorithm', *NeuroImage*, 38(1), pp. 95–113.

Ashburner, J. and Friston, K. J. (2005) 'Unified segmentation', *Neuroimage*, 26(3), pp. 839–851.

Aston-Jones, G. and Cohen, J. D. (2005) 'An integrative theory of locus coeruleus-norepinephrine function: Adaptive Gain and Optimal Performance', *Annual Review of Neuroscience*, 28(1), pp. 403–450. doi: 10.1146/annurev.neuro.28.061604.135709.

Atallah, H. E. *et al.* (2007) 'Separate neural substrates for skill learning and performance in the ventral and dorsal striatum', *Nature Neuroscience*, 10(1), pp. 126–131. doi: 10.1038/nn1817.

Avants, B. B. *et al.* (2008) 'Symmetric diffeomorphic image registration with cross-correlation: Evaluating automated labeling of elderly and neurodegenerative brain', *Medical Image Analysis*, 12(1), pp. 26–41. doi: 10.1016/j.media.2007.06.004.

Awh, E., Belopolsky, A. . and Theeuwes, J. (2012) 'Top-down versus bottom-up attentional control: a failed theoretical dichotomy', 16(8), pp. 437–443. doi: 10.1016/j.tics.2012.06.010.Top-down.

Baayen, R. H., Davidson, D. J. and Bates, D. M. (2008) 'Mixed-effects modelling with crossed random effects for subjects and items', *Journal of Memory and Language*, 59, pp. 390–412.

Bahill, A., Clark, M. and Stark, L. (1975) 'The main sequence, a tool for studying human eye movements', *Mathematical Biosciences*, 24, pp. 191–204.

Ball, K. and Sekuler, R. (1980) 'Models of stimulus uncertainty in motion perception', *Psychological Review*, 87(5), pp. 435–469. doi: 10.1037/0033-295X.87.5.435.

Balleine, B. W. (2005) 'Neural bases of food-seeking: Affect, arousal and reward in corticostriatolimbic circuits', *Physiology and Behavior*, 86(5), pp. 717–730. doi: 10.1016/j.physbeh.2005.08.061.

Balleine, B. W., Delgado, M. R. and Hikosaka, O. (2007) 'The role of the dorsal striatum in reward and decision-making.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(31), pp. 8161–8165. doi: 10.1523/JNEUROSCI.1554-07.2007.

Balleine, B. W. and Dickinson, A. (1998) 'Goal-directed instrumental action: Contingency and incentive learning and their cortical substrates', *Neuropharmacology*, 37(4–5), pp. 407–419.

Barr, D. *et al.* (2013) 'Random effects structure for confirmatory hypothesis testing: keep it maximal', *Journal of Memory and Language*, 68(3), pp. 255–278.

Barto, A. G. (1995) 'Adaptive critics and the basal ganglia', in *Models of Information Processing in the Basal Ganglia*. Cambridge, MA: MIT Press, pp. 215–232.

Barto, A. G., Sutton, R. S. and Anderson, C. W. (1983) 'Neuronlike adaptive elements that can solve

difficult learning control problems', *IEEE Transactions on Systems, Man and Cybernetics*, 13, pp. 834–846.

Bates, D. *et al.* (2014) 'lme4: Linear mixed-effects models using Eigen and S4. R package. http://CRAN.r-project.org'.

Batista, G. E., Bazzan, A. L. and Monard, M. C. (2003) 'Balancing training data for automated annotation of keywords: a case study', *WOB*, pp. 10–18.

Bayer, H. M. and Glimcher, P. W. (2005) 'Midbrain dopamine neurons encode a quantitative reward prediction error signal', *Neuron*, 47(1), pp. 129–141. doi: 10.1016/j.neuron.2005.05.020.

Beauchamp, M. S. *et al.* (2001) 'A parametric fMRI study of overt and covert shifts of visuospatial attention', *NeuroImage*, 14(2), pp. 310–321. doi: 10.1006/nimg.2001.0788.

Becker, S. I. (2008) 'Can Intertrial Effects of Features and Dimensions Be Explained by a Single Theory?', *Journal of Experimental Psychology: Human Perception and Performance*, 34(6), pp. 1417–1440. doi: 10.1037/a0011386.

Beckers, G. *et al.* (1992) 'Transcranial magnetic stimulation of human frontal and parietal cortex impairs programming of periodic saccades', *Neuro-Ophthalmology*, 12(5), pp. 289–295.

Beckstead, R. M., Domesick, V. B. and Nauta, W. J. H. (1979) 'Efferent connections of the substantia nigra and ventral tegmental area in the rat', *Brain Research*, 175(2), pp. 191–217. doi: 10.1016/0006-8993(79)91001-1.

Behrens, T. E. J. *et al.* (2007) 'Learning the value of information in an uncertain world', *Nature Neuroscience*, 10(9), pp. 1214–1221. doi: 10.1038/nn1954.

Behrens, T. E. J. *et al.* (2018) 'What Is a Cognitive Map? Organizing Knowledge for Flexible Behavior', *Neuron*. Elsevier Inc., 100(2), pp. 490–509. doi: 10.1016/j.neuron.2018.10.002.

Behzadi, Y. *et al.* (2007) 'A component based noise correction method (CompCor) for BOLD and perfusion based fMRI', *NeuroImage*. Elsevier Inc., 37(1), pp. 90–101. doi: 10.1016/j.neuroimage.2007.04.042.

Bell, A. H., Everling, S. and Munoz, D. P. (2000) 'Influence of stimulus eccentricity and direction on characteristics of pro- and antisaccades in non-human primates', *Journal of Neurophysiology*, 84(5), pp. 2595–2604.

van Belle, J. *et al.* (2014) 'Common and unique neural networks for proactive and reactive response inhibition revealed by independent component analysis of functional MRI data', *NeuroImage*. Elsevier Inc., 103, pp. 65–74. doi: 10.1016/j.neuroimage.2014.09.014.

Belopolsky, A. V. and Theeuwes, J. (2012) 'Updating the premotor theory: The allocation of attention is not always accompanied by saccade preparation.', *Journal of Experimental Psychology: Human Perception and Performance*, 38(4), pp. 902–914. doi: 10.1037/a0028662.

van Bergen, R. S. *et al.* (2015) 'Sensory uncertainty decoded from visual cortex predicts behavior', *Nature Neuroscience*, 18(12), pp. 1728–1730. doi: 10.1016/j.physbeh.2017.03.040.

Bernheimer, H. *et al.* (1973) 'Brain dopamine and the syndromes of Parkinson and Huntington Clinical, morphological and neurochemical correlations', *Journal of the Neurological Sciences*, 20(4),

pp. 415–455. doi: 10.1016/0022-510X(73)90175-5.

Berridge, C. and Waterhouse, B. (2003) 'The locus coeruleus–noradrenergic system: modulation of behavioral state and state-dependent cognitive processes', *Brain Research Reviews*, 42, pp. 1–8. doi: 10.1016/S0165-0173(03)00143-7.

Berridge, K. C. (2000) '<BerridgehedonicimpactNeurosciBiobehvRev2000.pdf>', 4(99).

Berridge, K. C. and Kringelbach, M. L. (2015) 'Pleasure Systems in the Brain', *Neuron*. Elsevier Inc., 86(3), pp. 646–664. doi: 10.1016/j.neuron.2015.02.018.

Berridge, K. C. and Robinson, T. E. (1998) 'What is the Role of Dopamine in Reward: Hedonics, Learning, or Incentive Salience?', *Brain Research Reviews*, 28, pp. 308–67.

Berridge, K. C. and Robinson, T. E. (2003) 'Parsing reward', *Trends in Neurosciences*, 26(9), pp. 507–513. doi: 10.1016/S0166-2236(03)00233-9.

Bisley, J. W. (2011) 'The neural basis of visual attention', *Journal of Physiology*, 589(1), pp. 49–57. doi: 10.1113/jphysiol.2010.192666.

Bisley, J. W. and Goldberg, M. E. (2010) 'Attention, Intention, and Priority in the Parietal Lobe', (Critchley 1953), pp. 1–21. doi: 10.1146/annurev-neuro-060909-152823.Attention.

Bloemendaal, M. *et al.* (2015) 'Dopaminergic modulation of distracter-resistance and prefrontal delay period signal', *Psychopharmacology*, 232(6), pp. 1061–1070. doi: 10.1007/s00213-014-3741-9.

Blumenfeld, R. S. *et al.* (2010) 'Putting the Pieces Together: The Role of Dorsolateral Prefrontal Cortex in Relational Memory Encoding Effects of spatial context, cue reminders, and expectations on reconsolidation View project Recollection and aging View project Putting the Pieces Togeth', *Article in Journal of Cognitive Neuroscience*, pp. 257–265. doi: 10.1162/jocn.2010.21459.

Bódi, N. *et al.* (2009) 'Reward-learning and the novelty-seeking personality: A between-and within-subjects study of the effects of dopamine agonists on young parkinsons patients', *Brain*, 132, pp. 2385–2395. doi: 10.1093/brain/awp094.

Böhler, N. *et al.* (2014) 'Reward prospect rapidly speeds up response inhibition via reactive control', *Cognitive, Affective, & Behavioral Neuroscience*, 14, pp. 593–609.

Bosboom, J. L. W., Stoffers, D. and Wolters, E. C. (2004) 'Cognitive dysfunction and dementia in Parkinson's disease', *Journal of Neural Transmission*, 111(10–11), pp. 1303–1315. doi: 10.1007/s00702-004-0168-1.

Botvinick, M. M. (2007) 'Multilevel structure in behaviour and in the brain: A model of Fuster's hierarchy', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 362(1485), pp. 1615–1626. doi: 10.1098/rstb.2007.2056.

Boulougouris, V., Glennon, J. C. and Robbins, T. W. (2008) 'Dissociable effects of selective 5-HT2A and 5-HT2C receptor antagonists on serial spatial reversal learning in rats', *Neuropsychopharmacology*, 33(8), pp. 2007–2019. doi: 10.1038/sj.npp.1301584.

Bouret, S. and Sara, S. J. (2005) 'Network reset: a simplified overarching theory of locus coeruleus noradrenaline function', *Trends in Cognitive Sciences*, 28(11), pp. 574–582.

Braem, S. *et al.* (2012) 'Reward modulates adaptations to conflict', *Cognition*. Elsevier B.V., 125(2), pp. 324–332. doi: 10.1016/j.cognition.2012.07.015.

Brascamp, J. W., Blake, R. and Kristjánsson, Á. (2011) 'Deciding Where to Attend: Priming of Pop-Out Drives Target Selection', *Journal of Experimental Psychology: Human Perception and Performance*, 37(6), pp. 1700–1707. doi: 10.1037/a0025636.

Braun, A., Urai, A. E. and Donner, T. H. (2018) 'Adaptive History Biases Result from Confidence-weighted Accumulation of Past Choices', *The Journal of Neuroscience*, pp. 2189–17. doi: 10.1523/JNEUROSCI.2189-17.2017.

Breiter, H. C. *et al.* (2001) 'Functional Imaging of Neural Responses to Monetary Gains and Losses', *Neuron*, 30(May), pp. 619-. doi: 10.1016/S0896-6273(01)00303-8.

Brodersen, K. H. *et al.* (2012) 'Decoding the perception of pain from fMRI using multivariate pattern analysis', *NeuroImage*. Elsevier Inc., 63(3), pp. 1162–1170. doi: 10.1016/j.neuroimage.2012.08.035.

Brodersen, K. H. *et al.* (2013) 'Variational Bayesian mixed-effects inference for classification studies', *NeuroImage*. Elsevier Inc., 76, pp. 345–361. doi: 10.1016/j.neuroimage.2013.03.008.

Bromberg-Martin, E. S., Matsumoto, M. and Hikosaka, O. (2010) 'Dopamine in Motivational Control: Rewarding, Aversive, and Alerting', *Neuron*. Elsevier Inc., 68(5), pp. 815–834. doi: 10.1016/j.neuron.2010.11.022.

Brown, J., Bullock, D. and Grossberg, S. (1999) 'How the basal ganglia use parallel excitatory and inhibitory learning pathways to selectively respond to unexpected rewarding cues.', *Journal of Neuroscience*, 19(23), pp. 10502–10511. doi: 10.1523/jneurosci.3883-12.2013.

Brown, J. W., Bullock, D. and Grossberg, S. (2004) 'How laminar frontal cortex and basal ganglia circuits interact to control planned and reactive saccades', *Neural Networks*, 17, pp. 471–510.

Browning, M. *et al.* (2015) 'Anxious individuals have difficulty learning the causal statistics of aversive environments', *Nature Neuroscience*. Nature Publishing Group, 18(4), pp. 590–596. doi: 10.1038/nn.3961.

Bucker, B. *et al.* (2015) 'Reward modulates oculomotor competition between differently valued stimuli', *Vision Research*. Elsevier Ltd, 108, pp. 103–112. doi: 10.1016/j.visres.2015.01.020.

Bucker, B., Belopolsky, A. V. and Theeuwes, J. (2014) 'Distractors that signal reward attract the eyes', *Visual Cognition*, (July 2015), pp. 1–24. doi: 10.1080/13506285.2014.980483.

Bucker, B. and Theeuwes, J. (2014) 'The effect of reward on orienting and reorienting in exogenous cuing.', *Cognitive, affective & behavioral neuroscience*, (2008), pp. 635–646. doi: 10.3758/s13415-014-0278-7.

Busemeyer, J. R. and Townsend, J. T. (1993) 'Decision field theory: A dynamic-cognitive approach to decision making in an uncertain environment', *Psychological Review*, 100(3), pp. 432–459. doi: 10.1037/0033-295X.100.3.432.

Calabresi, P. *et al.* (2014) 'Direct and indirect pathways of basal ganglia: A critical reappraisal', *Nature Neuroscience*. Nature Publishing Group, 17(8), pp. 1022–1030. doi: 10.1038/nn.3743.

Camara, E., Manohar, S. and Husain, M. (2013) 'Past rewards capture spatial attention and action

choices', *Experimental Brain Research*, 230(3), pp. 291–300. doi: 10.1007/s00221-013-3654-6.

Camerer, C. and Ho, T.-H. (1999) 'Experience-weighted attraction learning in normal form games', *Econometrica*, 67(4), pp. 827–874.

Cardinal, R. N. *et al.* (2002) 'Emotion and motivation: The role of the amygdala, ventral striatum, and prefrontal cortex', *Neuroscience and Biobehavioral Reviews*, 26(3), pp. 321–352. doi: 10.1016/S0149-7634(02)00007-6.

Carello, C. D. and Krauzlis, R. J. (2004) 'Manipulating intent: Evidence for a causal role of the superior colliculus in target selection', *Neuron*, 43(4), pp. 575–583. doi: 10.1016/j.neuron.2004.07.026.

Cavanaugh, J. and Wurtz, R. H. (2004) 'Subcortical modulation of attention counters change blindness', *Journal of Neuroscience*, 24(50), pp. 11236–11243. doi: 10.1523/JNEUROSCI.3724-04.2004.

Chao, L. L. and Knight, R. T. (1995) 'Human prefrontal lesions increase distractibility to irrelevant sensory inputs', *Neuroreport*, 6, pp. 1605–1610.

Chao, L. L. and Knight, R. T. (1998) 'Contribution of human prefrontal cortex to delay performance', *Journal of Cognitive Neuroscience*, 10(2), pp. 167–177.

Chelazzi, L. *et al.* (1998) 'Responses of neurons in inferior temporal cortex during memory-guided visual search', *Journal of Neurophysiology*, 80(6), pp. 2918–2940. doi: 10.1152/jn.1998.80.6.2918.

Chelazzi, L. *et al.* (2013) 'Rewards teach visual selective attention', *Vision Research*. Elsevier Ltd, 85, pp. 58–62. doi: 10.1016/j.visres.2012.12.005.

Chelazzi, L. *et al.* (2014) 'Altering Spatial Priority Maps via Reward-Based Learning.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 34(25), pp. 8594–604. doi: 10.1523/JNEUROSCI.0277-14.2014.

Chun, M. M. (2000) 'Contextual cueing of visual attention', *Trends in Cognitive Sciences*, 4, pp. 170–177.

Chung, Y. *et al.* (2013) 'A nondegenerative penalized likelihood estimator for variance parameters in multilevel models', *Psychometrika*, 78(4), pp. 685–709.

Clark, A. (2013) 'Whatever next? Predictive brains, situated agents, and the future of cognitive science', *Behavioral and Brain Sciences*, 36(3), pp. 181–204. doi: 10.1017/S0140525X12000477.

Clatworthy, P. L. *et al.* (2009) 'Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory', *Journal of Neuroscience*, 29(15), pp. 4690–4696. doi: 10.1523/JNEUROSCI.3266-08.2009.

Cools, R. *et al.* (2001) 'Enhanced or Impaired Cognitive Function in Parkinson ' s Disease as a Function of Dopaminergic Medication and Task Demands', *Cerebral Cortex*, 11(12), pp. 1136–1143.

Cools, R. (2006) 'Dopaminergic modulation of cognitive function-implications for L-DOPA treatment in Parkinson's disease', *Neuroscience and Biobehavioral Reviews*, 30(1), pp. 1–23. doi: 10.1016/j.neubiorev.2005.03.024.

Cools, R. *et al.* (2007) 'L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease', *Neuropsychopharmacology*, 32(1), pp. 180–189. doi: 10.1038/sj.npp.1301153.

Cools, R., Altamirano, L. and D'Esposito, M. (2006a) 'Reversal learning in Parkinson's disease depends on medication status and outcome valence.', *Neuropsychologia*, 44(10), pp. 1663–73. doi: 10.1016/j.neuropsychologia.2006.03.030.

Cools, R., Altamirano, L. and D'Esposito, M. (2006b) 'Reversal learning in Parkinson's disease depends on medication status and outcome valence', *Neuropsychologia*, 44(10), pp. 1663–1673. doi: 10.1016/j.neuropsychologia.2006.03.030.

Cools, R., Clark, L. and Robbins, T. W. (2004) 'Differential Responses in Human Striatum and Prefrontal Cortex to Changes in Object and Rule Relevance', *Journal of Neuroscience*, 24(5), pp. 1129–1135. doi: 10.1523/JNEUROSCI.4312-03.2004.

Corbett, D. and Wise, R. A. (1980) 'Intracranial self-stimulation in relation to the ascending dopaminergic systems of the midbrain: a moveable electrode mapping study', *Brain Research*, 185, pp. 1–15.

Corbetta, M. (1998) 'Frontoparietal cortical networks for directing attention and the eye to visual locations: Identical, independent, or overlapping neural systems?', *Proceedings of the National Academy of Sciences*, 95, pp. 831–838. doi: 10.1088/1674-1056/20/11/114702.

Corbetta, M. and Shulman, G. L. (2002) 'Control of goal-directed and stimulus-driven attention in the brain.', *Nature reviews. Neuroscience*, 3(3), pp. 201–15. doi: 10.1038/nrn755.

Coren, S. and Hoenig, P. (1972) 'Effect of non-target stimuli upon length of voluntary saccades', *Perceptual and Motor Skills*, 34, pp. 499–508.

Costa, V. D. *et al.* (2015) 'Reversal Learning and Dopamine: A Bayesian Perspective', *Journal of Neuroscience*, 35(6), pp. 2407–2416. doi: 10.1523/JNEUROSCI.1989-14.2015.

Cousineau, D. (2005) 'Confidence intervals in within-subject designs: A simpler solution to Loftus and Masson's method', *Tutorials in Quantitative Methods for Psychology*, 1(1), pp. 42–45. doi: no DOI found.

Cox, S. M. L. *et al.* (2015) 'Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes', *NeuroImage*. Elsevier B.V., 109, pp. 95–101. doi: 10.1016/j.neuroimage.2014.12.070.

Critchley, M. (1953) *The Parietal Lobes*. London: Edward Arnold.

Crockett, M. J., Clark, L. and Robbins, T. W. (2009) 'Reconciling the Role of Serotonin in Behavioral Inhibition and Aversion: Acute Tryptophan Depletion Abolishes Punishment-Induced Inhibition in Humans', *Journal of Neuroscience*, 29(38), pp. 11993–11999. doi: 10.1523/JNEUROSCI.2513-09.2009.

Curtis, C. E. and D'Esposito, M. (2003) 'Persistent activity in the prefrontal cortex during working memory', *Trends in Cognitive Sciences*, 7(9), pp. 415–423. doi: 10.1016/S1364-6613(03)00197-9.

Dale, A. M., Fischl, B. and Sereno, M. I. (1999) 'Cortical surface-based analysis: I. Segmentation and surface reconstruction', *NeuroImage*, 9(2), pp. 179–194. doi: 10.1006/nimg.1998.0395.

Damasio, A. and Carvalho, G. B. (2013) 'The nature of feelings: evolutionary and neurobiological origins', *Nature Reviews Neuroscience*, 14, pp. 143–152.

Dauer, W. and Przedborski, S. (2003) 'Parkinson's disease: Mechanisms and models', *Neuron*, 39(6), pp. 889–909. doi: 10.1016/S0896-6273(03)00568-3.

Davidson, M. C. *et al.* (2004) 'Differential cingulate and caudate activation following unexpected nonrewarding stimuli', *NeuroImage*, 23(3), pp. 1039–1045. doi: 10.1016/j.neuroimage.2004.07.049.

Daw, N. D. *et al.* (2011) 'Model-based influences on humans' choices and striatal prediction errors', *Neuron*. Elsevier Inc., 69(6), pp. 1204–1215. doi: 10.1016/j.neuron.2011.02.027.

Daw, N. D. (2011) 'Trial-by-trial data analysis using computational models', *Decision Making, Affect, and Learning*, 23, pp. 3–38. doi: 10.1093/acprof:oso/9780199600434.003.0001.

Daw, N. D., Kakade, S. and Dayan, P. (2002) 'Opponent interactions between serotonin and dopamine', *Neural Networks*, 15(4–6), pp. 603–616. doi: 10.1016/S0893-6080(02)00052-7.

Daw, Nathaniel D, Niv, Y. and Dayan, P. (2005) 'Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control.', *Nature neuroscience*, 8(12), pp. 1704–1711. doi: 10.1038/nn1560.

Daw, Nathaniel D., Niv, Y. and Dayan, P. (2005) 'Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control', *Nature Neuroscience*, 8(12), pp. 1704–1711. doi: 10.1038/nn1560.

Dayan, P. and Balleine, B. W. (2002) 'Reward, Motivation, and Reinforcement Learning', *Neuron*, 36, pp. 285–298.

Dayan, P. and Huys, Q. J. M. (2008) 'Serotonin, inhibition, and negative mood', *PLoS Computational Biology*, 4(2). doi: 10.1371/journal.pcbi.0040004.

Delgado, M. R. *et al.* (2005) 'An fMRI study of reward-related probability learning', *NeuroImage*, 24(3), pp. 862–873. doi: 10.1016/j.neuroimage.2004.10.002.

DeLong, M. R. and Georgopoulos, A. P. (1981) 'Motor functions of the basal ganglia', in *Handbook of Physiology, The Motor System, Motor Control*. Bethesda, MD, pp. 1017–1061.

den Ouden, H. E. M. *et al.* (2013) 'Dissociable Effects of Dopamine and Serotonin on Reversal Learning', *Neuron*, 80(6), p. 1572. doi: 10.1016/j.neuron.2013.12.008.

Deniau, J. M., Thierry, A. M. and Feger, J. (1980) 'Electrophysiological identification of mesencephalic ventromedial tegmental (VMT) neurons projecting to the frontal cortex, septum and nucleus accumbens', *Brain Research*, 189(2), pp. 315–326. doi: 10.1016/0006-8993(80)90093-1.

Desimone, R. (1998) 'Visual attention mediated by biased competition in extrastriate visual cortex', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 353(1373), pp. 1245–1255. doi: 10.1098/rstb.1998.0280.

Desimone, R. and Duncan, J. (1995) 'NEURAL MECHANISMS OF SELECTIVE VISUAL', pp. 193–222.

Deubel, H. (2008) 'The time course of presaccadic attention shifts', *Psychological Research*, 72, pp.

630–640.

Deubel, H. and Schneider, W. X. (1996) 'Saccade target selection and object recognition: Evidence for a common attentional mechanism', *Vision Research*, 36(12), pp. 1827–1837. doi: 10.1016/0042-6989(95)00294-4.

Dickinson, A. and Balleine, B. (1994) 'Motivational control of goal-directed action', *Animal Learning & Behavior*, 22(1), pp. 1–18. doi: 10.3758/BF03199951.

Dickinson, A., Nicholas, D. J. and Adams, C. D. (1983) 'The effect of the instrumental training contingency on susceptibility to reinforcer devaluation', *Quarterly Journal of Experimental Psychology Section B*, 35(1), pp. 35–51.

Ding, L. and Gold, J. I. (2012) 'Neural correlates of perceptual decision making before, during, and after decision commitment in monkey frontal eye field', *Cerebral Cortex*, 22(5), pp. 1052–1067. doi: 10.1093/cercor/bhr178.

Ding, L. and Hikosaka, O. (2006) 'Comparison of reward modulation in the frontal eye field and caudate of the macaque.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(25), pp. 6695–6703. doi: 10.1523/JNEUROSCI.0836-06.2006.

Doll, B. B. *et al.* (2016) 'Variability in Dopamine Genes Dissociates Model-Based and Model-Free Reinforcement Learning', *Journal of Neuroscience*, 36(4), pp. 1211–1222. doi: 10.1523/JNEUROSCI.1901-15.2016.

Doré-Mazars, K., Pouget, P. and Beauvillain, C. (2004) 'Attentional selection during preparation of eye movements.', *Psychological research*, 69(1–2), pp. 67–76. doi: 10.1007/s00426-003-0166-1.

Dorris, M. C., Olivier, E. and Munoz, D. P. (2007) 'Competitive integration of visual and preparatory signals in the superior colliculus during saccadic programming.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(19), pp. 5053–5062. doi: 10.1523/JNEUROSCI.4212-06.2007.

Doyle, M. and Walker, R. (2001) 'Curved saccade trajectories: voluntary and reflexive saccades curve away from irrelevant distractors.', *Experimental brain research. Experimentelle Hirnforschung. Experimentation cerebrale*, 139(3), pp. 333–344. doi: 10.1007/s002210100742.

Draganski, B. *et al.* (2008) 'Evidence for segregated and integrative connectivity patterns in the human basal ganglia', *Journal of Neuroscience*, 28(28), pp. 7143–7152. doi: 10.1523/JNEUROSCI.1486-08.2008.

Durstewitz, D. and Seamans, J. K. (2008) 'The Dual-State Theory of Prefrontal Cortex Dopamine Function with Relevance to Catechol-O-Methyltransferase Genotypes and Schizophrenia', *Biological Psychiatry*, 64(9), pp. 739–749. doi: 10.1016/j.biopsych.2008.05.015.

Edwards, M. J., Quinn, N. and Bhatia, K. P. (2008) *Parkinson's disease and other movement disorders*. Oxford University Press.

Egeth, H. E. and Yantis, S. (1997) 'Visual attention: Control, representation and time course', *Annual Review of Psychology*, 48, pp. 269–297.

Eickhoff, S. B. *et al.* (2005) 'A new SPM toolbox for combining probabilistic cytoarchitectonic maps and functional imaging data', *NeuroImage*, 5, pp. 1325–1335.

Engbert, R. and Kliegl, R. (2003) 'Microsaccades uncover the orientation of covert attention', *Vision Research*, 43, pp. 1035–1045. doi: 10.1016/S0042-6989(03)00084-1.

Engelmann, J. B. *et al.* (2009) 'Combined effects of attention and motivation on visual task performance: transient and sustained motivational effects.', *Frontiers in human neuroscience*, 3(March), p. 4. doi: 10.3389/neuro.09.004.2009.

Engelmann J and L, P. (2007) 'Motivation sharpens exogenous spatial attention', *Emotion*, 1, pp. 64–72.

Engels, Gwenda *et al.* (2018) 'Clinical pain and functional network topology in Parkinson's disease : a resting-state fMRI study', *Journal of Neural Transmission*. Springer Vienna, 0(0), p. 0. doi: 10.1007/s00702-018-1916-y.

Engels, G *et al.* (2018) 'Dynamic functional connectivity and symptoms of Parkinson's disease: a resting-state fMRI study', *Frontiers in Aging Neuroscience*, 10(388). doi: 10.3389/fnagi.2018.00388.

Eriksen, B. A. and Ericksen, C. W. (1974) 'Effects of noise letters upon identification of a target letter in a nonsearch task', *Perception & Psychophysics*, 16, pp. 143–149.

Eriksen, C. W. and Hoffman, J. E. (1973) 'The extent of processing of noise elements during selective coding from visual display', *Perception & Psychophysics*, 14, pp. 155–160.

Esteban, Oscar *et al.* (2018) 'FMRIPrep: a robust preprocessing pipeline for functional MRI', *bioRxiv*, p. 306951. doi: 10.1101/306951.

Esteban, O. *et al.* (2018) *poldracklab/fmriprep: 1.1.1*, *Zenodo*. Available at: https://doi.org/10.5281/zenodo.1285255.

Everitt, B. J. and Robbins, T. W. (2005) 'Neural systems of reinforcement for drug addiction: From actions to habits to compulsion', *Nature Neuroscience*, 8(11), pp. 1481–1489. doi: 10.1038/nn1579.

Everitt, B. J. and Robbins, T. W. (2013) 'From the ventral to the dorsal striatum: Devolving views of their roles in drug addiction', *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd, 37(9), pp. 1946–1954. doi: 10.1016/j.neubiorev.2013.02.010.

Fahn, S., Elton, R. . and Members of the UPDRS Development Committee (1987) 'Unified Parkinson's disease rating scale', in *Recent Developments in Parkinson's Disease*. 2nd edn. Florham Park, NJ: Macmillan Health Care Information, pp. 153–163.

Failing, M. *et al.* (2015) 'Oculomotor capture by stimuli that signal the availability of reward', pp. 2316–2327. doi: 10.1152/jn.00441.2015.

Failing, M. F. and Theeuwes, J. (2014) 'Exogenous visual orienting by reward', 14, pp. 1–9. doi: 10.1167/14.5.6.doi.

Failing, M. and Theeuwes, J. (2016) 'Reward alters the perception of time', *Cognition*. Elsevier B.V., 148, pp. 19–26. doi: 10.1016/j.cognition.2015.12.005.

Fearnley, J. M. and Lees, A. J. (1991) 'Ageing and Parkinson's disease: Substantia nigra regional selectivity', *Brain*, 114(1), pp. 2283–2301. doi: 10.1212/wnl.53.5.1158.

Fecteau, J. H. and Munoz, D. P. (2006) 'Salience, relevance, and firing: a priority map for target

selection', *Trends in Cognitive Sciences*, 10(8), pp. 382–390. doi: 10.1071/AH14258.

Fibiger, H. C. *et al.* (1987) 'The role of dopamine in intracranial self-stimulation of the ventral tegmental area', *Journal of Neuroscience*, 7(12), pp. 3888–3896. doi: 10.1523/jneurosci.07-12-03888.1987.

Findlay, J. M. (1982) 'Global processing for saccadic eye movements', *Vision Research*, 22, pp. 1033–1045.

Fischl, B. *et al.* (2002) 'Whole Brain Segmentation', *Neuron*, 33(3), pp. 341–355. doi: 10.1016/s0896-6273(02)00569-x.

de Fockert, J. *et al.* (2004) 'Neural correlates of attentional capture in visual search', *J Cogn Neurosci*, 16(5), pp. 751-9 ST-Neural correlates of attentional captu. doi: 10.1162/089892904970762.

Fonov, V. S. *et al.* (2009) 'Unbiased nonlinear average age-appropriate brain templates from birth to adulthood', *Neuroimage*, 47.

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, pp. 51–72. doi: 10.1162/0898929052880093.

Frank, M. J. *et al.* (2007) 'Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning.', *Proceedings of the National Academy of Sciences of the United States of America*, 104(41), pp. 16311–16316. doi: 10.1073/pnas.0706111104.

Frank, M. J. (2007) 'Hold your horses: Impulsivity, deep brain stimulation, and medication in Parkinsonism', *Science*, 318(23), pp. 1309–1312. doi: 10.1126/science.1146157.

Frank, M. J., Seeberger, L. C. and Reilly, R. C. O. (2004) 'By Carrot or by Stick : Cognitive Reinforcement Learning in Parkinsonism', *Science*, 306, pp. 1940–1943.

Friston, K. J. *et al.* (1995) 'Analysis of fMRI time-series revisited', *Neuroimage*, 2, pp. 45–53.

Friston, K. J., Jezzard, P. and Turner, R. (1994) 'Analysis of functional MRI time-series', *Human Brain Mapping*, 1, pp. 153–71.

Frund, I., Wichmann, F. A. and Macke, J. H. (2014) 'Quantifying the effect of intertrial dependence on perceptual decisions', *Journal of Vision*, 14(7), pp. 9–9. doi: 10.1167/14.7.9.

Gelman, A. and Rubin, D. B. (1992) 'Inference from Iterative Simulation Using Multiple Sequences', *Statistical Science*, 7(4), pp. 457–472.

Gläscher, J., Hampton, A. N. and O'Doherty, J. P. (2009) 'Determining a role for ventromedial prefrontal cortex in encoding action-based value signals during reward-related decision making', *Cerebral Cortex*, 19(2), pp. 483–495. doi: 10.1093/cercor/bhn098.

Glimcher, P. (2002) 'Decisions, Decisions, Decisions: Review Choosing a Biological Science of Choice', *Neuron*, 36(2), pp. 1–10. Available at: papers2://publication/uuid/EA593748-4D7D-4B7D-BED4-9BC5FC3C3D57.

Glimcher, P. W. (2003) 'The neurobiology of visual-saccadic decision making', *Annual Review of Neuroscience*, 26, pp. 133–179.

Godijn, R. and Theeuwes, J. (2002) 'Programming of endogenous and exogenous saccades: evidence for a competitive integration model.', *Journal of experimental psychology. Human perception and performance*, 28(5), pp. 1039–1054. doi: 10.1037/0096-1523.28.5.1039.

Goffart, L., Hafed, Z. M. and Krauzlis, R. J. (2012) 'Visual fixation as equilibrium: Evidence from superior colliculus inactivation', *Journal of Neuroscience*, 32(31), pp. 10627–10636.

Gold, J. N. and Shadlen, M. N. (2001) 'Neural computations that underlie decisions about sensory stimuli', *Trends in Cognitive Sciences*, 5, pp. 10–16.

Gorgolewski, K. *et al.* (2011) 'Nipype: A Flexible, Lightweight and Extensible Neuroimaging Data Processing Framework in Python', *Frontiers in Neuroinformatics*, 5(August). doi: 10.3389/fninf.2011.00013.

Gorgolewski, K. *et al.* (2017) *Nipype: a flexible, lightweight and extensible neuroimaging data processing framework in Python. 0.13.0*, *Zenodo*. Available at: 10.5281/zenodo.581704.

Gottlieb, J. (2012) 'Attention, learning, and the value of information.', *Neuron*. Elsevier Inc., 76(2), pp. 281–95. doi: 10.1016/j.neuron.2012.09.034.

Gottlieb, J. *et al.* (2014) 'Attention, Reward, and Information Seeking', *Journal of Neuroscience*, 34(46), pp. 15497–15504. doi: 10.1523/JNEUROSCI.3270-14.2014.

Gottlieb, J. P. and Kusunoki, M. (1998) 'The representation of visual salience in monkey parietal cortex', 791(1994), pp. 481–484.

Graef, S. *et al.* (2010) 'Differential Influence of Levodopa on Reward-Based Learning in Parkinson's Disease', *Frontiers in Human Neuroscience*, 4(October), pp. 1–13. doi: 10.3389/fnhum.2010.00169.

Gramfort, A. *et al.* (2013) 'MEG and EEG data analysis with MNE-Python', *Frontiers in Neuroscience*, 7, pp. 1–13. doi: 10.3389/fnins.2013.00267.

Gramfort, A. *et al.* (2014) 'MNE software for processing MEG and EEG data', *NeuroImage*. Elsevier Inc., 86, pp. 446–460. doi: 10.1016/j.neuroimage.2013.10.027.

Green, D. M. (1964) 'Consistency of auditory detection judgments', *Psychological Review*, 71(5), pp. 392–407.

Greve, D. N. and Fischl, B. (2009) 'Accurate and robust brain image alignment using boundary-based registration', *NeuroImage*. Elsevier Inc., 48(1), pp. 63–72. doi: 10.1016/j.neuroimage.2009.06.060.

Grill-Spector, K., Kourtzi, Z. and Kanwisher, N. (2001) 'The lateral occipital complex and its role in object recognition', *Vision Research*, 41(10–11), pp. 1409–1422. doi: 10.1016/S0042-6989(01)00073-6.

Grogan, J. P. *et al.* (2017) 'Effects of dopamine on reinforcement learning and consolidation in parkinson's disease', *eLife*, 6, pp. 1–23. doi: 10.7554/eLife.26801.

de Haan, B., Morgan, P. S. and Rorden, C. (2008) 'Covert orienting of attention and overt eye movements activate identical brain regions', *Brain Research*, 1204, pp. 102–111. doi: 10.1016/j.brainres.2008.01.105.

Haber, S. N. and Knutson, B. (2010) 'The reward circuit: Linking primate anatomy and human

imaging', *Neuropsychopharmacology*. Nature Publishing Group, 35(1), pp. 4–26. doi: 10.1038/npp.2009.129.

Hafed, Z. M., Goffart, L. and Krauzlis, R. J. (2009) 'A neural mechanism for microsaccade generation in the primate superior colliculus', *Science*, 323, pp. 940–943.

Hafting, T. *et al.* (2005) 'Microstructure of a spatial map in the entorhinal cortex', *Nature*, 436(7052), pp. 801–806. doi: 10.1038/nature03721.

Hallett, P. E. (1978) 'Primary and secondary saccades to goals defined by instructions', *Vision Research*, 18, pp. 1279–1296.

Hammers, A. *et al.* (2003) 'Three-dimensional maximum probability atlas of the human brain, with particular reference to the temporal lobe', *Human Brain Mapping*, 19(4), pp. 224–247. doi: 10.1002/hbm.10123.

Hanes, D. P. and Schall, J. D. (1996) 'Neural control of voluntary movement initiation', *Science*, 274, pp. 427–430.

Harley, C. W. (2004) 'Norepinephrine and dopamine as learning signals', *Neural Plasticity*, 11(3–4), pp. 191–204. doi: 10.1155/NP.2004.191.

Harrison, S. a and Tong, F. (2009) 'Decoding reveals the contents of visual working memory in early visual areas.', *Nature*. Nature Publishing Group, 458(7238), pp. 632–5. doi: 10.1038/nature07832.

Haruno, M. and Kawato, M. (2006) 'Different Neural Correlates of Reward Expectation and Reward Expectation Error in the Putamen and Caudate Nucleus During Stimulus-Action-Reward Association Learning', *Journal of Neurophysiology*, 95(2), pp. 948–959. doi: 10.1152/jn.00382.2005.

Haxby, J. V *et al.* (2001) 'Distributed and overlapping representations of faces and objects in ventral temporal cortex.', *Science*, 293(5539), pp. 2425–30. doi: 10.1126/science.1063736.

Hayden, B. Y. *et al.* (2011) 'Surprise Signals in Anterior Cingulate Cortex: Neuronal Encoding of Unsigned Reward Prediction Errors Driving Adjustment in Behavior', *Journal of Neuroscience*, 31(11), pp. 4178–4187. doi: 10.1523/JNEUROSCI.4652-10.2011.

Haynes, J.-D. and Rees, G. (2006) 'Decoding mental states from brain activity in humans.', *Nature reviews. Neuroscience*, 7(7), pp. 523–34. doi: 10.1038/nrn1931.

Hazy, T. E., Frank, M. J. and O'Reilly, R. C. (2010) 'Neural mechanisms of acquired phasic dopamine responses in learning', *Neuroscience and Biobehavioral Reviews*. Elsevier Ltd, 34(5), pp. 701–720. doi: 10.1016/j.neubiorev.2009.11.019.

Heitz, R. P. and Schall, J. D. (2012) 'Neural Mechanisms of Speed-Accuracy Tradeoff', *Neuron*. Elsevier Inc., 76(3), pp. 616–628. doi: 10.1016/j.neuron.2012.08.030.

Helmich, R. C. *et al.* (2010) 'Spatial remapping of cortico-striatal connectivity in parkinson's disease', *Cerebral Cortex*, 20(5), pp. 1175–1186. doi: 10.1093/cercor/bhp178.

Henik, A., Rafal, R. and Rhodes, D. (1994) 'Endogenously generated and visually guided saccades after lesions of the human frontal eye fields', *Journal of Cognitive Neuroscience*, 6(4), pp. 400–411.

Hickey, C., Chelazzi, L. and Theeuwes, J. (2010) 'Reward changes salience in human vision via the

anterior cingulate.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 30(33), pp. 11096–103. doi: 10.1523/JNEUROSCI.1026-10.2010.

Hickey, C., Chelazzi, L. and Theeuwes, J. (2014) 'Reward-priming of location in visual search', *PLoS ONE*, 9(7). doi: 10.1371/journal.pone.0103372.

Hickey, C. and Peelen, M. V. (2015) 'Neural Mechanisms of Incentive Salience in Naturalistic Human Vision', *Neuron*. Elsevier Inc., 85(3), pp. 512–518. doi: 10.1016/j.neuron.2014.12.049.

Hickey, C. and Theeuwes, J. (2011) 'Context and competition in the capture of visual attention.', *Attention, perception & psychophysics*, 73(7), pp. 2053–64. doi: 10.3758/s13414-011-0168-9.

Hickey, C. and Zoest, W. Van (2013) 'Reward-associated stimuli capture the eyes in spite of strategic attentional set'.

Hikosaka, O. *et al.* (2000) 'Role of the basal ganglia in the control of purposive saccadic eye movements', *Physiological Reviews*, 80(3), pp. 953–978. doi: http://physrev.physiology.org/content/80/3/953.

Hikosaka, O. (2014) 'Basal ganglia circuits for reward value-guided behavior', *Annual review of neuroscience*, pp. 289–306. doi: 10.1146/annurev-neuro-071013-013924.Basal.

Hikosaka, O. *et al.* (2018) 'Parallel basal ganglia circuits for decision making', *Journal of Neural Transmission*. Springer Vienna, 125(3), pp. 515–529. doi: 10.1007/s00702-017-1691-1.

Hikosaka, O., Nakamura, K. and Nakahara, H. (2006) 'Basal ganglia orient eyes to reward.', *Journal of neurophysiology*, 95(2), pp. 567–584. doi: 10.1152/jn.00458.2005.

Hikosaka, O. and Wurtz, R. . (1983) 'Visual and oculomotor functions of monkey substantia nigra pars reticulata. I. Relation of visual and auditory responses to saccades', *Journal of neurophysiology*, 49, pp. 1230–53.

Hillstrom, A. P. (2000) 'Repetition effects in visual search', *Perception & Psychophysics*, 62, pp. 800–817.

Hoehn, M. M. and Yahr, M. D. (1967) 'Parkinsonism : onset, progression, and mortality Parkinsonism: onset, progression, and mortality', *Neurology*, 17(5), pp. 427–442. doi: 10.1212/WNL.17.5.427.

Hoffman, M. D. and Gelman, A. (2016) 'The No-U-Turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo', *Journal of Machine Learning Research*, 15, pp. 1351–1381. Available at: https://www.youtube.com/watch?v=uDZmpmMcrUQ.

Holland, P. C. and Gallagher, M. (2004) 'Amygdala-frontal interactions and reward expectancy', *Current Opinion in Neurobiology*, 14(2), pp. 148–155. doi: 10.1016/j.conb.2004.03.007.

de Hollander, G. and Knapen, T. (2017) *nideconv*. Available at: https://response-fytter.readthedocs.io/en/latest/index.html.

Hommel, B. and Prinz, W. (1997) 'Theoretical issues in stimulus-response compatibility', *Advances in psychology*, 118.

Horga, G. *et al.* (2015) 'Changes in corticostriatal connectivity during reinforcement learning in

humans', *Human Brain Mapping*, 36(2), pp. 793–803. doi: 10.1002/hbm.22665.

Hornak, J. *et al.* (2004) 'Reward-related reversal learning after surgical excisions in orbito-frontal or dorsolateral prefrontal cortex in humans', *Journal of Cognitive Neuroscience*, 16(3), pp. 463–478. doi: 10.1162/089892904322926791.

Houk, J. C. (1995) 'Information processing in modular circuits linking basal ganglia and cerebral cortex', in Houk, J. C., Davis, J. L., and Beiser, D. G. (eds) *Models of Information Processing in the Basal Ganglia*. Cambridge, MA: The MIT Press, pp. 3–10.

Howard, J. D. *et al.* (2015) 'Identity-specific coding of future rewards in the human orbitofrontal cortex', *Proceedings of the National Academy of Sciences of the United States of America*, 112(16), pp. 5195–5200. doi: 10.1073/pnas.1503550112.

Huang, L., Holcombe, A. O. and Pashler, H. (2004) 'Repetition priming in visual search: Episodic retrieval, not feature priming', *Memory & Cognition*, 32(1), pp. 12–20.

Hunt, A. R. and Kingstone, A. (2003) 'Covert and overt voluntary attention: Linked or independent?', *Cognitive Brain Research*, 18, pp. 102–105. doi: 10.1016/j.cogbrainres.2003.08.006.

Huntenburg, J. M. *et al.* (2012) *Evaluating nonlinear coregistration of BOLD EPI and T1 images*, *Proceedings of the Organisation of Human Brain Mapping*. Available at: http://hdl.handle.net/11858/00-001M-0000-002B-1CB5-A.

Huys, Q. J. M., Maia, T. V and Frank, M. J. (2016) 'Computational psychiatry as a bridge between neuro- science and clinical applications', (February), pp. 1–21. doi: 10.1038/nn.4238.

Itti, L. and Koch, C. (2000) 'A saliency-based search mechanism for overt and covert shifts of visual attention', *Vision Research*, 40, pp. 1489–1506. doi: 10.1016/S0042-6989(99)00163-7.

Itti, L. and Koch, C. (2001) 'Computational modelling of visual attention', *Nature Reviews Neuroscience*, 2(3), pp. 194–203. doi: 10.1038/35058500.

Jahfari, S. *et al.* (2018) 'Cross-task contributions of fronto-basal ganglia circuitry in response inhibition and conflict-induced slowing', *Cerebral Cortex*, (April), pp. 1–15. doi: 10.1093/cercor/bhy076.

Jahfari, S. and Theeuwes, J. (2017) 'Sensitivity to value-driven attention is predicted by how we learn from value', *Psychonomic Bulletin & Review*. Psychonomic Bulletin & Review. doi: 10.3758/s13423-016-1106-6.

Jahfari, S., Theeuwes, J. and Knapen, T. (2019) 'Learning in visual regions as support for the bias in future value-driven choice', *biorxiv*, p. 523340. doi: 10.1101/523340.

Jeffreys, H. (1998) *The theory of probability*. Oxford, UK: Oxford University Press.

Jenkinson, M. *et al.* (2002) 'Improved optimization for the robust and accurate linear registration and motion correction of brain images', *NeuroImage*, 17(2), pp. 825–841. doi: 10.1016/S1053-8119(02)91132-8.

Jocham, G., Klein, T. a and Ullsperger, M. (2011) 'Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(5), pp. 1606–1613.

doi: 10.1523/JNEUROSCI.3904-10.2011.

Joel, D., Niv, Y. and Ruppin, E. (2002) 'Actor-critic models of the basal ganglia: new anatomical and computational perspectives', *Neural Networks*, 15(4–6), pp. 535–547. doi: 10.1016/S0893-6080(02)00047-3.

Kable, J. W. and Glimcher, P. W. (2007) 'The neural correlates of subjective value during intertemporal choice', *Nature Neuroscience*, 10(12), pp. 1625–1633. doi: 10.1038/nn2007.

Kable, J. W. and Glimcher, P. W. (2009) 'The Neurobiology of Decision: Consensus and Controversy', *Neuron*. Elsevier Inc., 63(6), pp. 733–745. doi: 10.1016/j.neuron.2009.09.003.

Kahneman, D. (2011) *Thinking, Fast and Slow*. Farrar, Straus and Giroux.

Kahnt, T. *et al.* (2009) 'Dorsal Striatal–midbrain Connectivity in Humans Predicts How Reinforcements Are Used to Guide Decisions', *Journal of Cognitive Neuroscience*, 21(7), pp. 1332–1345. doi: 10.1162/jocn.2009.21092.

Kamitani, Y. and Tong, F. (2005) 'Decoding the visual and subjective contents of the human brain.', *Nature neuroscience*, 8(5), pp. 679–85. doi: 10.1038/nn1444.

Kanwisher, N., McDermott, J. and Chun, M. M. (1997) 'The fusiform face area: a module in human extrastriate cortex specialized for face perception', *Journal of Neuroscience*, 17, pp. 4302–4311.

Kastner, S. *et al.* (1998) 'Attentional response modulation in human extrastriate cortex as revealed by functional MRI', *Science*, 282, pp. 108–111. doi: 10.1016/s1053-8119(18)30877-2.

Kawagoe, R., Takikawa, Y. and Hikosaka, O. (1998) 'Expectation of reward modulates cognitive signals in the basal ganglia', *Nature Neuroscience*, 1(5), pp. 411–416. doi: 10.1038/1625.

Kim, H. *et al.* (2009) 'Role of Striatum in Updating Values of Chosen Actions', *Journal of Neuroscience*, 29(47), pp. 14701–14712. doi: 10.1523/jneurosci.2728-09.2009.

Klein, A. *et al.* (2017) *Mindboggling morphometry of human brains*, *PLoS Computational Biology*. doi: 10.1371/journal.pcbi.1005350.

Klein, R. (1980) 'Does oculomotor readiness mediate cognitive control of visual attention?', *Attention and Performance*, 8, pp. 259–276.

Koller, W. C. and Melamed, E. (2007) 'Parkinson's disease and related disorders: part 1', in *Handbook of clinical neurology*. Philadelphia: Elsevier.

Konkle, T. *et al.* (2010) 'Conceptual distinctiveness supports detailed visual long-term memory for real-world objects', *Journal of Experimental Psychology: General*, 139(3), pp. 558–578. doi: 10.1037/a0019165.

Kourtzi, Z. *et al.* (2003) 'Human Lateral Occipital Complex Representation of Perceived Object Shape by the Human Lateral Occipital Complex', *Cerebral Cortex*, 13(9), pp. 911–920. doi: 10.1126/science.1061133.

Kowler, E. *et al.* (1995) 'The Role of saccades in programming attention', *Vision Res.*, pp. 1897–1916.

Krauzlis, R. J., Goffart, L. and Hafed, Z. M. (2017) 'Neuronal control of fixation and fixational eye movements', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372.

Kriegeskorte, N., Goebel, R. and Bandettini, P. (2006) 'Information-based functional brain mapping.', *Proceedings of the National Academy of Sciences of the United States of America*, 103(10), pp. 3863–8. doi: 10.1073/pnas.0600244103.

Kruijne, W. and Meeter, M. (2015) 'The long and the short of priming in visual search', *Attention, Perception, and Psychophysics*, 77(5), pp. 1558–1573. doi: 10.3758/s13414-015-0860-2.

Kruschke, J. (2015) *Doing Bayesian data analysis: A tutorial introduction with R, JAGS and Stan*. 2nd edn. London, UK: Academic Press/Elsevier.

Lavie, N. and de Fockert, J. (2006) 'Frontal control of attentional capture in visual search', *Visual Cognition*, 14(4–8), pp. 863–876. doi: 10.1080/13506280500195953.

Lawson, R. P., Mathys, C. and Rees, G. (2017) 'Adults with autism overestimate the volatility of the sensory environment', *Nature Neuroscience*, 20(9), pp. 1293–1299. doi: 10.1038/nn.4615.

Lee, C. (1988) 'Population coding of saccadic eye movements by neurons in the superior colliculus'.

Lee, M. D. (2011) 'How cognitive modeling can benefit from hierarchical Bayesian models', *Journal of Mathematical Psychology*. Elsevier Inc., 55(1), pp. 1–7. doi: 10.1016/j.jmp.2010.08.013.

Lefebvre, G. *et al.* (2017) 'Behavioural and neural characterization of optimistic reinforcement learning', *Nature Human Behaviour*. Macmillan Publishers Limited, part of Springer Nature., 1(4), pp. 1–9. doi: 10.1038/s41562-017-0067.

Leigh, R. J. and Zee, D. S. (1999) *The neurology of eye movements*. New York: Oxford University Press.

Della Libera, C. and Chelazzi, L. (2006) 'Visual selective attention and the effects of monetary rewards.', *Psychological science : a journal of the American Psychological Society / APS*, 17(3), pp. 222–227. doi: 10.1111/j.1467-9280.2006.01689.x.

Libera, C. Della and Chelazzi, L. (2009) 'Learning to attend and to ignore is a matter of gains and losses', *Psychological Science*, 20(6), pp. 778–784. doi: 10.1111/j.1467-9280.2009.02360.x.

Liljeholm, M. *et al.* (2011) 'Neural correlates of instrumental contingency learning: Differential effects of action-reward conjunction and disjunction', *Journal of Neuroscience*, 31(7), pp. 2474–2480. doi: 10.1523/JNEUROSCI.3354-10.2011.

Liljeholm, M. *et al.* (2013) 'Neural correlates of the divergence of instrumental probability distributions', *Journal of Neuroscience*, 33(30), pp. 12519–12527. doi: 10.1523/JNEUROSCI.1353-13.2013.

Loftus, G. . and Masson, M. E. (1994) 'Using confidence intervals in within-subject designs', *Psychonomic Bulletin & Review*, 1, pp. 476–490. Available at: papers3://publication/uuid/0DFB63AE-D78E-44E2-B870-1341DCA37BBC.

Van Loon, A. M. *et al.* (2016) 'NMDA Receptor Antagonist Ketamine Distorts Object Recognition by Reducing Feedback to Early Visual Cortex', *Cerebral Cortex*, 26(5), pp. 1986–1996. doi: 10.1093/cercor/bhv018.

Lotharius, J. and Brundin, P. (2002) 'Pathogenesis of parkinson's disease: Dopamine, vesicles and α-synuclein', *Nature Reviews Neuroscience*, 3(12), pp. 932–942. doi: 10.1038/nrn983.

Luck, S. J. *et al.* (1989) 'Independent hemispheric attentional systems mediate visual search in split-brain patients.', *Nature*, 342(6249), pp. 543–545. doi: 10.1038/342543a0.

Maia, T. V. and Conceição, V. A. (2017) 'The Roles of Phasic and Tonic Dopamine in Tic Learning and Expression', *Biological Psychiatry*. Elsevier Inc, 82(6), pp. 401–412. doi: 10.1016/j.biopsych.2017.05.025.

Maljkovic, V. and Nakayama, K. (1994) 'Priming ofpop-out: I. Role offeatures', *Memory & Cognition*, 22(6), pp. 657–672.

Maril, S. *et al.* (2013) 'Effects of asymmetric dopamine depletion on sensitivity to rewarding and aversive stimuli in Parkinson's disease.', *Neuropsychologia*. Elsevier, 51(5), pp. 818–24. doi: 10.1016/j.neuropsychologia.2013.02.003.

Marsman, M. and Wagenmakers, E. J. (2017) 'Three Insights from a Bayesian Interpretation of the One-Sided P Value', *Educational and Psychological Measurement*, 77(3), pp. 529–539. doi: 10.1177/0013164416669201.

Martinez-Conde, S., Macknik, S. L. and Hubel, D. H. (2004) 'The role of fixational eye movements in visual perception.', *Nature reviews. Neuroscience*, 5(March), pp. 229–240. doi: 10.1038/nrn1348.

Martinez-Conde, S., Otero-Millan, J. and Macknik, S. L. (2013) 'The impact of microsaccades on vision: towards a unified theory of saccadic function.', *Nature reviews. Neuroscience*. Nature Publishing Group, 14(2), pp. 83–96. doi: 10.1038/nrn3405.

Mathar, D. *et al.* (2017) 'The role of dopamine in positive and negative prediction error utilization during incidental learning – Insights from Positron Emission Tomography, Parkinson's disease and Huntington's disease', *Cortex*. Elsevier Ltd, 90, pp. 149–162. doi: 10.1016/j.cortex.2016.09.004.

Mathôt, S., Schreij, D. and Theeuwes, J. (2012) 'OpenSesame: An open-source, graphical experiment builder for the social sciences', *Behavior Research Methods*, 44(2), pp. 314–324. doi: 10.3758/s13428-011-0168-7.

Maunsell, J. H. R. (2004) 'Neuronal representations of cognitive state: reward or attention?', *Trends in cognitive sciences*, 8(6), pp. 261–5. doi: 10.1016/j.tics.2004.04.003.

Maunsell, J. H. R. and Treue, S. (2006) 'Feature-based attention in visual cortex', *Trends in Neurosciences*. Elsevier Ltd, 29(6), pp. 317–322. doi: 10.1016/j.tins.2006.04.001.

McClure, S. M., Berns, G. S. and Montague, P. R. (2003) 'Temporal prediction errors in a passive learning task activate human striatum', *Neuron*, 38(2), pp. 339–346. doi: 10.1016/S0896-6273(03)00154-5.

McCoy, B. *et al.* (2019) 'Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease', *Brain*, pp. 1–16. doi: 10.1093/brain/awz276.

McCoy, B. and Theeuwes, J. (2016) 'Effects of reward on oculomotor control', *Journal of neurophysiology*, p. jn.00498.2016. doi: 10.1152/jn.00498.2016.

McPeek, R. M., Maljkovic, V. and Nakayama, K. (1999) 'Saccades require focal attention and are facilitated by a short-term memory system', *Vision Research*, 39(8), pp. 1555–1566. doi: 10.1016/S0042-6989(98)00228-4.

Meeter, M. and Van der Stigchel, S. (2014) 'Visual priming through a boost of the target signal: Evidence from saccadic landing positions', *Attention, Perception, and Psychophysics*, 76(2), p. 643. doi: 10.3758/s13414-013-0622-y.

Meeter, M., Van Der Stigchel, S. and Theeuwes, J. (2010) 'A competitive integration model of exogenous and endogenous eye movements', *Biological Cybernetics*, 102(4), pp. 271–291. doi: 10.1007/s00422-010-0365-y.

Miller, E., Erickson, C. and Desimone, R. (1996) 'Neural mechanisms of visual working memory in prefrontal cortex of the macaque', *Journal of Neuroscience*, 16, pp. 5154–5167.

Milstein, D. M. and Dorris, M. C. (2007) 'The influence of expected value on saccadic preparation.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(18), pp. 4810–8. doi: 10.1523/JNEUROSCI.0577-07.2007.

Mir, P. *et al.* (2011) 'Motivation and movement: the effect of monetary incentive on performance', *Experimental Brain Research*, 209, pp. 551–559.

Montagnini, A. and Castet, E. (2007) 'Spatiotemporal dynamics of visual attention during saccade preparation: Independence and coupling between attention and movement planning', *Journal of Vision*, 7(14), pp. 1–16. doi: 10.1167/7.14.8.

Montague, P. R., Dayan, P. and Sejnowski, T. J. (1996) 'A framework for mesencephalic dopamine systems based on predictive Hebbian learning.', *Journal of neuroscience*, 16(5), pp. 1936–1947. doi: 10.1.1.156.635.

Moore, T. and Fallah, M. (2001) 'Control of eye movements and spatial attention.', *Proceedings of the National Academy of Sciences of the United States of America*, 98(3), pp. 1273–1276. doi: 10.1073/pnas.021549498.

Mulckhuyse, M., Van Der Stigchel, S. and Theeuwes, J. (2009) 'Early and late modulation of saccade deviations by target distractor similarity', *Journal of Neurophysiology*, 102(3), pp. 1451–1458. doi: 10.1152/jn.00068.2009.

Mumford, J. a *et al.* (2012) 'Deconvolving BOLD activation in event-related designs for multivoxel pattern classification analyses.', *NeuroImage*. Elsevier Inc., 59(3), pp. 2636–43. doi: 10.1016/j.neuroimage.2011.08.076.

Munoz, D. P. and Everling, S. (2004) 'Look away: the anti-saccade task and the voluntary control of eye movement.', *Nature reviews. Neuroscience*, 5(3), pp. 218–28. doi: 10.1038/nrn1345.

Munoz, D. P., Pelisson, D. and Guitton, D. (1991) 'Movement of neural activity on the superior colliculus motor map during gaze shifts', *Science*, 251(4999), pp. 1358–1360.

Munoz, D. P. and Wurtz, R. H. (1993a) 'Fixation cells in monkey superior colliculus. I. Characteristics of cell discharge', *Journal of Neurophysiology*, 70, pp. 559–575.

Munoz, D. P. and Wurtz, R. H. (1993b) 'Fixation cells in superior colliculus. II. Reversible activation and deactivation', *Journal of neurophysiology*, 70, pp. 576–589.

Mur, M., Bandettini, P. A. and Kriegeskorte, N. (2009) 'Revealing representational content with pattern-information fMRI - An introductory guide', *Social Cognitive and Affective Neuroscience*, 4(1), pp. 101–109. doi: 10.1093/scan/nsn044.

Nakahara, H., Amari, S. and Hikosaka, O. (2002) 'Self-organization in the basal ganglia with modulation of reinforcement signals', *Neural Computation*, 14(4), pp. 819–844.

Nakamura, K. and Hikosaka, O. (2006a) 'Role of dopamine in the primate caudate nucleus in reward modulation of saccades', *Journal of Neuroscience*, 26, pp. 5360–5369.

Nakamura, K. and Hikosaka, O. (2006b) 'Role of Dopamine in the Primate Caudate Nucleus in Reward Modulation of Saccades', *Journal of Neuroscience*, 26(20), pp. 5360–5369. doi: 10.1523/JNEUROSCI.4853-05.2006.

von Neumann, J. and Morgenstern, O. (1944) *The Theory of Games and Economic Behavior*. Princeton: Princeton University Press.

Nickl-Jockschat, T. *et al.* (2012) 'Brain Structure Anomalies in Autism Spectrum Disorder—A Meta-Analysis of VBM Studies Using Anatomic Likelihood Estimation', *Human Brain Mapping*, 33, pp. 1470–1489. doi: 10.1016/j.physbeh.2017.03.040.

Niv, Y. (2009) 'Reinforcement learning in the brain', *Journal of Mathematical Psychology*. Elsevier Inc., 53(3), pp. 139–154. doi: 10.1016/j.jmp.2008.12.005.

Nobre, A. C. *et al.* (1997) 'Functional localization of the system for visuospatial attention using positron emission tomography', *Brain*, 120(3), pp. 515–533. doi: 10.1093/brain/120.3.515.

Noudoost, B. *et al.* (2010) 'Top-down control of visual attention', *Current Opinion in Neurobiology*. Elsevier Ltd, 20(2), pp. 183–190. doi: 10.1016/j.conb.2010.02.003.

Noudoost, B. and Moore, T. (2011) 'Control of visual cortical signals by prefrontal dopamine', *Nature*. Nature Publishing Group, 474(7351), pp. 372–375. doi: 10.1038/nature09995.

Nummenmaa, L. and Hietanen, J. K. (2006) 'Gaze distractors influence saccadic curvature: Evidence for the role of the oculomotor system in gaze-cued orienting', *Vision Research*, 46(21), pp. 3674–3680. doi: 10.1016/j.visres.2006.06.004.

O'Callaghan, C. *et al.* (2013) 'Fronto-striatal gray matter contributions to discrimination learning in Parkinson's disease', *Frontiers in Computational Neuroscience*, 7(December), pp. 1–10. doi: 10.3389/fncom.2013.00180.

O'Craven, K. M. and Kanwisher, N. (2000) 'Mental imagery of faces and places activates corresponding stiimulus-specific brain regions.', *Journal of cognitive neuroscience*, 12(6), pp. 1013–23. Available at: http://www.ncbi.nlm.nih.gov/pubmed/11177421.

O'Doherty, John *et al.* (2003) 'Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 23(21), pp. 7931–7939. doi: 23/21/7931 [pii].

O'Doherty, J *et al.* (2003) 'Temporal Difference Models and Reward-Related Learning in the Human Brain', 28, pp. 329–337. doi: 10.1016/s0896-6273(03)00169-7.

O'Doherty, J. *et al.* (2004) 'Dissociable roles of ventral and dorsal striatum in instrumental conditioning.', *Science (New York, N.Y.)*, 304(2001), pp. 452–454. doi: 10.1126/science.1094285.

O'Doherty, J. P., Cockburn, J. and Pauli, W. M. (2017) 'Learning, Reward, and Decision Making', *Annual Review of Psychology*, 68(1), pp. 73–100. doi: 10.1146/annurev-psych-010416-044216.

O'Keefe, J. and Dostrovsky, J. (1971) 'The hippocampus as a spatial map. Preliminary evidence from unit activity in the freely-moving rat', *Brain Research*, 34, pp. 171–175.

O'Reilly, R. C. and Frank, M. J. (2006) 'Making working memory work: A computational model of learning in the prefrontal cortex and basal ganglia', *Neural Computation*, 18(2), pp. 283–328. doi: 10.1162/089976606775093909.

Ogawa, S. *et al.* (1990) 'Functional brain mapping by blood oxygenation level dependent contrast magnetic resonance imaging: a comparison of signal characters with a biophysical model', *Proceedings of the National Academy of Sciences*, 87, pp. 9868–9872.

Olshausen, B. A., Anderson, C. H. and Van Essen, D. C. (1993) 'A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information', *Journal of Neuroscience*, 13(11), pp. 4700–4719. doi: 10.1523/jneurosci.13-11-04700.1993.

Opris, I., Lebedev, M. and Nelson, R. (2011) 'Motor planning under unpredictable reward: modulations of movement vigor and primate striatum activity', *Frontiers in neuroscience*, 5, p. 61.

Ostlund, S. B. and Balleine, B. W. (2007) 'The Contribution of Orbitofrontal Cortex to Action Selection', *Annals of the New York Academy of Sciences*, 1121(1), pp. 174–192.

Otero-Millan, J. *et al.* (2008) 'Saccades and microsaccades during visual fixation, exploration, and search: foundations for a common saccadic generator.', *Journal of vision*, 8(2008), pp. 21.1–18. doi: 10.1167/9.8.447.

Ott, T. and Nieder, A. (2019) 'Dopamine and cognitive control in prefrontal cortex.', *Trends in Cognitive Sciences*. Elsevier Ltd, 23(3), pp. 213–234. doi: 10.1016/j.tics.2018.12.006.

den Ouden, H. E. *et al.* (2013) 'Dissociable effects of dopamine and serotonin on reversal learning', *Neuron*, 80(4), pp. 1090–100. doi: 10.1016/j.neuron.2013.08.030.

Owen, A. M. *et al.* (1992) 'Fronto-striatal cognitive deficits at different stages of parkinson's disease', *Brain*, 115(6), pp. 1727–1751. doi: 10.1093/brain/115.6.1727.

Packard, M. G., Hirsh, R. and White, N. M. (1989) 'Differential effects of fornix and caudate nucleus lesions on two radial maze tasks: evidence for multiple memory systems', *J Neurosci*, 9(5), pp. 1465–1472. doi: 0270-6474/89/051465-08.

Packard, M. G. and Knowlton, B. J. (2002) 'Learning and memory functions of the Basal Ganglia.', *Annual review of neuroscience*, 25, pp. 563–93. doi: 10.1146/annurev.neuro.25.112701.142937.

Palminteri, S. *et al.* (2009) 'Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes', *Proceedings of the National Academy of Sciences*, 106(45), pp. 19179–19184. doi: 10.1073/pnas.0904035106.

Panksepp, J. (2011) 'The basic emotional circuits of mammalian brains: Do animals have affective lives?', *Neuroscience & Biobehavioral Reviews*, 35(9), pp. 1791–1804.

Pauli, W. M., Nili, A. N. and Tyszka, J. M. (2018) 'A high-resolution probabilistic in vivo atlas of human subcortical brain nuclei', *Scientific Data*, pp. 1–13. doi: 10.1038/sdata.2018.63.

Pavlov, I. (1927) *Conditioned Reflexes: An Investigation ofthe Physiological Activity ofthe Cerebral Cortex*. London: Oxford University Press.

Pedersen, M. L., Frank, M. J. and Biele, G. (2016) 'The drift diffusion model as the choice rule in reinforcement learning', *Psychonomic Bulletin & Review*. Psychonomic Bulletin & Review. doi: 10.3758/s13423-016-1199-y.

Le Pelley, M. E. *et al.* (2015) 'When goals conflict with values: Counterproductive attentional and oculomotor capture by reward-related stimuli.', *Journal of Experimental Psychology: General*, 144(1), pp. 158–171. doi: 10.1037/xge0000037.

Perugini, A. *et al.* (2018) 'Paradoxical Decision-Making: A Framework for Understanding Cognition in Parkinson's Disease', *Trends in Neurosciences*, 41(8). doi: 10.1016/j.tins.2018.04.006.

Perugini, A., Ditterich, J. and Basso, M. A. (2016) 'Patients with Parkinson's Disease Show Impaired Use of Priors in Conditions of Sensory Uncertainty', *Current Biology*. Elsevier Ltd., 26(14), pp. 1902–1910. doi: 10.1016/j.cub.2016.05.039.

Pessiglione, M. *et al.* (2006) 'Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans', *Nature*, 442(7106), pp. 1042–1045. doi: 10.1038/nature05051.

Pessoa, L. and Engelmann, J. B. (2010) 'Embedding reward signals into perception and cognition.', *Frontiers in neuroscience*, 4(September). doi: 10.3389/fnins.2010.00017.

Pierrot-Deseilligny, C. *et al.* (1995) 'Cortical control of saccades', *Annals of Neurology*, 37(5), pp. 557–567.

Poldrack, R. A. (2011) 'Inferring mental states from neuroimaging data: From reverse inference to large-scale decoding', *Neuron*, 72(5), pp. 692–697. doi: 10.1038/jid.2014.371.

Posner, M. I. (1978) *Chronometric explorations of mind.* Hillsdale, NJ: Erlbaum.

Posner, M. I. (1980) 'Orienting of attention', *The Quarterly journal of experimental psychology*, 32(1), pp. 3–25. doi: 10.1080/00335558008248231.

Posner, M. I. *et al.* (1985) 'Inhibition of return: Neural basis and function', *Cognitive Neuropsychology*, 2(3), pp. 211–228. doi: 10.1080/02643298508252866.

Power, J. D. *et al.* (2014) 'Methods to detect, characterize, and remove motion artifact in resting state fMRI', *NeuroImage*. Elsevier Inc., 84, pp. 320–341. doi: 10.1016/j.neuroimage.2013.08.048.

Pulcu, E. and Browning, M. (2017) 'Affective bias as a rational response to the statistics of rewards and punishments', *eLife*, 6, pp. 1–15. doi: 10.7554/elife.27879.

R Development Core and Team (2017) 'R: A language and environment for statistical computing'.

Rangel, A., Camerer, C. and Montague, P. R. (2008) 'A framework for studying the neurobiology of value-based decision making', *Nature Reviews Neuroscience*, 9(7), pp. 545–556. doi: 10.1038/nrn2357.

Ratcliff, R. (1978) 'A theory of memory retrieval', *Psychological Review*, 85, pp. 59–108.

Ratcliff, R. and Childers, R. (2015) 'Individual Differences and Fitting Methods for the Two-Choice Diffusion Model of Decision Making', *Decision*. doi: 10.1037/dec0000030.

Ratcliff, R. and McKoon, G. (2008) 'The Diffusion Decision Model: Theory and Data for Two-Choice Decision Tasks', *Neural Computation*, 20(4), pp. 873–922. doi: 10.1162/neco.2008.12-06-420.

Ratcliff, R. and Rouder, J. N. (1998) 'Modeling Response Times for Two-Choice Decisions', *Psychological Science*, 9(5), pp. 347–356. doi: 10.1111/1467-9280.00067.

Ratcliff, R. and Rouder, J. N. (2000) 'A diffusion model account of masking in two-choice letter identification', *Journal of Experimental Psychology: Human Perception and Performance*, 26(1), pp. 127–140. doi: 10.10371/0096-1523.26.1.127.

Ratcliff, R., Van Zandt, T. and McKoon, G. (1999) 'Connectionist and diffusion models of reaction time', *Psychological Review*, 106, pp. 261–300.

Rescorla, R. A. and Wagner, A. . (1972) 'A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement', in *Classical Conditioning II: Current Research and Theory*. Appleton-Century-Crofts, pp. 64–99.

Reynolds, J. H. and Chelazzi, L. (2004) 'Attentional Modulation of Visual Processing', *Annual Review of Neuroscience*, 27(1), pp. 611–647. doi: 10.1146/annurev.neuro.26.041002.131039.

Reynolds, J. H., Chelazzi, L. and Desimone, R. (1999) 'Competitive mechanisms subserve attention in macaque areas V2 and V4', *Journal of Neuroscience*, 19(5), pp. 1736–1753. doi: 10.1523/jneurosci.19-05-01736.1999.

Rizzolatti, G. (1983) 'Mechanisms of selective attention in mammals', in *Advances in vertebrate neuroethology*. London: Plenum Press, pp. 261–297.

Rizzolatti, G. *et al.* (1987) 'Reorienting attention across the horizontal and vertical meridians: Evidence in favor of a premotor theory of attention', *Neuropsychologia*, 25(1), pp. 31–40. doi: 10.1016/0028-3932(87)90041-8.

Rizzolatti, G. and Gallese, V. (1988) 'Mechanisms and theories of spatial neglect', in *Handbook of neuropsychology*. Amsterdam: Elsevier, pp. 223–246.

Rizzolatti, G., Riggio, L. and Sheliga, B. (1994) 'Space and selective attention', in *Attention and performance XV*. Cambridge: MIT Press, pp. 231–265.

Robbins, T. W. and Everitt, B. J. (1996) 'Neurobehavioural Mechanisms of Reward and Motivation', *Current Opinion in Neurobiology*, 6, pp. 228–236.

Robinson, D. A. and Fuchs, A. F. (1969) 'Eye movements evoked by stimulation of frontal eye fields', *Journal of neurophysiology*, 32(5), pp. 637–48.

Robinson, T. E. and Berridge, K. C. (2003) 'Addiction. Annual Review of Psychology', *Annual Review of Psychology*, 54, pp. 25–53.

Rutledge, R. B. *et al.* (2009) 'Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 29(48), pp. 15104–15114. doi: 10.1523/JNEUROSCI.3524-09.2009.

Saan, R. J. and Deelman, B. G. (1986) *De Nieuwe 15-Woordentest (A en B). Een Handleiding.*

Sara, S. J. and Bouret, S. (2012) 'Orienting and Reorienting: The Locus Coeruleus Mediates Cognition through Arousal', *Neuron*. Elsevier Inc., 76(1), pp. 130–141. doi: 10.1016/j.neuron.2012.09.011.

Sato, M. and Hikosaka, O. (2002) 'Role of primate substantia nigra pars reticulata in reward-

oriented saccadic eye movement', *Journal of Neuroscience*, 22(6), pp. 2363–2373. doi: 10.1523/jneurosci.22-06-02363.2002.

Sawaguchi, T. and Goldman-Rakic, P. S. (1991) 'D1 dopamine receptors in pre-frontal cortex: involvement in working memory', *Science*, 251, pp. 947–50.

Scarpazza, C. *et al.* (2015) 'False positive rates in voxel-based morphometrystudies ofthe human brain: should we be worried?', *Neuroscience & Biobehavioral Reviews*, 52, pp. 49–55.

Van Der Schaaf, M. E. *et al.* (2014) 'Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning', *Cerebral Cortex*, 24(3), pp. 633–642. doi: 10.1093/cercor/bhs344.

Schall, J. D. (1991) 'Neuronal activity related to visually guided saccades in the frontal eye fields of rhesus monkeys: comparison with supplementary eye fields', *Journal of neurophysiology*, 66, pp. 559–579.

Schall, J. D. (1995) 'Neural basis of saccade target selection', *Reviews in the Neurosciences*, 6, pp. 63–85.

Schall, J. D. and Thompson, K. G. (1999) 'Neural Selection and Control of Visually Guided Eye Movements'. doi: 10.1146/annurev.neuro.22.1.241.

Schmidt, L. *et al.* (2014) 'Mind matters: placebo enhances reward learning in Parkinson's disease', *Nature Neuroscience*. Nature Publishing Group, 17(October), pp. 1793–1797. doi: 10.1038/nn.3842.

Schoenbaum, G., Roesch, M.R., S. and Stalnaker, T. A. (2006) 'Orbitofrontal cortex, decision-making and drug addiction', *Trends in Neurosciences*, 29, pp. 116–124.

Schönberg, T. *et al.* (2007) 'Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making.', *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(47), pp. 12860–12867. doi: 10.1523/JNEUROSCI.2496-07.2007.

Schuck, N. W. *et al.* (2016) 'Human Orbitofrontal Cortex Represents a Cognitive Map of State Space', *Neuron*. Elsevier Inc., 91(6), pp. 1402–1412. doi: 10.1016/j.neuron.2016.08.019.

Schultz, W. (2006) 'Behavioral theories and the neurophysiology of reward.', *Annual review of psychology*, 57, pp. 87–115. doi: 10.1146/annurev.psych.56.091103.070229.

Schultz, W. (2007) 'Behavioral dopamine signals', *Trends in Neurosciences*, 30(5), pp. 203–210. doi: 10.1016/j.tins.2007.03.007.

Schultz, W., Dayan, P. and Montague, P. R. (1997) 'A Neural Substrate of Prediction and Reward', *Science*, 275, pp. 1593–1600.

Seamans, J. K. and Yang, C. R. (2004) 'The principal features and mech- anisms of dopamine modulation in the prefrontal cortex', *Progress in Neurobiology*, 74, pp. 1–57.

Serences, J. T. *et al.* (2004) 'Preparatory activity in visual cortex indexes distractor suppression during covert spatial orienting', *Journal of Neurophysiology*, 92(6), pp. 3538–3545. doi: 10.1152/jn.00435.2004.

Serences, J. T. (2008) 'Value-Based Modulations in Human Visual Cortex', *Neuron*. Elsevier Ltd, 60(6), pp. 1169–1181. doi: 10.1016/j.neuron.2008.10.051.

Serences, J. T. and Saproo, S. (2010) 'Population Response Profiles in Early Visual Cortex Are Biased in Favor of More Valuable Stimuli', pp. 76–87. doi: 10.1152/jn.01090.2009.

Serences, J. T. and Yantis, S. (2006) 'Selective visual attention and perceptual coherence', *Trends in Cognitive Sciences*, 10(1), pp. 38–45. doi: 10.1016/j.tics.2005.11.008.

Sereno, A. B. *et al.* (2006) 'Disruption of reflexive attention and eye movements in an individual with a collicular lesion', *Journal of Clinical and Experimental Neuropsychology*, 28(1), pp. 145–166.

Sereno, A. B. and Kosslyn, S. M. (1991) 'Discrimination within and between hemifields: A new constraint on theories of attention', *Neuropsychologia*, 29(7), pp. 659–675. doi: 10.1016/0028-3932(91)90100-M.

Seymour, B. *et al.* (2004) 'ID 1627 letters to nature Temporal difference models describe higher-order learning in humans', *Nature*, 429(June), pp. 664–667. doi: 10.1038/nature02636.1.

Sharot, T. and Garrett, N. (2016) 'Forming Beliefs: Why Valence Matters', *Trends in Cognitive Sciences*. Elsevier Ltd, 20(1), pp. 25–33. doi: 10.1016/j.tics.2015.11.002.

Sharp, M. E. *et al.* (2016) 'Dopamine selectively remediates "model-based" reward learning: A computational approach', *Brain*, 139(2), pp. 355–364. doi: 10.1093/brain/awv347.

Shimo, Y. and Hikosaka, O. (2001) 'Role of tonically active neurons in primate caudate in reward-oriented saccadic eye movement.', *The Journal of Neuroscience*, 21(19), pp. 7804–7814. doi: 21/19/7804 [pii].

Shiner, T. *et al.* (2012) 'Dopamine and performance in a reinforcement learning task: Evidence from Parkinson's disease', *Brain*, 135(6), pp. 1871–1883. doi: 10.1093/brain/aws083.

Skinner, B. F. (1938) *The behavior of organisms*. New York: Appleton-Century-Crofts.

Van Slooten, J. C. *et al.* (2018) 'How pupil responses track value-based decision-making during and after reinforcement learning', *PLoS Computational Biology*, 14(11), pp. 1–24. doi: 10.1371/journal.pcbi.1006632.

Van Slooten, J. C., Jahfari, S. and Theeuwes, J. (2019) 'Spontaneous eye blink rate predicts individual differences in exploration and exploitation during reinforcement learning', *BioRxiv*.

Smit, A. C. and Van Gisbergen, J. A. M. (1990) 'An analysis of curvature in fast and slow human saccades', *Experimental Brain Research*, 81(2), pp. 335–345. doi: 10.1007/BF00228124.

Smith, D. T. and Schenk, T. (2012) 'The Premotor theory of attention: Time to move on?', *Neuropsychologia*. Elsevier Ltd, 50(6), pp. 1104–1114. doi: 10.1016/j.neuropsychologia.2012.01.025.

Smith, P. L. and Ratcliff, R. (2004) 'Psychology and neurobiology of simple decisions', *Trends in Neurosciences*, 27(3), pp. 161–168. doi: 10.1016/j.tins.2004.01.006.

Smith, P. L. and Vickers, D. (1988) 'The accumulator model of two-choice discrimination', *Journal of Mathematical Psychology*, 32(2), pp. 135–168. doi: 10.1016/0022-2496(88)90043-0.

Smittenaar, P. *et al.* (2012) 'Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection - learning or performance?', *European Journal of Neuroscience*, 35(7), pp. 1144–1151. doi: 10.1111/j.1460-9568.2012.08043.x.

Sokol-Hessner, P. *et al.* (2009) 'Thinking like a trader selectively reduces individuals' loss aversion', *Proceedings of the National Academy of Sciences of the United States of America*, 106(13), pp. 5035–5040. doi: 10.1073/pnas.0806761106.

Soto, D. *et al.* (2008) 'Automatic guidance of attention from working memory', *Trends in Cognitive Sciences*, 12(9), pp. 342–348.

Soto, D. and Blanco, M. J. (2004) 'Spatial attention and object-based attention: a comparison within a single task', *Vision Research*, 44(1), pp. 69–81. doi: 10.1016/j.visres.2003.08.013.

Sparks, D. L. and Mays, L. E. (1981) 'he role of the superior colliculus in the control of saccadic eye movements: a current perspective', in *Progress in Oculomotor Research*. Amsterdam: Elsevier, pp. 137–144.

Squire, L. R. *et al.* (1992) 'Activation of the hippocampus in normal humans: a functional anatomical study of memory', *Proceedings of the National Academy of Sciences*, 89, pp. 1837–41.

Stan Development Team (2014) 'RStan: The R interface to Stan (Version 2.17.0)'. Available at: http://mc-stan.org.

Stan Development Team (2015) 'Stan Modeling Language User ' s Guide and Reference Manual (v. 2.6.2)'.

Stankevich, B. a and Geng, J. J. (2014) 'Reward associations and spatial probabilities produce additive effects on attentional selection.', *Attention, perception & psychophysics*. doi: 10.3758/s13414-014-0720-5.

Stanton, G. B., Bruce, C. J. and Goldberg, M. E. (1993) 'Topography of projections to the frontal lobe from the macaque frontal eye fields', *Journal of Computational Neurology*, 330(2), pp. 286–301.

Steiner, J. E. *et al.* (2001) 'Comparative expression of hedonic impact: Affective reactions to taste by human infants and other primates', *Neuroscience and Biobehavioral Reviews*, 25(1), pp. 53–74. doi: 10.1016/S0149-7634(00)00051-8.

Steingroever, H., Wetzels, R. and Wagenmakers, E. J. (2013) 'Validating the PVL-Delta model for the Iowa gambling task', *Frontiers in Psychology*, 4(DEC), pp. 1–17. doi: 10.3389/fpsyg.2013.00898.

van der Stigchel, S. and Theeuwes, J. (2005) 'Our eyes deviate away from a location where a distractor is expected to appear', *Experimental Brain Research*, 169, p. 338.

Stoermer, V. S., Alvarez, G. A. and Cavanagh, P. (2014) 'Within-hemifield competition in early visual areas limits the ability to track multiple objects with attention', *Journal of Neuroscience*, 34(35), pp. 11526–11533. doi: 10.1523/JNEUROSCI.0980-14.2014.

Sugrue, L. P., Corrado, G. S. and Newsome, W. T. (2004) 'Matching behavior and the representation of value in the parietal cortex', *Science*, 304(5678), pp. 1782–1787. doi: 10.1126/science.1094765.

Sugrue, L. P., Corrado, G. S. and Newsome, W. T. (2005) 'Choosing the greater of two goods: neural currencies for valuation and decision making', *Nature Reviews Neuroscience*, 6(5), pp. 363–375. doi:

10.1038/nrn1666.

Suri, R. E. and Schultz, W. (1999) 'A neural network model with dopamine-like reinforcement signal that learns a spatial delayed response task', *Neuroscience*, 91(3), pp. 871–890. doi: 10.1016/S0306-4522(98)00697-6.

Surmeier, D. J. *et al.* (2007) 'D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons', *Trends in Neurosciences*, 30(5), pp. 228–235. doi: 10.1016/j.tins.2007.03.008.

Sutton, R. S. and Barto, A. . (1998) *Reinforcement Learning: An Introduction*. Cambridge, MA: MIT Press.

Swainson, R. *et al.* (2000) 'Probablisitic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: Possible adverse effects of dopaminergic medication', *Neuropsychologia*, 38(5), pp. 596–612.

Swanson, L. W. (1982) 'The Projections of the Ventral Tegmental Area and Adjacent Regions: A Combined Flourescent Retrograde Tracer and Immunofluorescence Study in the Rat', *Brain Res.Bull.*, 9, pp. 321–353.

Takikawa, Y. *et al.* (2002) 'Modulation of saccadic eye movements by predicted reward outcome', *Experimental Brain Research*, 142(2), pp. 284–291. doi: 10.1007/s00221-001-0928-1.

Talluri, B. C. *et al.* (2018) 'Confirmation Bias through Selective Overweighting of Choice-Consistent Evidence', *Current Biology*. Elsevier Ltd., 28(19), pp. 3128-3135.e8. doi: 10.1016/j.cub.2018.07.052.

Talsma, D. *et al.* (2010) 'Brain structures involved in visual search in the presence and absence of color singletons', *Journal of Cognitive Neuroscience*, 22(4), pp. 761–774. doi: 10.1162/jocn.2009.21223.

Theeuwes, J. (1991) 'Exogenous and endogenous control of attention: The effect of visual onsets and offsets', *Perception & psychophysics*, 349(1), pp. 83–90. doi: 10.3758/BF03211619.

Theeuwes, J. (1992) 'Perceptual selectivity for color and form', *Perception & Psychophysics*, 51(6), pp. 599–606. doi: 10.3758/BF03211656.

Theeuwes, J. *et al.* (1998) 'Our Eyes do Not Always Go Where we Want Them to Go: Capture of the Eyes by New Objects', *Psychological Science*, 9(5), pp. 379–385. doi: 10.1111/1467-9280.00071.

Theeuwes, J. *et al.* (1999a) 'Influence of attentional capture on oculomotor control', *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), pp. 1595–1608. doi: 10.1037/0096-1523.25.6.1595.

Theeuwes, J. *et al.* (1999b) 'Influence of attentional capture on oculomotor control', *Journal of Experimental Psychology: Human Perception and Performance*, 25(6), pp. 1595–1608. doi: 10.1037/0096-1523.25.6.1595.

Theeuwes, J. (2010) 'Top-down and bottom-up control of visual selection', *Acta Psychologica*. Elsevier B.V., 135(2), pp. 77–99. doi: 10.1016/j.actpsy.2010.02.006.

Theeuwes, J. (2018) 'Visual Selection: Usually Fast and Automatic; Seldom Slow and Volitional; A Reply to Commentaries', *Journal of Cognition*, 1(1), pp. 1–15. doi: 10.5334/joc.32.

Theeuwes, J. and Belopolsky, A. V. (2012) 'Reward grabs the eye: Oculomotor capture by rewarding stimuli', *Vision Research*. Elsevier Ltd, 74, pp. 80–85. doi: 10.1016/j.visres.2012.07.024.

Theeuwes, J. and Burger, R. (1998a) 'Attentional Control during Visual Search: The Effect of Irrelevant Singletons', *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), pp. 1342–1353. doi: 10.1037/0096-1523.24.5.1342.

Theeuwes, J. and Burger, R. (1998b) 'Attentional Control during Visual Search: The Effect of Irrelevant Singletons', *Journal of Experimental Psychology: Human Perception and Performance*, 24(5), pp. 1342–1353. doi: 10.1037/0096-1523.24.5.1342.

Thompson, K. G. and Bichot, N. P. (2005) 'A visual salience map in the primate frontal eye field', *Progress in brain research*, 147, pp. 251–62.

Thomson, D. R. and Milliken, B. (2012) 'Revisiting the time course of inter-trial feature priming in singleton search', *Psychological Research*, 77(5), pp. 637–650. doi: 10.1007/s00426-012-0455-7.

Thomson, D. R. and Milliken, B. (2013) 'Contextual distinctiveness produces long-Lasting priming of pop-out', *Journal of Experimental Psychology: Human Perception and Performance*, 39(1), pp. 201–215. doi: 10.1037/a0028069.

Thorndike, E. L. (1911) *Animal intelligence*. New York: Macmillian.

Tomlinson, C. L. *et al.* (2010) 'Systematic review of levodopa dose equivalency reporting in Parkinson's disease', *Movement Disorders*, 25(15), pp. 2649–2653. doi: 10.1002/mds.23429.

Tommasi, G. *et al.* (2015) 'Disentangling the Role of Cortico-Basal Ganglia Loops in Top–Down and Bottom–Up Visual Attention: An Investigation', *Journal of Cognitive Neuroscience*, 27(6), pp. 1215–1237. doi: 10.1162/jocn.

Torralbo, A. and Beck, D. M. (2008) 'Perceptual-load-induced selection as a result of local competitive interactions in visual cortex', *Psychological Science*, 19(10), pp. 1045–1050. doi: 10.1111/j.1467-9280.2008.02197.x.

Trappenberg, T. P. *et al.* (2001) 'A model of saccade initiation based on the competitive integration of exogenous and endogenous signals in the superior colliculus', *Journal of cognitive neuroscience*, 13(2), pp. 256–271. doi: 10.1162/089892901564306.

Treiber, J. M. *et al.* (2016) 'Characterization and correction of geometric distortions in 814 Diffusion Weighted Images', *PLoS ONE*, 11(3), pp. 1–9. doi: 10.1371/journal.pone.0152472.

Treisman, A. (1998) 'Feature binding, attention and object perception.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 353(1373), pp. 1295–306. doi: 10.1098/rstb.1998.0284.

Treisman, A. and Gelade, G. (1980) 'A Feature-Integration of Attention', *Cognitive Psychology*, 12, pp. 97–136.

Tustison, N. J. *et al.* (2010) 'N4ITK: Improved N3 bias correction', *IEEE Transactions on Medical Imaging*, 29(6), pp. 1310–1320. doi: 10.1109/TMI.2010.2046908.

Verharen, J. P. H. *et al.* (2018) 'A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states', *Nature Communications*. Springer US, 9(1), pp. 1–15. doi:

10.1038/s41467-018-03087-1.

Voon, V. *et al.* (2010) 'Mechanisms Underlying Dopamine-Mediated Reward Bias in Compulsive Behaviors', *Neuron*. Elsevier Inc., 65(1), pp. 135–142. doi: 10.1016/j.neuron.2009.12.027.

Walker, R. *et al.* (1997) 'Effect of Remote Distractors on Saccade Programming : Evidence for an Extended Fixation Zone', *Journal of neurophysiology*, pp. 1108–1119.

Wang, S. *et al.* (2017) 'Evaluation of Field Map and Nonlinear Registration Methods for Correction of Susceptibility Artifacts in Diffusion MRI', *Frontiers in Neuroinformatics*, 11(February), pp. 1–9. doi: 10.3389/fninf.2017.00017.

Watkins, C. J. C. H. (1989) *Learning from delayed rewards*. Cambridge, England.

Watkins, C. J. C. H. and Dayan, P. (1992) 'Q Learning', *Machine Learning*, 8, pp. 279–292. doi: 10.1023/A:1022676722315.

Wetzels, R. *et al.* (2010) 'Bayesian parameter estimation in the Expectancy Valence model of the Iowa gambling task', *Journal of Mathematical Psychology*. Elsevier Inc., 54(1), pp. 14–27. doi: 10.1016/j.jmp.2008.12.001.

Wiecki, T. V., Sofer, I. and Frank, M. J. (2013) 'HDDM: Hierarchical Bayesian estimation of the Drift-Diffusion Model in Python', *Frontiers in Neuroinformatics*, 7(August), p. 14. doi: 10.3389/fninf.2013.00014.

Wolfe, J. M. (1994) 'Guided Search 2.0: A revised model of visual search JEREMY', *Psychonomic Bulletin & Review*, 1(2), pp. 202–238.

Wolosin, S. M., Zeithamova, D. and Preston, A. . (2012) 'Reward modulation of hippocampal subfield activation during successful associative encoding and retrieval', *Journal of Cognitive Neuroscience*, 24, pp. 1532–1547.

Worsley, K. J. and Friston, K. J. (1995) 'Analysis of fMRI time-series revisited – again', *Neuroimage*, 2, pp. 173–81.

Wunderlich, K., Smittenaar, P. and Dolan, R. J. (2012) 'Dopamine Enhances Model-Based over Model-Free Choice Behavior', *Neuron*, 75, pp. 418–424. doi: 10.1016/j.neuron.2012.03.042.

Wurtz, R. H. (2015) 'Using perturbations to identify the brain circuits underlying active vision', *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370(1677). doi: 10.1098/rstb.2014.0205.

Wylie, S. A. *et al.* (2009) 'The effect of speed-accuracy strategy on response interference control in Parkinson's disease', *Neuropsychologia*, 47(8–9), pp. 1844–1853. doi: 10.1016/j.neuropsychologia.2009.02.025.

Wyvell, C. L. and Berridge, K. C. (2000) 'Intra-Accumbens Amphetamine Increases the Conditioned Incentive Salience of Sucrose Reward: Enhancement of Reward "Wanting" without Enhanced "Liking" or Response Reinforcement', *The Journal of Neuroscience*, 20(21), pp. 8122–8130. doi: 10.1523/JNEUROSCI.20-21-08122.2000.

Yamamoto, S. *et al.* (2012) 'What and Where Information in the Caudate Tail Guides Saccades to Visual Objects', *Journal of Neuroscience*, 32(32), pp. 11005–11016. doi: 10.1523/JNEUROSCI.0828-

12.2012.

Yamamoto, S., Kim, H. F. and Hikosaka, O. (2013) 'Reward Value-Contingent Changes of Visual Responses in the Primate Caudate Tail Associated with a Visuomotor Skill', *Journal of Neuroscience*, 33(27), pp. 11227–11238. doi: 10.1523/JNEUROSCI.0318-13.2013.

Yasuda, M., Yamamoto, S. and Hikosaka, O. (2012) 'Robust representation of stable object values in the oculomotor basal ganglia', *Journal of Neuroscience*, 32(47), pp. 16917–16932. doi: 10.1523/JNEUROSCI.3438-12.2012.

Yin, H. H., Knowlton, B. J. and Balleine, B. W. (2004) 'Lesions of dorsolateral striatum preserve outcome expectancy but disrupt habit formation in instrumental learning', *European Journal of Neuroscience*, 19, pp. 181–189.

Yin, H. H., Ostlund, S. B. and Balleine, B. W. (2005) 'The role of the dorsomedial striatum in instrumental conditioning', *European Journal of Neuroscience*, 28(8), pp. 513–523. doi: 10.1111/j.1460-9568.2008.06422.x.

Yoon, J. H., Curtis, C. E. and D'Esposito, M. (2006) 'Differential effects of distraction during working memory on delay-period activity in the pre-frontal cortex and the visual association cortex', *Neuroimage*, 29, pp. 1117–1126.

Yu, A. J. and Dayan, P. (2005) 'Uncertainty, neuromodulation, and attention', *Neuron*, 46(4), pp. 681–692. doi: 10.1016/j.neuron.2005.04.026.

Zaldivar, D. *et al.* (2014) 'Dopamine-induced dissociation of BOLD and neural activity in macaque visual cortex', *Current Biology*. Elsevier Ltd, 24(23), pp. 2805–2811. doi: 10.1016/j.cub.2014.10.006.

Zhang, Y., Brady, M. and Smith, S. (2001) 'Segmentation of brain MR images through a hidden Markov random field model and the expectation-maximization algorithm', *IEEE Transactions on Medical Imaging*, 20(1), pp. 45–57. doi: 10.1109/42.906424.

Zoest, W. Van, Donk, M. and Theeuwes, J. (2004) 'The Role of Stimulus-Driven and Goal-Driven Control in Saccadic Visual Selection', *Journal of experimental psychology. Human perception and performance*, 30(4), pp. 746–759. doi: 10.1037/0096-1523.30.4.746.

List of Publications

- **McCoy, B.**, Jahfari, S., Engels, G., Knapen, T.*, & Theeuwes, J.* (2019). Dopaminergic medication reduces striatal sensitivity to negative outcomes in Parkinson's disease. *Brain*, 142(11), 3605–3620. https://doi.org/ 10.1093/brain/awz276.

- Engels, G., A. Vlaar, **McCoy, B.**, Scherder, E., & Douw, L. (2018). Dynamic functional connectivity and symptoms of Parkinson's disease: a resting-state fMRI study. *Frontiers in Aging Neuroscience*, 10:388. https://doi.org/10.3389/fnagi.2018.00388.

- Engels, G., **McCoy, B.**, Vlaar, A., Theeuwes, J., Weinstein, H., Scherder, E. (2018). *Journal of Neural Transmission*, 125, 1449-1459. https://doi.org/10.1007/s00702-018-1916-y.

- **McCoy, B.** & Theeuwes, J. (2018). Overt and covert attention to location-based reward. *Vision Research*, 142, 27-39. https://doi.org/10.1016/j.visres.2017.10.003.

- **McCoy, B.** & Theeuwes, J. (2016). Effects of reward on oculomotor control. Journal of Neurophysiology, 116(5), 2453-2466. https://doi.org/10.1152/jn.00498.2016.

In Preparation

- **McCoy, B.** & Theeuwes, J. (2020). Distractor inhibition and reinforcement learning in Parkinson's disease.

Acknowledgements

There are many people who made my years at the VU an incredibly enjoyable and positive experience. The people and culture have shaped who I am, and now that I am back in my "almost" native culture, I hold these experiences close to me. Upon finishing my PhD position, I had spent seven years of my life in the Netherlands, a country I'm proud to view as my second home. There are many transitionary periods in our lives, with us millenials being particularly slow to grow up, but looking back, I can see that my PhD training truly delivered me as a fully-equipped adult, with all of the understanding and wisdom that entails. Not least of which is the knowledge that there is always more to learn!

I am writing this under exceptional circumstances, with the entire globe undergoing an unprecedented pandemic. For this reason, my PhD defense will take place as a video call. After attending very many of my colleague's defenses, and imagining my own, this really is quite an extraordinary experience.

Jan, thank you for giving me such a great opportunity. I have always appreciated your direct and honest approach. You have been nothing but kind to me throughout my whole PhD. You immediately accepted me as a person and not just an employee, and quickly understood my need for independence. You were constantly responsive and on top of things; these are admirable qualities. I will always be grateful to you for trusting that I would come to you when I needed to, and knowing your door was always open has meant a lot to me. I wish you much happiness in the rest of your (busy!) career and in life.

Tomas, thank you for being my mentor in science. I learned so much from you, in terms of both the science itself but also in how we approach it, which is (at least!) half the battle. You are a true geek who pays attention to every detail! Motivating others with your enthusiasm, but also showing us that enthusiasm is not enough; if we want to get close to the truth, we must be dually meticulous and able to think about the big picture, which forces us to think many steps ahead. You are a beautiful writer - your way with words has helped reignite my own appreciation of the magic that lies in that. Although most of us will never achieve your prowess, you truly are an inspiration to us to always keep trying.

I would like to thank the members of my committee for taking the time and effort to read my thesis. You are the perfect mixture of my old and new life, from my days living in Nijmegen, Utrecht, and Amsterdam to my new home here in Cambridge. I look forward to discussions with you all at the defense.

Gwendie, my partner in crime. We spent so many weekends in that scanning room. We stressed out so much together, and you put up with me during my most rigid, nit-picky moments; not only has our friendship survived this, but it has thrived! Your creative and flexible approach to life was always a breath of fresh air to me. Thank you for your ridiculous jokes, your music, your comfort, and your sofa-bed! You are truly a friend for life.

Sara, thank you for your supervision and clear vision for the Parkinson's project. I learned a lot from you about the scientific method, from the analysis stages to a publication-ready manuscript.

You were always motivated by the project and played an important role in the quality of the final output. I wish you the best in your new ventures outside of academia!

To the entire VU crew, who I shared so many fun and silly conversations with over Friday drinks at the red table. Thank you Daniel, Lisette, Eduard, Michel, Wouter, Joshua, Benchi, Berno, Katya (also for the fun choir times!), Joanne, Anouk, Ingmar, Kiki, Iliana, Jessica, Paul, Nicki, Jeroen, Dirk, Daan, Tom, Martin, Kim, Sebastiaan, Daniel Schreij, Sylco, Jan, Chris, Sander, Richard, Erik, and Mieke. Thanks to Elle and Inês for all the light-heartedness and for making the VU still like home whenever I return. Thanks to Gilles for your level-headed and calming influence, whether it was about fMRI deconvolution methods or chatting about life. Thank you, Barbara, for always being so helpful and kind. Finally, thank you to all those (many of the above) who joined me for all the NVP frolics, VSS 2016 (and the road-trip!), and the attention workshop in Rovereto (at the very beginning of my PhD, when I had hardly a penny to my name). Thanks to my book club peeps, particularly the staples: Kayta, Yannick, Eduard, Daniel, Elle, and Kim. I enjoyed the dinners, diversions, and each time finding out what percentage of the book people actually read and having them relay the story and their opinions accordingly.

Eduard, I am incredibly glad to have had someone as fun, adventurous and spontaneous as you in my life during my PhD, who somehow also managed to be one of the most dedicated people I've ever met, in everything you did. I think the story of you picking up an old 80's family portrait by the kerb-side and forcing me to bring it home best sums you up. Always attentive, random, and determined that nothing should go to waste - everything can be repurposed. Thank you for all the life lessons you taught me.

Lisette, I am very wistful for our long runs in Amsterdamse Bos. You are my favourite running partner! So many chats; if trees had ears, they would have heard our life stories. I loved every minute of you describing your latest training and race improvements. You are a strong and determined person, and always made me feel motivated. Running aside, thank you for the many dinners and fun nights out we shared; there will be more to come.

I am grateful to my old Nijmegen friends, my first Dutch family, who have given me a sense of home and stability in NL. In particular, thanks to Suus, for all the chats, comfort, and for always telling the truth – even when it is hard to hear. Thanks to Matthias, who started my PhD journey with me; I learned a lot from you, in all aspects of life.

I would also like to thank all my new friends and colleagues in Cambridge, for helping to get me over the final hurdle: Áine, Alex, Petra, Edwin, Kanad, Max, Helen, Reuben, Lizzie, Becky, Ellie, and Sandy. I am truly happy to have every one of you in my life and look forward to all the laughs and research that we'll share together in the future.

Thanks to my oldest friends Mossy, Brendan and Vanessa, who will always be a part of me.

Thank you, Ronan, for being one of the loveliest and most understanding people I've ever known. We could analyze our (and everyone else's) lives until the cows come home. How lucky am I that you're stuck with me as family!

Anthony, you're the only person in here where I think words are not even necessary. But I'll put down a few! Sometimes I've thought that we are like twins; it would explain my unbreakable connection to you even when we don't speak for months on end. Our shared background, both familial and in the maths/physics/AI/philosophy/random ideas sense, means that you are simply with me in everything I do and achieve. Thank you for all the chats and laughs over the years, and also for having such fun kids! MC, thank you for being lovely to me, and for always so willingly accepting me into your home. Breedge, thank you for being so warm and friendly towards me since I was a teenager. Your strength, level-headedness, and support, of both me and my father, has meant so much to me.

Dad, thank you for being my rock. You have constantly highlighted the need for me to be kind to myself. That has not always been easy, but you understand that it takes time, and have supported me every step of the way. Thank you for always promoting my self-development and growth and for helping me to see that our work and life selves are the same person. Your acceptance of me, no matter what, has been the greatest comfort to me over the years.

Finally - thank you Mam, a Mhamaí, tá tú très speisialta à moi. Thanks for your unwavering belief in me and for your positive energy. Thank you for your kookiness and ability to adapt to any circumstances; you are forever youthful. I enjoyed you visiting me in all my homes in NL, and am very grateful that you've always wanted to be part of my life. You have always welcomed me home with open arms; this thesis would not have been possible without you.