doi:10.1093/brain/aws083 Brain 2012: 135; 1871–1883 | 1871

BRAIN
A JOURNAL OF NEUROLOGY

Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease

Tamara Shiner,[1] Ben Seymour,[1] Klaus Wunderlich,[1] Ciaran Hill,[1] Kailash P. Bhatia,[2] Peter Dayan[3] and Raymond J. Dolan[1]

1 Wellcome Trust Centre for Neuroimaging, UCL, London, WC1N 3BG, UK
2 Sobell Department of Motor Neuroscience and Movement Disorders, UCL, London, WC1N 3BG, UK
3 Gatsby Computational Neuroscience Unit, UCL, London, WC1N 3AR, UK

Correspondence to: Tamara Shiner,
12 Queen Square London,
WCN 3BG, UK
E-mail: tamarashiner@gmail.com

The role dopamine plays in decision-making has important theoretical, empirical and clinical implications. Here, we examined its precise contribution by exploiting the lesion deficit model afforded by Parkinson's disease. We studied patients in a two-stage reinforcement learning task, while they were ON and OFF dopamine replacement medication. Contrary to expectation, we found that dopaminergic drug state (ON or OFF) did not impact learning. Instead, the critical factor was drug state during the performance phase, with patients ON medication choosing correctly significantly more frequently than those OFF medication. This effect was independent of drug state during initial learning and appears to reflect a facilitation of generalization for learnt information. This inference is bolstered by our observation that neural activity in nucleus accumbens and ventromedial prefrontal cortex, measured during simultaneously acquired functional magnetic resonance imaging, represented learnt stimulus values during performance. This effect was expressed solely during the ON state with activity in these regions correlating with better performance. Our data indicate that dopamine modulation of nucleus accumbens and ventromedial prefrontal cortex exerts a specific effect on choice behaviour distinct from pure learning. The findings are in keeping with the substantial other evidence that certain aspects of learning are unaffected by dopamine lesions or depletion, and that dopamine plays a key role in performance that may be distinct from its role in learning.

Keywords: Parkinson's disease; learning; functional MRI; dopamine
Abbreviations: BIC = Bayesian Information Criterion

Introduction

Dopamine is strongly implicated in reward signalling, playing a central role in reward learning in animals (Wise and Rompre, 1989; Schultz *et al*., 1997; Schultz, 1998; Wise, 2004; Bayer and Glimcher, 2005) and humans (Pessiglione *et al*., 2006). Accumulating evidence from pharmacological interventions in healthy subjects (Pessiglione *et al*., 2006) and patients with Parkinson's disease studied ON and OFF medication (Frank *et al*., 2004, 2007*b*) indicate that manipulating dopamine neurotransmission in humans influences reward-related reinforcement learning and decision-making. An assumption arising from these data is that dopamine exerts a direct effect on instrumental learning, a form of learning that links actions and their outcomes. At a mechanistic level, activity in dopaminergic neurons express a prediction error believed to mediate learning and updating the reward value of predictive stimuli (Schultz *et al*., 1997). The idea that prediction

Received August 10, 2011. Revised February 1, 2012. Accepted February 3, 2012. Advance Access publication April 15, 2012

error-based learning is computationally implemented via activity patterns within the dopaminergic system is supported by a substantial body of experimental work across a range of species (Haber and Knutson, 2010).

However, the functions of dopamine are known to extend beyond reinforcement learning. First, considerable evidence points to a contribution to the control of Pavlovian approach behaviour (Ikemoto and Panksepp, 1999; Parkinson *et al*., 2002; Day and Carelli, 2007) as well as in motivational engagement and vigour (Ahlenius *et al*., 1977; Beninger and Phillips, 1981; Berridge and Robinson, 1998; McClure *et al*., 2003; Niv, 2007; Niv *et al*., 2007; Bardgett *et al*., 2009; Lex and Hauber, 2010; Boureau and Dayan, 2011). These influences are distinct from learning (Yin *et al*., 2008), even in cases where they arise from a signal that actually reports a prediction error (McClure *et al*., 2003). Secondly, many aspects of appetitive learning can progress normally in the absence of dopamine, most dramatically in the case of genetically engineered dopamine-deficient mice (Palmiter, 2008). Thirdly, a number of previous studies that investigated the effect of dopamine in humans could not easily distinguish between action learning and action performance (Frank *et al*., 2004, 2007*b*; Pessiglione *et al*., 2006). Consequently, in assessing dopamine's impact on behaviour, it is necessary to distinguish influences on learning from influences on the modulation of the expression of learning, i.e. an effect on actual choice behaviour or performance.

Parkinson's disease is a common neurological disorder characterized by neuronal loss in the substantia nigra (Edwards *et al*., 2008), which leads to depleted levels of striatal dopamine (Koller and Melamed, 2007). Parkinson's disease results in deficits across several cognitive domains, including probabilistic learning and classification tasks (Knowlton *et al*., 1996; Graef *et al*., 2010), with dopamine replacement therapy having distinct effects on these behaviours. For example, when Parkinson's disease patients are OFF dopamine replacement therapy, it is reported that their expression of learning from positive feedback is impaired (Frank *et al*., 2004, 2007*b*), while when ON dopamine replacement therapy they show impaired performance in learning from negative outcomes (Frank *et al*., 2004; Bodi *et al*., 2009). This behavioural pattern has been attributed to increased levels of striatal dopamine when patients are ON medication boosting prediction error signals resulting in enhanced learning from positive outcomes. In contrast, a prevention of dips in dopaminergic activity, as occurs with omission of expected outcomes, is suggested to worsen learning from negative outcomes (Frank *et al*., 2004; Frank, 2007*b*; Maia and Frank, 2011).

Here, we sought to dissociate dopaminergic effects on learning from effects on choice (performance) by acquiring neuroimaging data during a reinforcement learning task in patients with Parkinson's disease. We employed a two-stage learning task that involves separate phases of (i) acquisition and (ii) a subsequent performance testing involving generalization of learning. This task has previously been shown to provide an effective means of examining the neural mechanisms underlying cognitive deficits in Parkinson's disease (Frank *et al*., 2004, 2007*b*). These previous studies focused on learning, while here we also probed the effect of dopaminergic status (ON medication, and OFF medication) on test 'performance', i.e. on the expression of learning during behavioural extinction. Crucially, this dissociation between learning and performance has not been explicitly explored in previous human investigations.

Materials and methods

The study and its procedures were approved by the National Research Ethics Service, The Joint UCL/UCLH Committees on the Ethics of Human Research (Committee A).

Participants

Fourteen early-to-moderate stage (Hoehn and Yahr stage: mean 1.69, SE 0.26) patients with idiopathic Parkinson's disease (10 males) (as per UK Brain Bank criteria) aged between 44 and 81 years (mean 61.8 years, SE 3.3 years) participated in and completed the study. Patients were recruited from the movement disorder clinic at the National Hospital for Neurology and Neurosurgery.

We obtained written informed consent from all subjects and transport costs were reimbursed.

Subjects were interviewed for psychiatric and neurological history as well as current and past medication. They were also examined by a clinician and asked to complete several questionnaires, including a health questionnaire, a Mini-Mental State Examination and an impulse control disorder screening questionnaire (Supplementary Table 1).

One subject had difficulty understanding the task demands and adopted an incorrect strategy for stimulus selection, whereby he explicitly believed the incorrect stimulus to be correct and continued to select it despite ongoing negative feedback resulting in significantly worse than chance performance. Data from this subject are not included in any analyses. Another subject was excluded from the imaging analysis due to an incidental finding of abnormally large ventricles, compromising normalization of this data set to a standard coordinate space. Hence, 13 subjects were analysed behaviourally and 12 subjects were analysed in the functional MRI study.

Twelve of the subjects were right-handed and one was left-handed. All were fluent English speakers. The duration of Parkinson's disease varied from 1 to 10 years from the time of initial diagnosis (mean 4.9 years, SE 0.96 years). Subjects had no history of other major neurological or psychiatric disease. Patients were all on levodopa/carbidopa combinations; eight patients were also on dopamine agonists; total daily dose of levodopa/carbidopa varied from 50/12.5 mg to 1000/255 mg (mean 400/100 mg, SE 74.4/18.6 mg) (Supplementary Table 2). We did not recruit patients on trihexyphenidyl, benzhexol or high-dose tolterodine due to possible confounding effects of high-dose anti-cholinergic medication, or patients on amantadine due to its effect on multiple neuromodulators.

Stimuli

We used a version of the generalization task introduced by Frank *et al*. (2004, 2007*b*). Stimuli consisted of Hiragana symbols presented in white fonts on a black background where each stimulus had a different probability of being correct when selected. These probabilities ranged from 80% to 20%. In the first, or acquisition, stage of the task, the symbols were paired to form three sets: the 80% stimulus was paired with the 20% stimulus; the 70% stimulus was paired with the 30% stimulus and the 60% stimulus with the 40% stimulus. The sets were presented in a randomized order. In the second, or performance phase, along with all the training pairings, the best stimulus (the one with 80% chance of being correct) and the worst stimulus

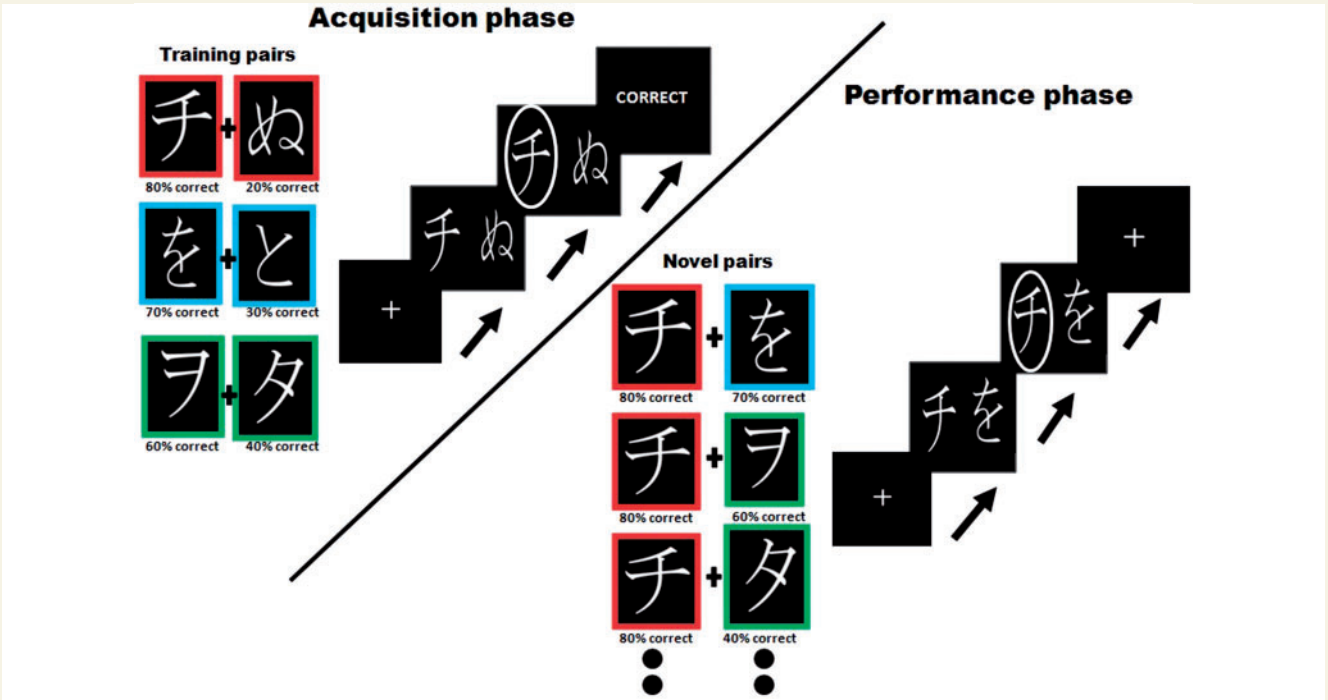

Figure 1 Task. Stimuli consisted of Hiragana symbols which were presented in white fonts on a black background. Each stimulus had a different probability of being correct when selected. In the first, or acquisition stage of the task, symbols were paired to form three 'training pairs' that remained the same throughout this phase: the 80% stimulus was paired with the 20% stimulus (highlighted for illustration purposes with a red border); the 70% stimulus was paired with the 30% stimulus (blue border) and the 60% stimulus with the 40% stimulus (green border). Subjects selected the left or right stimulus by button presses and, during the acquisition phase, also received information about the outcome (correct/incorrect). In the second, or performance phase, along with all the training pairings, the best stimulus (the one with 80% chance of being correct) and the worst stimulus (the one with only 20% chance of being correct) were presented in novel pairings with all the other stimuli. During this phase subjects did not receive information about the outcome of their choice. During this phase, subjects were also presented again with the three sets of 'training pairs', which were interspersed with the novel pairs.

(the one with only 20% chance of being correct) were presented in novel pairings with all the other stimuli (Fig. 1).

Procedure

Overview

Each patient participated in three separate sessions on different days, which were a minimum of 1 week apart (i.e. a within-subject design). Each session involved different Hiragana symbols (Fig. 1). All patients performed the task in three different drug states (Supplementary Table 3): acquisition and performance in the ON state (ON–ON), acquisition and performance in the OFF state (OFF–OFF) and acquisition of the stimulus contingencies in the OFF state but performance in the ON medication state (OFF–ON). The order of the different drug states in which patients performed the task was randomized. The OFF state in two of the conditions was achieved by a minimum of 12 h withdrawal from all dopaminergic medication and omission of all slow release preparations for a minimum of 18 h. On the remaining day (ON–ON), patients were asked to take their morning dopaminergic medication as usual. We were unable to test patients in the ON–OFF state, i.e. acquisition in the ON state and performance in the OFF state, due to the half-life of levodopa/carbidopa combinations, which would require a minimum of 7.5 h to be metabolized and excreted resulting in too long an interval between the acquisition and performance phases. All patients were tested at similar times in the morning to equalize washout times and to control for diurnal symptom fluctuations.

To familiarize subjects with the structure of the task, we undertook a short practice block before the first scanning session. During the practice session, patients worked on an identical task as in the main study except for the fact they were presented with different Hiragana symbols. The main session began with two functional scans (Scans 1 and 2, acquisition sessions). Most subjects completed a third acquisition session on a laptop. In OFF–ON condition, patients took their medication following this training. All patients then waited for 45–60 min before undergoing a third functional scanning session (Scan 3, performance session) for performance testing.

On one of the 3 days, after the training and performance stages were complete, the patients also underwent a structural scan, a Mini-Mental State Examination and completed questionnaires as detailed above.

Acquisition phase of the task: Scans 1 and 2

Scan sessions 1 and 2 (acquisition phase of the task), lasted ~16 min, and consisted of 120 trials of 8 s each. On each trial, two Hiragana characters appeared on the screen side by side, presented via a mirror mounted on the head coil. Subjects' task was to select one of the characters on each trial by pressing either the right or the left key on a button box. The stimuli remained on the screen for 4 s, followed

by presentation of the outcome (either 'correct' or 'incorrect') for 2 s. The likelihood of being correct or incorrect was probabilistically determined for each stimulus (see above). If subjects did not respond within 4 s that the stimuli were on the screen the message 'no key pressed' was presented and the trial was excluded from the analysis. A fixation cross was presented for 2 s during the intertrial interval.

Performance phase of the task: Scan 3

Scanning session 3 (performance phase) was 10-min long and consisted of 110 trials of 6 s each. Similar to the acquisition phase, two Hiragana characters were presented side by side on each trial and subjects had to select one of the characters by pressing either the right or the left key. As before, characters remained on screen for 4 s. This time subjects did not receive feedback after making a response and the trial instead progressed immediately to the presentation of a fixation cross during the 2 s intertrial interval.

Importantly, in addition to the stimulus pairs used during training (80% with 20%; 70% with 30% and 60% with 40%), the symbols were shown in eight novel pairings. Four of the pairings had the 'best' stimulus paired with all other stimuli (80% with 70%; 80% with 60%; 80% with 40% and 80% with 30%), and the other four pairings compared the 'worst' stimulus to all other stimuli (20% with 70%; 20% with 60%; 20% with 40% and 20% with 30%). All pairs were presented 10 times each in randomized order, resulting in 110 pairs overall (Fig. 1).

Magnetic resonance imaging

The study was conducted at the Wellcome Trust Centre for Neuroimaging at University College London using a 3 T (Siemens TRIO) scanner equipped with a Siemens 12-channel phased array head coil. Anatomical images were acquired using modified equilibrium Fourier transform T_1 gradient echo scans, which were followed by 1-mm thick axial slices parallel to the anterior commissure–posterior commissure plane. Functional scans used a gradient echo sequence; repetition time, 2.04 s; echo time 30 ms; flip angle 90°; matrix size 64 × 64; field of view 192 mm; slice thickness, 2 mm. A total of 30 axial slices were sampled. The in-plane resolution was 2 × 2 mm.

Functional imaging data were analysed using statistical parametric mapping software (SPM5; Wellcome Trust Centre for Neuroimaging; http://www.fil.ion.ucl.ac.uk/spm). During preprocessing, images were realigned with the first volume (after discarding six volumes to allow for T_1 equilibration effects), and unwarped. For each subject, the mean functional image was coregistered to a high resolution T_1 structural image. This image was then spatially normalized to standard Montreal Neurological Institute (MNI) space using the 'unified segmentation' algorithm available within SPM5 (Ashburner and Friston, 2005) with the resulting deformation field applied to the functional imaging data. These data were then spatially smoothed using an isotropic 6-mm full-width half-maximum Gaussian kernel.

Data analysis

Behavioural analysis

Acquisition sessions 1–3

All subjects reached at least 65% accuracy for the easiest pairing or after completion of three acquisition sessions had a minimum accuracy of 60% over all training pairs before proceeding to the performance phase. Accuracy levels in the acquisition phase were then separately computed for each drug state by averaging the overall accuracy across all acquisition sessions on that day. Accuracy was defined as percentage of trials on which the correct stimulus, i.e. the stimulus with the highest probabilistic contingency in each training pair was selected. We then compared overall accuracy during acquisition in the ON condition to overall accuracy in the two OFF medication states using paired *t*-tests and a linear mixed model to detect differences in accuracy in the acquisition phase between different drug states. We also tested for differences in the acquisition rate between the different drug states by comparing learning rates in a reinforcement learning model (see below). For this test, we individually fitted the parameters of the reinforcement learning model to subjects' choices in the ON and OFF medication condition, comparing the resulting learning rates using a paired *t*-test.

Performance session

Data from the performance session were separated into trials in which the 'best' stimulus (80% chance of being correct) was presented, and trials in which the 'worst' stimulus (20% chance of being correct) was presented. We calculated the percentage of times subjects picked the best stimulus and the percentage of times the subjects avoided the worst stimulus in these novel pairings and tested for any differences in performance between the different medication conditions.

Reinforcement learning model

We used a simple prediction error-based reinforcement learning model (Sutton and Barto, 1998) to estimate a trial-by-trial measure of stimulus value, and thus an outcome prediction error δ defined as the difference between the actual observed outcome R (correct/incorrect = 1/0) and the current expected value of the chosen stimulus.

For each pair of stimuli A and B, the model estimates the expected values of choosing A, (Q_A) and choosing B (Q_B), on the basis of individual sequences of choices and outcomes. The expected values were set to zero before learning. After every trial $t > 0$ the value of the chosen stimulus (e.g. 'A') was updated according to the rule $Q_A(t + 1) = Q_A(t) + \alpha \times \delta(t)$. The outcome prediction error is the difference between the actual and the expected outcome, $\delta(t) = R(t) - Q_A(t)$ with the actual outcome being either 'Correct' or 'Incorrect' (1 or 0). Values of stimuli that were not shown on a trial were not updated.

Given the expected values, the probability (or likelihood) of the observed choice was estimated using the softmax rule: $P_A(t) = \exp[Q_A(t)/\beta]/\{\exp[Q_A(t)/\beta] + \exp[Q_B(t)/\beta]\}$. The parameters α (learning rate) and β (temperature) were adjusted to maximize the likelihood of the actual choices under the model, for all subjects. Trial-by-trial outcome prediction errors estimated by the model were then used as parametric regressors in the imaging data.

We also considered an alternative reinforcement learning model, which allowed for separate learning rates α^+ on positive updates (increasing the predicted value) and α^- on negative updates (decreasing the predicted value). We then compared model likelihoods of the models with separate learning rates and the original reduced model on an individual subject level using Bayesian Information Criterion (BIC), which corrects for the different complexity in models (smaller values indicate better fit) (Schwarz, 1978), and population level using Bayesian model comparison (Stephan *et al.*, 2009).

Functional magnetic resonance imaging: whole-brain general linear model parametric analysis

Acquisition session

Functional MRI time series were regressed onto a composite general linear model containing four regressors: trial onset time (the

appearance of the hiragana characters), outcome onset time, motor response time and fixation cross presentation time. The outcome onset was parametrically modulated by the prediction error as estimated by the reinforcement learning model. We also composed another general linear model in which there were four regressors: correct trial onset time, incorrect trial onset time, motor response time and fixation cross presentation time. The actual value of the chosen cue in each trial was entered as a parametric modulator of the two trial onset regressors.

Performance session

Four regressors were entered into the functional MRI model: correct trial onset time, incorrect trial onset time, motor response time and fixation cross presentation time. The actual value of the chosen cue in each trial was entered as a parametric modulator of the two trial onset regressors.

The regressors were convolved with the canonical haemodynamic response function, and low frequency drifts were excluded with a high-pass filter (128-s cut-off). Short-term temporal autocorrelations were modelled using an autoregressive [AR(1)] process. Motion correction regressors estimated from the realignment procedure were entered as covariates of no interest. Statistical significance was assessed using linear compounds of the regressors in the general linear model, generating statistical parametric maps of *t*-values across the brain for each subject and contrast of interest. These contrast images were then entered into a second-level random-effects analysis using a one-sample *t*-test against zero.

Anatomical localization was carried out by overlaying the *t*-maps on a normalized structural image averaged across subjects, and with reference to an anatomical atlas (Naidich *et al*., 2009). All coordinates are reported in MNI space (Mazziotta *et al*., 1995).

Region of interest analysis

We extracted data for all region of interest analyses using a cross-validation leave-one-out procedure: we re-estimated our main second-level analysis 12 times, always leaving out one subject. Starting at the peak voxel for the chosen cue value signal in ventromedial prefrontal cortex and nucleus accumbens, which was identified by looking over all correct trials (in both the ON and OFF drug states), we selected the nearest maximum in these cross-validation second-level analyses. Using that new peak voxel, we then extracted the data from the left-out subject and calculated a representative time-course for each region of interest as first eigenvariate from data in all voxels within a 4-mm sphere around that peak. We then performed a small volume correction on the striatal activations in the putamen using an anatomical region of interest defined according to the Talairach Daemon atlas (Lancaster *et al*., 1997) using the SPM WFU PickAtlas tool (Maldjian *et al*., 2003).

Results

We used a within-subject design to study a single group of patients with Parkinson's disease [early-to-moderate stage (Hoehn and Yahr stage: mean 1.69, SE 0.26)] in a generalization task introduced by Frank *et al.* (2004, 2007*b*) in three separate drug states (Fig. 1). We employed a within-subject design given the inherent difficulty in accurately matching patients with Parkinson's disease with different disease severity. We also believe that this design allowed us to minimize and control, as far as possible, for individual cognitive and genetic differences that may exist in our cohort, allowing us to look at the within-subject effects of drug on behaviour. In parallel with our behavioural analysis, we also acquired neural data using functional MRI. Thus, this design enabled us to explore the effect of dopamine on behaviour and on the brain by testing patients in three different drug states: acquisition and performance ON medication; acquisition and performance OFF medication and acquisition in an OFF medication state and performance in an ON medication state. The inclusion of the latter condition specifically enabled us to probe whether dopamine's effects are expressed during task acquisition (learning) or task performance.

Acquisition phase

Behavioural results

At the end of the acquisition phase, average choice accuracy on the training pairs did not differ between groups across the different drug states [paired *t*-tests: comparing ON–ON with OFF–ON: $t(1,12) = 0.15$, $P = 0.87$; comparing ON–ON with OFF–OFF: $t(1,12) = 0.095$, $P = 0.92$; comparing OFF–ON with OFF–OFF: $t(1,12) = -0.079$, $P = 0.93$] (Supplementary Table 4). Similarly, we found no significant difference in learning rates between patients when they were in an OFF compared to ON medication state [ON: mean 0.25, SE 0.02; OFF: mean 0.24, SE 0.01; paired *t*-test $t(1,12) = 0.117$, $P = 0.90$] (Supplementary Table 5), or in the number of sessions required to reach criteria [ON: mean 1.23, SE 0.12; OFF: mean 1.38, SE 0.16 paired *t*-test $t(1,12) = -0.69$, $P = 0.50$]. When the three types of training pairs were examined separately, there were no differences in choice accuracy between the different drug states. When we compared log evidences from a learning model with separate learning rates for positive and negative updates (posneg) with a single learning rate model (single), we found that the more complex model did not explain behaviour any better than the simple model (average $BIC_{single} = 260.9$ versus $BIC_{posneg} = 261.4$; posterior probability $P_{single} = 0.72$; exceedance probability single versus posneg model $P > 0.99$). Thus, subjects' behaviour could not be explained better with separate learning rates for positive and negative updating.

Neuroimaging data

Here, we examined brain responses that correlated with outcome prediction errors computed from a reinforcement learning model, fit to subjects' behaviour during the acquisition phase. We found that bilateral responses in the striatum (central coordinates right putamen $x = 26$, $y = 0$, $z = -4$; left putamen $x = -28$, $y = -12$, $z = -2$) (Fig. 2) strongly correlated with reward prediction errors [small volume corrected using anatomical region of interest for false discovery rate (FDR); left putamen $P = 0.007$, right putamen $P = 0.006$], consistent with many previous results (McClure *et al*., 2003; O'Doherty *et al*., 2003; Schonberg *et al*., 2010). However, similar to our behavioural findings, we found no differences in prediction error-related brain activation between the different drug states during acquisition [paired *t*-test ON compared with OFF: $t(1,11) = -0.076$, $P = 0.46$] (see Supplementary Table 6 for a whole-brain activation table for prediction error-related activity across all drug conditions). We separately examined neural responses to positive and negative prediction errors but did not

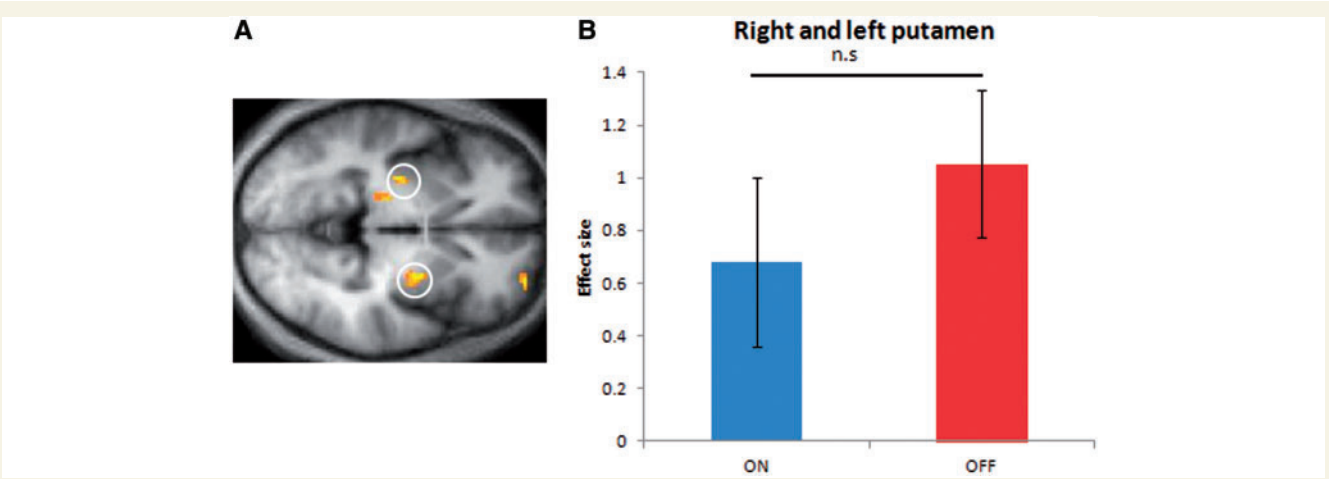

Figure 2 Prediction error-related activity during acquisition. (**A**) Brain activity in putamen correlated with magnitude of outcome prediction errors across all trials during the acquisition phase. Activations are thresholded at $P < 0.001$ uncorrected. (**B**) Correlation between outcome prediction errors and blood oxygen level-dependent activity in the two different drug states. Data in the 'ON' state was averaged across the two acquisition sessions performed in the scanner under this medication state and data in the 'OFF' state across the four acquisition sessions performed in this medication state. Error bars represent SEM. n.s. = not significant.

find any differences between the different drug states [paired *t*-tests comparing positive prediction errors ON compared with OFF: $t(1,11) = -0.083$, $P = 0.42$; and comparing negative prediction errors ON with OFF: $t(1,11) = -0.051$, $P = 0.614$]. Perhaps most surprisingly, at the time of cue onset, we did not observe any correlation between brain activity and the value of the chosen cue in any of the drug states.

Performance phase

Behavioural results

In the performance phase, after patients had acquired the task contingencies, along with all the training pairings, we presented the best (the one with 80% chance of being correct) and worst stimulus (the one with only 20% chance of being correct) in novel pairings with all the other stimuli (Fig. 1). We found that patients ON their dopamine replacement therapy performed significantly better than patients OFF dopamine replacement therapy [main effect comparing accuracy of the mean of ON–ON/OFF–ON with OFF–OFF, paired *t*-test, $t(1,12) = 2.8$, $P = 0.01$]. Crucially, a separate examination of the three drug states revealed a main effect of drug on performance but not on acquisition (Fig. 3A). Subjects who acquired the contingencies in an OFF medication state and received their dopamine replacement therapy after the acquisition phase, but before the performance phase, had the same level of overall accuracy as subjects who both acquired the contingencies ON medication and performed ON medication [paired *t*-test comparing ON–ON with OFF–ON, $t(1,12) = -0.03$, $P = 0.97$]. Both the ON–ON and OFF–ON groups were significantly more accurate than the OFF–OFF group [paired *t*-test comparing ON–ON with OFF–OFF, $t(1,12) = 2.17$, $P = 0.05$; and comparing OFF–ON with OFF–OFF, $t(1,12) = 2.28$ $P = 0.04$]. A mixed effects linear model showed a significant effect of drug state on performance accuracy at the performance phase [$F(1,36) = 5.38$, $P = 0.02$] but not at the acquisition phase [$F(1,36) = 0.002$, $P = 0.96$].

In addition to the novel pairings, we also presented subjects with the three stimulus pairs on which they had been trained during acquisition. Interestingly, we found no difference in accuracy levels on these training pairs across the different drug states [paired *t*-tests comparing ON–ON with OFF–ON, $t(1,12) = -1.36$, $P = 0.19$; comparing ON–ON with OFF–OFF, $t(1,12) = -0.64$, $P = 0.52$; comparing OFF–ON with OFF–OFF $t(1,12) = 1.26$, $P = 0.23$] (Fig. 3B), even when the three types of training pairs were examined separately. Indeed, there was no difference in the accuracies for training pairs versus novel pairs in ON–ON or OFF–ON drug states [paired *t*-tests comparing training pair accuracy with novel pair accuracy in ON–ON, $t(1,12) = 0.61$, $P = 0.55$ and OFF–ON drug states, $t(1,12) = -1.75$, $P = 0.10$]. However, as expected, in the OFF–OFF drug state, accuracy for training pairs was significantly higher than for novel pairs [$t(1,12) = -2.28$, $P = 0.04$]. Note that our results cannot be explained by a faster extinction effect in the patients when they were OFF dopamine replacement therapy. An effect of this sort would predict an overall gradual performance decrement over time. Instead, we found that patients OFF medication maintained their performance in the training pairs, which were randomly interspersed with the novel pairs, throughout the test session. One difference between our study and that of Frank *et al.* (2004, 2007*b*) is that in our study there was an added time delay between acquisition and transfer, whilst L-DOPA took effect. Since dopamine influences processes such as working memory (Sawaguchi and Goldman-Rakic, 1991; Watanabe *et al.*, 1997; Fuster, 2001; Stuss and Knight, 2002) and enhancing dopamine signalling will have had an effect on these processes (Sawaguchi, 2001; Wang *et al.*, 2004; Gibbs and D'Esposito, 2005; Cools and D'Esposito, 2011), one could argue that the poor generalization in the OFF group could stem from the delay. However, we believe this is unlikely given evidence that the OFF group performed just as well as the ON group on training pairs.

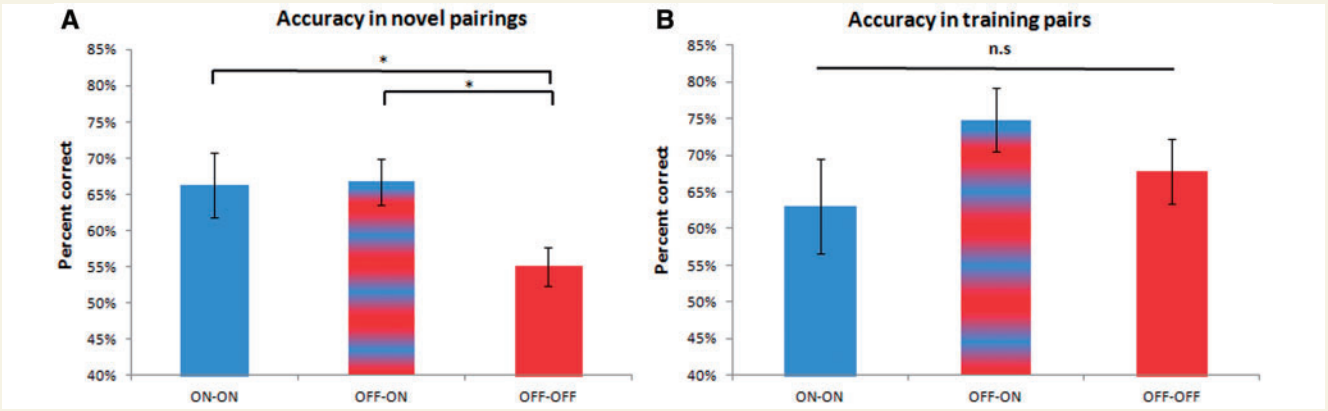

Figure 3 Accuracy during performance phase. (**A**) Accuracy in novel pairings was significantly higher when subjects were ON dopamine replacement therapy during the performance phase than when they were OFF. This effect is independent of drug state during the previous acquisition phase. Shown is the combined accuracy in selecting the best stimulus and avoiding the worst stimulus over the three drug states when subjects had to pick the stimulus with the highest likelihood of being correct when presented in novel parings. ON–ON session (blue), when patients took their usual dopamine replacement therapy; OFF–ON session (blue/red stripe), when patients took their dopamine replacement therapy only after completing the acquisition phase; OFF–OFF session (red), when patients abstained from their dopamine replacement therapy throughout the task. (**B**) Accuracy in selecting the better stimulus among the training pairs during the performance phase did not differ between drug states. n.s. = not significant.

We next tested for differential performance in selecting the best, and avoiding the worst, stimulus within the novel pairings. Interestingly, being in the ON dopamine replacement therapy state during the performance phase selectively improved accuracy in selecting the best stimulus compared to avoiding the worst stimulus for novel stimuli pairs [paired *t*-tests comparing ON accuracy for picking the best compared with avoiding worse stimulus $t(1,12) = 2.16$, $P = 0.05$]. This performance difference between selecting the best and avoiding the worst stimulus was not evident when subjects both acquired and performed the task in the OFF medication state [paired *t*-tests comparing OFF accuracy for picked best compared with avoiding worse stimulus $t(1,12) = 0.58$, $P = 0.56$], although their overall performance was worse (Fig. 4). Of note, there was no interaction between the medication status (ON versus OFF) during performance and the ability to pick the best compared with avoiding the worst stimulus as has previously been reported (Frank *et al*., 2004, 2007*b*). We only found this selective improvement in picking the best stimulus compared with avoiding the worst stimulus within the ON group. We observed the same pattern when analysing the ON–ON and OFF–ON groups separately. However, although a trend was evident this did not reach full significance [paired *t*-tests comparing ON accuracy for picking the best compared with avoiding worse stimulus in the ON–ON group alone, $t(1,12) = 1.5$, $P = 0.15$; and in the OFF–ON group alone, $t(1,12) = 1.68$, $P = 0.11$]. When we separately compared accuracy for picking the best stimuli, and for avoiding the worst stimuli, across groups there were no significant differences [paired *t*-tests comparing 'pick best' accuracy; ON–ON with OFF–ON, $t(1,12) = -0.26$, $P = 0.98$; comparing ON–ON with OFF–OFF, $t(1,12) = 1.79$, $P = 0.09$; comparing OFF–ON with OFF–OFF, $t(1,12) = 1.73$, $P = 0.1$; paired *t*-tests comparing 'avoid worse' accuracy: ON–ON with OFF–ON, $t(1,12) = -0.07$, $P = 0.94$; comparing ON–ON with OFF–OFF, $t(1,12) = 1.27$, $P = 0.22$; comparing OFF–ON with OFF–OFF, $t(1,12) = 1.6$, $P = 0.13$]. This indicates that when subjects were ON dopaminergic medication there was an asymmetry in performance between picking the best compared with avoiding the worst stimuli. Our data do not show that medication selectively improves generalization only for the best stimuli but rather that it affects the ability to generalize learnt information overall. When we examined reaction times, we found that across all groups there was a significant difference in reaction times between pick best and avoid worse stimuli trials [paired *t*-test, $t(1,12) = -2.52$, $P = 0.027$] with subjects being faster for the pick best trials. There were, however, no significant between group differences in reaction times.

Neuroimaging data

To investigate possible neural mechanisms underlying the observed behavioural effects during the performance phase, we next tested for differences in the degree at which functional MRI blood oxygen level-dependent activity correlated with decision variables between different drug states. We tested whether neural representations of stimulus values at the time of cue presentation differed between drug states. We found that blood oxygen level-dependent activity in nucleus accumbens (central coordinates $x = 8$, $y = 12$, $z = -4$) correlated with the value of the chosen cue, but this effect was only evident in the ON medication state for correct novel trials [one sample *t*-test, $t(1,11) = 2.7$, $P = 0.01$]. Cue-evoked blood oxygen level-dependent activity did not correlate with the value of the chosen cue when patients were OFF their dopamine replacement therapy [one sample *t*-test, $t(1,11) = 0.98$, $P = 0.34$] or made an incorrect choice [one sample *t*-test ON incorrect, $t(1,11) = -2.12$, $P = 0.06$; OFF incorrect $t(1,11) = -0.06$, $P = 0.94$] (Fig. 5A and B). We found an identical effect in ventromedial prefrontal cortex ($x = -2$, $y = 38$, $z = 0$), where blood oxygen level-dependent activity varied with the value of the chosen cue when patients were both ON medication and made the correct choice [one sample *t*-test, $t(1,11) = 2.52$, $P = 0.02$], but

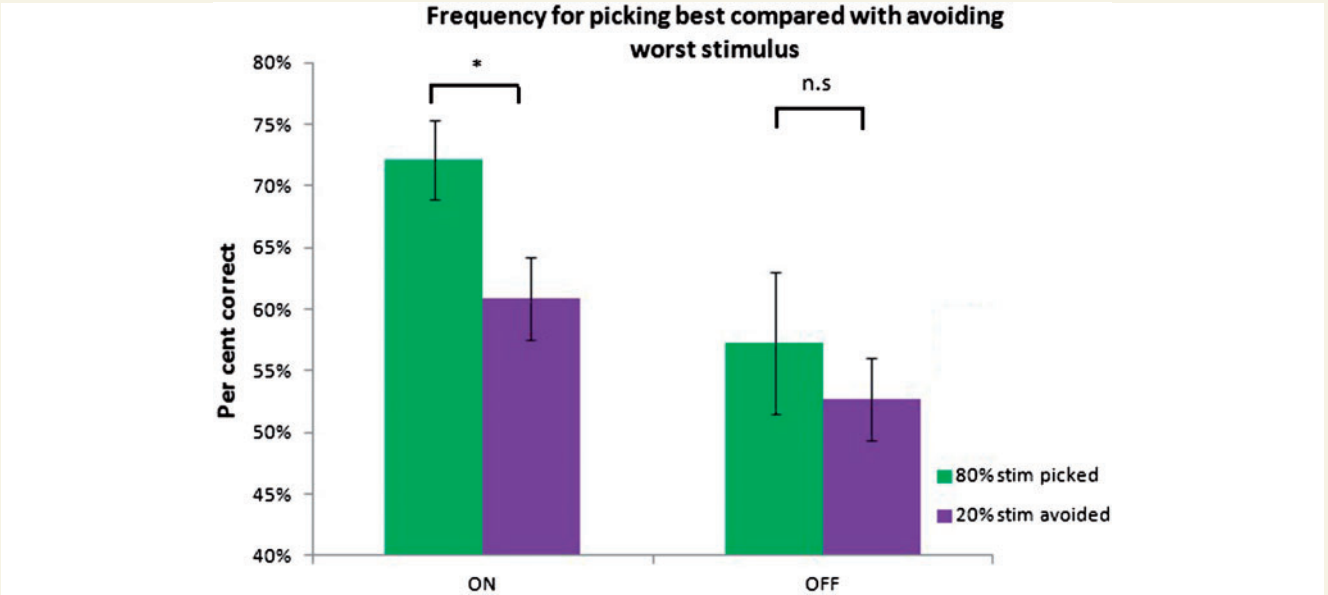

Figure 4 Differences in accuracy at picking best compared with avoiding worst stimuli. ON state during performance phase selectively improved accuracy for picking the best stimulus (the 80%) compared to avoiding the worst stimulus (the 20% stimulus) in novel pairings. The data in the 'ON' state comes from the two performance bouts performed in this medication state and the data in the 'OFF' state comes from the single performance bout performed in this medication state.

not when they were OFF their dopamine replacement therapy [one sample *t*-test, $t(1,11) = 0.31$, $P = 0.76$], or made an incorrect choice [one sample *t*-test ON incorrect, $t(1,11) = -1.28$, $P = 0.22$; OFF incorrect $t(1,11) = 0.76$, $P = 0.46$] (Fig. 5C and D). These findings show that activity in nucleus accumbens and ventromedial prefrontal cortex successfully reflect the values of the most rewarding cue only in an ON medication state, a characteristic that precisely mirrors patients' improved performance in this state. We found that this effect was not driven solely by the ON–ON group. When the OFF–ON group are examined separately, cue-evoked blood oxygen level-dependent activity correlated with the value of the chosen cue with patients made the correct choice both in the nucleus accumbens [one sample *t*-test, $t(1,11) = 2.86$, $P = 0.015$] and in the ventromedial prefrontal cortex [one sample *t*-test, $t(1,11) = 2.93$ $P = 0.014$], however, when we directly compared the activations in OFF–ON with the OFF–OFF group this did not reach statistical significance [paired *t*-test comparing OFF–ON with OFF–OFF, nucleus accumbens $t(1,11) = 1.62$, $P = 0.13$, ventromedial prefrontal cortex $t(1,11) = 1.02$, $P = 0.32$].

Akin to the neuroimaging findings from the acquisition phase, when we examined value-related neural activity at the time of cue onset during presentation of the training pairs at the performance phase, we did not find a significant correlation between blood oxygen level-dependent activity and the value of the chosen cue in either of the drug states.

Discussion

We show a striking effect of dopamine replacement therapy on the ability of patients with Parkinson's disease to select the correct stimulus in a probabilistic reinforcement learning task. Importantly, our data show that medication status at the acquisition task phase does not impact successful task learning. Instead, the data show that the critical factor is medication status at the performance phase, by which time stimulus values must already have been successfully acquired. The findings challenge a proposal that the impact of dopaminergic status on this form of decision-making solely reflects its involvement in learning.

Our key observation was that patients who were OFF dopamine during the second task phase performed significantly worse when stimuli occurred in novel pairings. However, dopaminergic drug state did not impact their ability to choose when confronted with pairs on which they had been trained in the first phase of the task. This indicates that the subjects OFF medication could successfully retrieve learnt contingencies but were unable to use this knowledge to make correct choices when they had to select between novel stimulus pairings. There was no difference in learning rates or accuracy during the acquisition phase between the different drug conditions, indicating that dopamine did not affect the ability to learn stimulus values. Consequently, our data indicate that dopamine replacement therapy influenced the ability to generalize, in a context, where subjects needed to select the best stimulus in a state characterized by presentation of novel pairings.

A mechanistic basis for our behavioural findings is provided by our functional MRI data, which specifically addressed the neural representation of stimulus value during the performance phase. Even when subjects had learned stimuli OFF dopamine replacement therapy, and were only given their dopamine replacement therapy after learning had occurred, activity in nucleus accumbens and ventromedial prefrontal cortex encoded the value of the chosen stimulus during the performance phase, allowing the brain to compare those values in novel pairings. This suggests that, in contrast

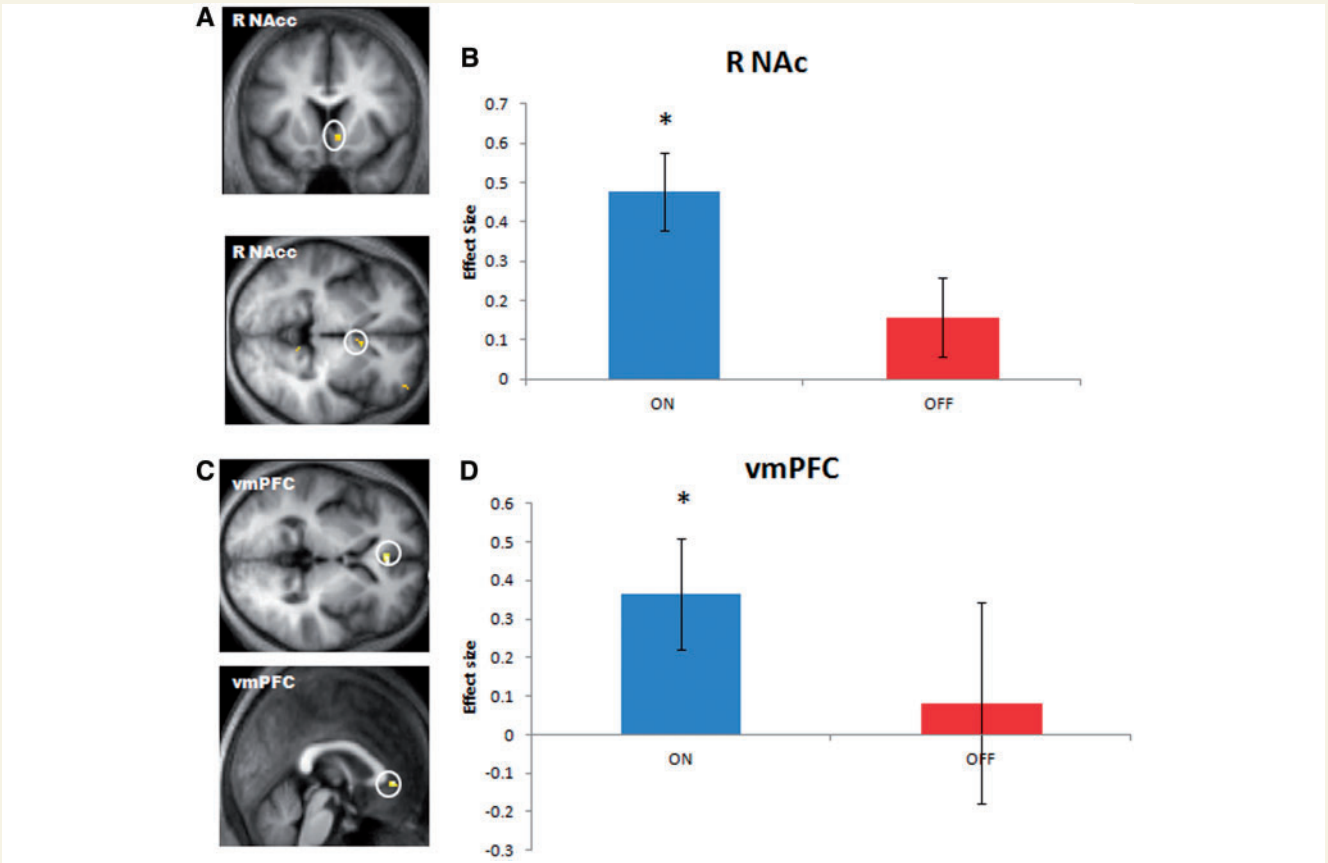

Figure 5 Brain activity correlating with the value of the chosen cue during performance phase. (**A**) Brain activity in right nucleus accumbens (R NAc) correlated with the value of the chosen cue. Analysis performed over all correct trials (both ON and OFF) in a context where novel parings are presented. (**B**) A differential analysis between drug states reveals this correlation was selective to the ON state. (**C**) Brain activity in ventromedial prefrontal cortex (vmPFC) also correlated with the value of the chosen cue. Whole brain analysis performed over all correct trials (both ON and OFF). (**D**) Similar to activity in nucleus accumbens, the correlation between blood oxygen level-dependent values in ventromedial prefrontal cortex and the value of the chosen cue was only evident in ON but not in OFF state. The error bars represent SEM. Thresholds in statistical parametric map images set to $P < 0.005$ uncorrected.

to previous accounts (Schultz *et al*., 1997; O'Doherty *et al*., 2003; Bayer and Glimcher, 2005; Pessiglione *et al*., 2006), reduced dopamine availability during learning did not impair value acquisition. Instead, our data show that decreased dopamine during performance resulted in an impoverished neural representation of stimulus value. This we suggest impaired an ability to compare cue values in novel pairings, in other words to generalize from learnt information. On the other hand, non-novel pairings (i.e. the acquisition pairs) could still be answered correctly if a stimulus–response association had been learned; this would not depend directly on value representations [and could for example be achieved by fixed stimulus–response associations, or explicit (episodic) memory retrieval], and could therefore operate successfully even when dopamine levels were low. This explanation best accounts for why subjects who were OFF medication throughout the task were equally successful at choosing the best stimuli in the context of the training pairs. It is of interest that the two structures highlighted in our data, the nucleus accumbens and ventromedial prefrontal cortex, are strongly associated with various forms of value prediction and prediction errors in reinforcement learning contexts (Matsumoto *et al*., 2003; Day and Carelli, 2007; Luk and Wallis, 2009). The pattern of findings we observed, whereby stimulus value correlated with activity in these two regions in the ON state, implies that these brain areas can successfully represent the reward value of cues when patients are ON medication enabling successful performance for novel pairings. However, when this signal is degraded as seen in the OFF state, performance is impaired.

The fact that patients in all three drug states performed equally well when they were selecting the best cue for sets on which they had been previously trained further indicates that dopamine did not influence patients' accuracy by a direct influence on learning. Levodopa medication in patients with Parkinson's disease has previously been shown to have a positive effect on generalization of learnt information in novel contexts; however, those observations were on a background of impaired learning and therefore crucially different from our current findings (Myers *et al*., 2003; Shohamy *et al*., 2006). Of course, many different systems are likely to be involved in learning, only some of which depend directly on dopamine (Beninger, 1983; Dickinson *et al*., 2000; Daw *et al*., 2005; Palmiter, 2008), and we cannot discount the

possibility that a more complex learning task, such as one involving sequences of choices, might be necessary to fully reveal effects of dopamine on learning.

Beyond its putative role in learning, dopamine is implicated in a number of distinct processes related to motivation, including the control of Pavlovian conditioned responses and motivational vigour (Dickinson *et al*., 2000; Parkinson, *et al*., 2002; Salamone *et al*., 2003; Berridge, 2007; Mazzoni *et al*., 2007; Niv, 2007; Bardgett *et al*., 2009; Beeler *et al*., 2010; Boureau and Dayan, 2011). The strong influence of dopamine on performance, separate to that on learning, is well known from animal data. For example, dopamine-deficient mice retain the ability to pick the most rewarding drink (sucrose compared with water) when presented in a discrimination task (Cannon and Palmiter, 2003). Genetic dopamine-deficient mice when tested in a maze task appears initially impaired but when subsequently treated with L-DOPA (Robinson *et al*., 2005), or caffeine (Hnasko *et al*., 2005) can be shown to have learned, consistent with an effect of dopamine on the expression of learning rather than on learning itself. In addition, dopamine is implicated in controlling movement rate and vigour (Ungerstedt, 1971; Salamone *et al*., 2003; Cagniard *et al*., 2006) with dopamine depletion causing decreased motivation to work for rewards under demanding reinforcement schedules (Salamone and Correa, 2002; Niv, 2007).

Importantly, these other roles remain consistent with the fact that phasic activity of dopamine neurons codes for an appetitive prediction error (McClure *et al*., 2003). However, our study has enabled us to disentangle these effects from a mere effect on learning in a manner that provides compelling evidence that dopamine has a specific role on the expression of learning that is distinct from any effect it may have on learning itself. We are, however, unable to comment on whether the drug manipulation, which included the withdrawal and then reinstatement of both L-DOPA and dopamine agonists, primarily exerted its main effect on tonic or phasic levels of dopamine although we infer that it is likely to have an effect on both.

The involvement of the nucleus accumbens during successful performance is particularly notable, since this structure is well known to control the immediate effects of dopamine on numerous aspects of performance (Berridge and Robinson, 1998; Ikemoto and Panksepp, 1999; Berridge, 2009; Lex and Hauber, 2010). The nucleus accumbens is a site where the predicted values of stimuli are transformed into preparatory Pavlovian responses under a modulatory influence of dopamine (Berridge and Robinson, 1998). We suggest that a preparatory response of approach is likely to be a key substrate for the behavioural patterns we observed in our task (Dayan *et al*., 2006). This provides another reminder of the complexities inherent in a single neuromodulator (dopamine) supporting two apparently independent roles, namely reporting on appetitive prediction errors and influencing motivation and vigour (Ikemoto and Panksepp, 1999; Niv *et al*., 2007; Boureau and Dayan, 2011; Cools *et al*., 2011).

A further important finding is the engagement of ventromedial prefrontal cortex in a context in which subjects made the correct choice between novel pairings of stimuli in the ON state, but not when subjects made incorrect choices in the ON state. This region is strongly implicated in valuation (Gottfried *et al*., 2003; Seymour and McClure, 2008; Boorman *et al*., 2009; Kable and Glimcher, 2009; Fitzgerald *et al*., 2010; Plassmann *et al*., 2010) across a range of experimental manipulations, with mounting evidence pointing to a specific role when subjects have to choose between distinct options with different values (Padoa-Schioppa and Assad, 2006; FitzGerald *et al*., 2009; Wunderlich *et al*., 2010). This fits neatly with our observation that this region was engaged when subjects generated correct choices based upon an assessment of a learnt value difference between novel pairings. However, our data are intriguing in suggesting that the integrity of a dopamine input to this region is important for this form of value-based decision.

Of course, we cannot be certain as to dopamine's precise role in our task. However, two possibilities are immediately apparent: dopamine is either necessary for a stable value representation that can support generalization, or alternatively, for taking the difference between the values of the available stimuli in order to choose. We were unable to dissociate whether the neural value correlates were precursors to choice (stimulus values) or the output of the choice process (chosen values) (Wunderlich *et al*., 2009). It remains an open question for future research as to whether the deficit is due to a misrepresentation of pre-choice values that are fed into a decision comparator, or reflect a problem at the value comparison stage itself or indeed a combination of both.

Our study involved testing patients with Parkinson's disease, which although providing the best human model of dopamine depletion, there is by necessity the problem of whether observations in this population can be generalized to the healthy population. Despite this caveat, our findings do lend support to the hypotheses (Berridge and Robinson, 1998; Berridge, 2007) and animal studies (Ahlenius *et al*., 1977; Beninger and Phillips, 1981; Wyvell and Berridge, 2000; Cannon and Palmiter, 2003; Denenberg *et al*., 2004; Robinson *et al*., 2005) that stress a major role for dopamine outside of learning.

A significant finding from our study is that when patients were ON their dopamine replacement therapy, they were worse at avoiding stimuli with the poorest probabilistic contingencies than at choosing the stimuli with the best probabilistic outcomes. This is in keeping with previous research showing a similar outcome valence performance asymmetry, whereby patients ON their dopamine replacement therapy are impaired at avoiding the least rewarding stimuli (Frank *et al*., 2004, 2007*b*). It has been postulated that this worsening in performance is due to 'overdosing' of the striatum, which interferes with the dips in dopamine that express negative prediction errors (Frank *et al*., 2004, 2007*b*). However, in our study, as in several others (de Wit *et al*., 2011; Jocham *et al*., 2011), we did not find a direct effect of medication on learning, and we postulate that the worsened performance may reflect some other mechanism, perhaps an impaired expression of avoidance behaviour in a high dopamine state.

Of note, we did not find the interaction between medication state and picking the best compared with avoiding the worst stimulus that has been reported in some previous studies (Frank *et al*., 2004, 2007*b*; Voon *et al*., 2010). We found an overall improvement in performance when subjects were ON medication and an asymmetry in this performance accuracy between picking the best compared with avoiding the worst stimuli within this group.

We did not find a significant difference between picking the best stimulus in the ON state and picking the best stimulus in the OFF state. Thus, although we can conclude that when subjects

were ON dopaminergic medication, there was an asymmetry in performance between picking the best compared with avoiding the worst stimuli, we cannot be certain whether this improvement in performance in the ON compared with the OFF group is solely due to an improved ability to select the most rewarding stimuli. Since the OFF–OFF patients generalized so poorly, there remains a possibility that we were unable to detect an asymmetry in performance in the OFF group due to floor effects. However, the clear asymmetry observed in the ON group is of interest as it provides a hint as to where dopamine may exert an important influence.

Several studies have also found differences in striatal activations in response to wins and losses when comparing patients with Parkinson's disease with compulsive behaviour ON medication to patients with idiopathic Parkinson's disease (Steeves *et al*., 2009; Voon *et al*., 2010) or in healthy subjects who were given dopaminergic modulating drugs (Pessiglione *et al*., 2006; Cools *et al*., 2007). We did not find this pattern in our study, possibly because of the unique feature of our design in making within-subject comparisons in patients with idiopathic Parkinson's disease. In particular, acute pharmacological manipulations in healthy volunteers may have very different effects to those found in patients previously exposed to dopaminergic agents. Another potential explanation for these differences is that in contrast to the studies that found differences in striatal activations in response to gains and losses (Pessiglione *et al*., 2006; Bodi *et al*., 2009; Palminteri *et al*., 2009; Voon *et al*., 2010), we did not have actual losses as outcomes, only stimuli that were probabilistically more or less likely to be correct. We caution against a conclusion that our data indicate that prediction errors do not play an important role in learning. Indeed, we observed prediction errors during the task acquisition. We were, however, unable to detect differences in the magnitude of these activations when dopaminergic drugs were given to this patient group, consistent with a suggestion that for successful choice, it is critical that dopamine levels are high during actual performance. We also cannot discount the fact that there might remain some residual activity within the dopaminergic system of patients with Parkinson's disease, resulting in adequate levels of dopamine to signal reward prediction errors in some structures, but not enough to form adequate cue value representations to allow successful generalization in those, or other, structures. Furthermore, in some contexts the relationship between dopamine appears to follow an inverted U-shaped function whereby the optimum level of performance exists at a certain level of dopaminergic stimulation and moving off that peak, either by reducing or increasing the levels of dopamine, leads to worsened task performance (Robbins, 2000; Rowe *et al*., 2008; Cools and D'Esposito, 2011) rendering dopaminergic manipulations crucially dependant on baseline levels. There is a possibility that high level of cognitive processing, such as working memory, required for this task may obscure differences in striatal activations. Indeed, differing performance in tasks of this type has been shown in genetic studies to have a differential impact on prefrontal and striatal dopamine (Frank *et al*., 2007*a*; Klein *et al*., 2007). However, in our task if dopamine had in fact boosted the prediction error magnitude in a way that impacted learning, we would have expected to see this in improved behaviour in the performance phase. This would result in the ON–ON group performing the best, which was in fact not the case.

By teasing apart learning and performance in patients with Parkinson's disease, we found that dopaminergic medication impacted the latter, but not the former. At the neural level, this improved performance in the ON medication state was associated with enhanced nucleus accumbens and ventromedial prefrontal cortex activity for the chosen cue value, an effect that was absent in the OFF medication state. Thus, the improved performance in patients ON medication cannot solely be attributed to an effect on learning and must reflect some other effect of dopamine, perhaps Pavlovian appetitive approach or a modulation of the motivational impact of cues associated with improved neural representation of cue value in a high dopamine state. By isolating the processes on which dopamine has the greatest impact, our findings point to likely mechanisms that underlie common behavioural deficits seen in patients with Parkinson's disease, both clinically and in various laboratory tasks, as well as providing a basis for future cognitive-oriented therapies.

Acknowledgements

The authors thank the reviewers for their insightful and helpful comments. They also thank Tali Sharot, Rosalyn Moran and the emotion group for their helpful input on this work.

Funding

Wellcome Trust [Ray Dolan programme grant number 078865/Z/05/Z], an MRC grant (to T.S.), Gatsby Charitable Foundation (to P.D.), Wellcome Trust (091593/Z/10/Z).

Supplementary material

Supplementary material is available at *Brain* online.

References

Ahlenius S, Engel J, Zoller M. Effects of apomorphine and haloperidol on exploratory behavior and latent learning in mice. Physiol Psychol 1977; 5: 290–4.

Ashburner J, Friston KJ. Unified segmentation. Neuroimage 2005; 26: 839–51.

Bardgett ME, Depenbrock M, Downs N, Points M, Green L. Dopamine modulates effort-based decision making in rats. Behav Neurosci 2009; 123: 242–51.

Bayer HM, Glimcher PW. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 2005; 47: 129–41.

Beeler JA, Daw N, Frazier CR, Zhuang X. Tonic dopamine modulates exploitation of reward learning. Front Behav Neurosci 2010; 4: 170.

Beninger RJ. The role of dopamine in locomotor activity and learning. Brain Res 1983; 287: 173–96.

Beninger RJ, Phillips AG. The effects of pimozide during pairing on the transfer of classical conditioning to an operant discrimination. Pharmacol Biochem Behav 1981; 14: 101–5.

Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology 2007; 191: 391–431.
Berridge KC. 'Liking' and 'wanting' food rewards: brain substrates and roles in eating disorders. Physiol Behav 2009; 97: 537–550.
Berridge KC, Robinson TE. What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? Brain Res Brain Res Rev 1998; 28: 309–69.
Bodi N, Keri S, Nagy H, Moustafa A, Myers CE, Daw N, et al. Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain 2009; 132: 2385–95.
Boorman ED, Behrens TE, Woolrich MW, Rushworth MF. How green is the grass on the other side? Frontopolar cortex and the evidence in favor of alternative courses of action. Neuron 2009; 62: 733–43.
Boureau YL, Dayan P. Opponency revisited: competition and cooperation between dopamine and serotonin. Neuropsychopharmacology 2011; 36: 74–97.
Cagniard B, Balsam PD, Brunner D, Zhuang X. Mice with chronically elevated dopamine exhibit enhanced motivation, but not learning, for a food reward. Neuropsychopharmacology 2006; 31: 1362–70.
Cannon CM, Palmiter RD. Reward without dopamine. J Neurosci 2003; 23: 10827–31.
Cools R, D'Esposito M. Inverted-U-shaped dopamine actions on human working memory and cognitive control. Biol Psychiatry 2011; 69: e113–e125.
Cools R, Nakamura K, Daw ND. Serotonin and dopamine: unifying affective, activational, and decision functions. Neuropsychopharmacology 2011; 36: 98–113.
Cools R, Sheridan M, Jacobs E, D'Esposito M. Impulsive personality predicts dopamine-dependent changes in frontostriatal activity during component processes of working memory. J Neurosci 2007; 27: 5506–14.
Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 2005; 8: 1704–11.
Day JJ, Carelli RM. The nucleus accumbens and Pavlovian reward learning. Neuroscientist 2007; 13: 148–59.
Dayan P, Niv Y, Seymour B, Daw ND. The misbehavior of value and the discipline of the will. Neural Netw 2006; 19: 1153–60.
de Wit S, Barker RA, Dickinson AD, Cools R. Habitual versus goal-directed action control in Parkinson disease. J Cogn Neurosci 2011; 23: 1218–29.
Denenberg VH, Kim DS, Palmiter RD. The role of dopamine in learning, memory, and performance of a water escape task. Behav Brain Res 2004; 148: 73–8.
Dickinson A, Smith J, Mirenowicz J. Dissociation of Pavlovian and instrumental incentive learning under dopamine antagonists. Behav Neurosci 2000; 114: 468–83.
Edwards M, Quinn N, Bhatia K. Parkinson's disease and other movement disorders. Oxford: Oxford University Press; 2008.
Fitzgerald TH, Seymour B, Bach DR, Dolan RJ. Differentiable neural substrates for learned and described value and risk. Curr Biol 2010; 20: 1823–9.
FitzGerald TH, Seymour B, Dolan RJ. The role of human orbitofrontal cortex in value comparison for incommensurable objects. J Neurosci 2009; 29: 8388–95.
Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci USA 2007a; 104: 16311–16.
Frank MJ, Samanta J, Moustafa AA, Sherman SJ. Hold your horses: impulsivity, deep brain stimulation, and medication in parkinsonism. Science 2007b; 318: 1309–12.
Frank MJ, Seeberger LC, O'Reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 2004; 306: 1940–3.
Fuster JM. The prefrontal cortex—an update: time is of the essence. Neuron 2001; 30: 319–33.
Gibbs SE, D'Esposito M. Individual capacity differences predict working memory performance and prefrontal activity following dopamine stimulation. Cogn Affect Behav Neurosci 2005; 5: 212–221.
Gottfried JA, O'Doherty J, Dolan RJ. Encoding predictive reward value in human amygdala and orbitofrontal cortex. Science 2003; 301: 1104–7.
Graef S, Biele G, Krugel LK, Marzinzik F, Wahl M, Wotka J, et al. Differential influence of levodopa on reward-based learning in Parkinson's disease. Front Hum Neurosci 2010; 4: 169.
Haber SN, Knutson B. The reward circuit: linking primate anatomy and human imaging. Neuropsychopharmacology 2010; 35: 4–26.
Hnasko TS, Sotak BN, Palmiter RD. Morphine reward in dopamine-deficient mice. Nature 2005; 438: 854–857.
Ikemoto S, Panksepp J. The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. Brain Res Brain Res Rev 1999; 31: 6–41.
Jocham G, Klein TA, Ullsperger M. Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. J Neurosci 2011; 31: 1606–13.
Kable JW, Glimcher PW. The neurobiology of decision: consensus and controversy. Neuron 2009; 63: 733–45.
Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M. Genetically determined differences in learning from errors. Science 2007; 318: 1642–5.
Knowlton BJ, Mangels JA, Squire LR. A neostriatal habit learning system in humans. Science 1996; 273: 1399–402.
Koller WC, Melamed E. Parkinson's disease and related disorders: part 1. In: Aminoff MJ, Boller E, Swaab DE, editors. Handbook of clinical neurology. Vol. 83. Philadephia: Elsevier; 2007.
Lancaster JL, Rainey LH, Summerlin JL, Freitas CS, Fox PT, Evans AC, et al. Automated labeling of the human brain: a preliminary report on the development and evaluation of a forward-transform method. Hum Brain Mapp 1997; 5: 238–42.
Lex B, Hauber W. The role of nucleus accumbens dopamine in outcome encoding in instrumental and Pavlovian conditioning. Neurobiol Learn Mem 2010; 93: 283–90.
Luk CH, Wallis JD. Dynamic encoding of responses and outcomes by neurons in medial prefrontal cortex. J Neurosci 2009; 29: 7526–39.
Maia TV, Frank MJ. From reinforcement learning models to psychiatric and neurological disorders. Nat Neurosci 2011; 14: 154–62.
Maldjian JA, Laurienti PJ, Kraft RA, Burdette JH. An automated method for neuroanatomic and cytoarchitectonic atlas-based interrogation of fMRI data sets. Neuroimage 2003; 19: 1233–9.
Matsumoto K, Suzuki W, Tanaka K. Neuronal correlates of goal-based motor selection in the prefrontal cortex. Science 2003; 301: 229–32.
Mazziotta JC, Toga AW, Evans A, Fox P, Lancaster J. A probabilistic atlas of the human brain: theory and rationale for its development. The International Consortium for Brain Mapping (ICBM). Neuroimage 1995; 2: 89–101.
Mazzoni P, Hristova A, Krakauer JW. Why don't we move faster? Parkinson's disease, movement vigor, and implicit motivation. J Neurosci 2007; 27: 7105–16.
McClure SM, Berns GS, Montague PR. Temporal prediction errors in a passive learning task activate human striatum. Neuron 2003; 38: 339–46.
McClure SM, Daw ND, Montague PR. A computational substrate for incentive salience. Trends Neurosci 2003; 26: 423–428.
Myers CE, Shohamy D, Gluck MA, Grossman S, Kluger A, Ferris S, et al. Dissociating hippocampal versus basal ganglia contributions to learning and transfer. J Cogn Neurosci 2003; 15: 185–93.
Naidich TP, Duvernoy HM, Delman BN, Sorensen AG, Kollias SS, Haacke EM. Duvernoy's Atlas of the Human Brain Stem and Cerebellum. NewYork: SpringerWien; 2009.
Niv Y. Cost, benefit, tonic, phasic: what do response rates tell us about dopamine and motivation? Ann NY Acad Sci 2007; 1104: 357–76.
Niv Y, Daw ND, Joel D, Dayan P. Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology 2007; 191: 507–20.

O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. Neuron 2003; 38: 329–37.

Padoa-Schioppa C, Assad JA. Neurons in the orbitofrontal cortex encode economic value. Nature 2006; 441: 223–26.

Palminteri S, Lebreton M, Worbe Y, Grabli D, Hartmann A, Pessiglione M. Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. Proc Natl Acad Sci USA 2009; 106: 19179–84.

Palmiter RD. Dopamine signaling in the dorsal striatum is essential for motivated behaviors: lessons from dopamine-deficient mice. Ann NY Acad Sci 2008; 1129: 35–46.

Parkinson JA, Dalley JW, Cardinal RN, Bamford A, Fehnert B, Lachenal G, et al. Nucleus accumbens dopamine depletion impairs both acquisition and performance of appetitive Pavlovian approach behaviour: implications for mesoaccumbens dopamine function. Behav Brain Res 2002; 137: 149–63.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 2006; 442: 1042–1045.

Plassmann H, O'Doherty JP, Rangel A. Appetitive and aversive goal values are encoded in the medial orbitofrontal cortex at the time of decision making. J Neurosci 2010; 30: 10799–808.

Robbins TW. Chemical neuromodulation of frontal-executive functions in humans and other animals. Exp Brain Res 2000; 133: 130–8.

Robinson S, Sandstrom SM, Denenberg VH, Palmiter RD. Distinguishing whether dopamine regulates liking, wanting, and/or learning about rewards. Behav Neurosci 2005; 119: 5–15.

Rowe JB, Hughes L, Ghosh BC, Eckstein D, Williams-Gray CH, Fallon S, et al. Parkinson's disease and dopaminergic therapy—differential effects on movement, reward and cognition. Brain 2008; 131: 2094–105.

Salamone JD, Correa M. Motivational views of reinforcement: implications for understanding the behavioral functions of nucleus accumbens dopamine. Behav Brain Res 2002; 137: 3–25.

Salamone JD, Correa M, Mingote S, Weber SM. Nucleus accumbens dopamine and the regulation of effort in food-seeking behavior: implications for studies of natural motivation, psychiatry, and drug abuse. J Pharmacol Exp Ther 2003; 305: 1–8.

Sawaguchi T. The effects of dopamine and its antagonists on directional delay-period activity of prefrontal neurons in monkeys during an oculomotor delayed-response task. Neurosci Res 2001; 41: 115–28.

Sawaguchi T, Goldman-Rakic PS. D1 dopamine receptors in prefrontal cortex: involvement in working memory. Science 1991; 251: 947–50.

Schonberg T, O'Doherty JP, Joel D, Inzelberg R, Segev Y, Daw ND. Selective impairment of prediction error signaling in human dorsolateral but not ventral striatum in Parkinson's disease patients: evidence from a model-based fMRI study. Neuroimage 2010; 49: 772–81.

Schultz W. Predictive reward signal of dopamine neurons. J Neurophysiol 1998; 80: 1–27.

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997; 275: 1593–1599.

Schwarz G. Estimating the dimension of a model. Ann Statis 1978; 6: 461–4.

Seymour B, McClure SM. Anchors, scales and the relative coding of value in the brain. Curr Opin Neurobiol 2008; 18: 173–8.

Shohamy D, Myers CE, Geghman KD, Sage J, Gluck MA. L-dopa impairs learning, but spares generalization, in Parkinson's disease. Neuropsychologia 2006; 44: 774–84.

Steeves TD, Miyasaki J, Zurowski M, Lang AE, Pellecchia G, Van Eimeren T, et al. Increased striatal dopamine release in Parkinsonian patients with pathological gambling: a [11C] raclopride PET study. Brain 2009; 132: 1376–85.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ. Bayesian model selection for group studies. Neuroimage 2009; 46: 1004–17.

Stuss DT, Knight RT, editors. Principles of frontal lobe function. New York: Oxford University Press; 2002.

Sutton R, Barto AG. Reinforcement learning, an introduction. Cabridge, MA: MIT press; 1998.

Ungerstedt U. Striatal dopamine release after amphetamine or nerve degeneration revealed by rotational behaviour. Acta Physiol Scand Suppl 1971; 367: 49–68.

Voon V, Pessiglione M, Brezing C, Gallea C, Fernandez HH, Dolan RJ, et al. Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. Neuron 2010; 65: 135–42.

Wang M, Vijayraghavan S, Goldman-Rakic PS. Selective D2 receptor actions on the functional circuitry of working memory. Science 2004; 303: 853–6.

Watanabe M, Kodama T, Hikosaka K. Increase of extracellular dopamine in primate prefrontal cortex during a working memory task. J Neurophysiol 1997; 78: 2795–8.

Wise RA. Dopamine, learning and motivation. Nat Rev Neurosci 2004; 5: 483–494.

Wise RA, Rompre PP. Brain dopamine and reward. Annu Rev Psychol 1989; 40: 191–225.

Wunderlich K, Rangel A, O'Doherty JP. Neural computations underlying action-based decision making in the human brain. Proc Natl Acad Sci USA 2009; 106: 17199–204.

Wunderlich K, Rangel A, O'Doherty JP. Economic choices can be made using only stimulus values. Proc Natl Acad Sci USA 2010; 107: 15005–10.

Wyvell CL, Berridge KC. Intra-accumbens amphetamine increases the conditioned incentive salience of sucrose reward: enhancement of reward "wanting" without enhanced "liking" or response reinforcement. J Neurosci 2000; 20: 8122–30.

Yin HH, Ostlund SB, Balleine BW. Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. Eur J Neurosci 2008; 28: 1437–48.

Supplementary Material

Table S1: Neuropsychological data sets

	Patients (n=13)
Age	61.8 (3.3)
Education (years since age 16)	4.3 (1)
MMSE	29 (0.32)
ICD	1.6 (0.8)

Values represent mean (SE). BDI = Beck Depression Inventory; MMSE=Mini Mental State Examination; ICD = impulse control disorder questionnaire.

Table S2: medications

	Patients (n=13)
levodopa/carbidopa	13
Stalevo	1
Ropinirole	5
Pramipexole	2
Selegeline	2
Rasalgiline	2
Anti-hypertensives	3
Anti-depressants (SSRI/SNRI)	1
Gliclazide	1
Omeprazole	1
Ceterizine	1
Detrusitol	1
Voltarol	1
Sildefinil	1
Aspirin	1

Table S3: Drug states in which the task was carried out

Each patient returned three times to perform the task. In ‘state 1’ the subjects undertook both the first and second phases of the task in an OFF medication state. In ‘state 2’ subjects undertook both the first and second phases of the task in an ON medication state. In ‘state 3’ subjects undertook the first phase, the acquisition phase, in an OFF

medication state. Following completion of the first phase, patients then received their dopaminergic medication and undertook the second phase of the task, the performance phase, in an ON medication state. The order of the states was randomised across subjects. On all three days there was a break of 50 minutes between the first and second phases of the task to allow for dopaminergic medication to be given after the first phase in 'state 3' and to allow for adequate absorption time but to ensure consistency across all 3 days.

	Phase 1: Acquisition	**Break**	**Phase 2: Performance**
State 1	OFF	50 mins	OFF
State 2	ON	50 mins	ON
State 3	OFF	50 mins	ON

Table S4: Choice accuracy

Average choice accuracy on the training pairs after the final training session in each of the medication groups. All subjects reached at least 65% accuracy in the easiest pairing or after completion of 3 acquisition sessions had a minimum accuracy of 60%.

Subject	**ON_ON**	**OFF_ON**	**OFF_OFF**
1	77%	85%	71%
2	71%	87%	87%
3	70%	71%	65%
4	70%	62%	82%
5	97%	100%	92%
6	80%	79%	86%
7	63%	53%	84%
8	56%	62%	68%
9	94%	53%	65%
10	67%	62%	64%
11	34%	70%	65%
12	92%	63%	59%
13	75%	89%	54%

Table S5: Learning rates and softmax beta

Learning rates were fitted separately per subject in the different drug conditions. All bouts were considered equally during fitting. The data in the 'ON' group came from the 3 bouts they performed in the ON medication state; the data in the 'OFF' group came from the 6 bouts they performed in the OFF medication state.

	ON		**OFF**	
Subject	alpha	beta	alpha	beta
1	0.175	2.7	0.3221	3.2
2	0.2324	1.4	0.2351	3.3
3	0.2267	1.0	0.2012	2.7
4	0.2797	2.1	0.3064	2.1
5	0.1486	3.9	0.1779	2.3
6	0.3678	3.8	0.1712	4.8
7	0.3443	3.3	0.2851	2.9
8	0.2004	1.5	0.2348	2.6
9	0.1935	4.0	0.219	2.2
10	0.473	3.8	0.3216	8.4
11	0.2527	1.0	0.2078	0.8
12	0.1828	3.0	0.2062	2.6
13	0.175	3.9	0.3221	0.4

Table S6: Whole brain activation table

Whole brain activation table for prediction error related activity across all drug conditions during the acquisition phase of the task.

Cluster level		**Voxel level**		**Description of region**
$P_{corrected}$	K_E	*x, y, z*	*Z peak*	
0.003	90	4, -46, -12	4.77	Right cerebellum
0.347	28	-28,- 12, -2	8.17	Left posterior putamen
0.671	19	-44, 48, 4	4.13	Left inferior frontal Gyrus, Brodmann area 10
0.968	10	28, -14, 0	4.08	Right posterior putamen
0.927	12	-22, -86, 6	4.00	Left occipital lobe (white matter)
0.320	29	-20 -26 -2	4.00	Left thalamus
0.052	51	26, -4, -4	3.97	Right putamen
0.100	43	30, 60, -4	3.94	Right superior frontal gyrus, Brodmann area 10
0.995	7	40, 58, 8	3.89	Right middle frontal gyrus
0.092	44	20, -12, 8	3.8	Left thalamus
0.998	6	26, 48, -10	3.75	Right middle frontal gyrus, Brodmann area 11