Visual Attention and Dopamine in Value-Based Learning

McCoy, B.

2020

document version
Publisher's PDF, also known as Version of record

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.