Brain Advance Access published December 17, 2015

BRAIN
A JOURNAL OF NEUROLOGY

REPORT

Dopamine selectively remediates 'model-based' reward learning: a computational approach

Madeleine E. Sharp,[1] Karin Foerde,[2] Nathaniel D. Daw[3] and Daphna Shohamy[4]

Patients with loss of dopamine due to Parkinson's disease are impaired at learning from reward. However, it remains unknown precisely which aspect of learning is impaired. In particular, learning from reward, or reinforcement learning, can be driven by two distinct computational processes. One involves habitual stamping-in of stimulus-response associations, hypothesized to arise computationally from 'model-free' learning. The other, 'model-based' learning, involves learning a model of the world that is believed to support goal-directed behaviour. Much work has pointed to a role for dopamine in model-free learning. But recent work suggests model-based learning may also involve dopamine modulation, raising the possibility that model-based learning may contribute to the learning impairment in Parkinson's disease. To directly test this, we used a two-step reward-learning task which dissociates model-free versus model-based learning. We evaluated learning in patients with Parkinson's disease tested ON versus OFF their dopamine replacement medication and in healthy controls. Surprisingly, we found no effect of disease or medication on model-free learning. Instead, we found that patients tested OFF medication showed a marked impairment in model-based learning, and that this impairment was remediated by dopaminergic medication. Moreover, model-based learning was positively correlated with a separate measure of working memory performance, raising the possibility of common neural substrates. Our results suggest that some learning deficits in Parkinson's disease may be related to an inability to pursue reward based on complete representations of the environment.

1 Department of Neurology, Columbia University Medical Centre, New York NY, USA
2 Department of Psychology, New York University, New York NY, USA
3 Center for Neural Science, New York University, New York NY, USA
4 Department of Psychology and Kavli Center for Brain Sciences, Columbia University, New York NY, USA

Correspondence to: Daphna Shohamy,
Columbia University,
406 Schermerhorn Hall,
1190 Amsterdam Avenue, MC5501,
New York, NY 10027
E-mail: shohamy@psych.columbia.edu

Keywords: reinforcement learning; Parkinson's disease; dopamine; model-free; model-based

Abbreviation: UPDRS = Unified Parkinson's Disease Rating Scale

Introduction

It is widely accepted that phasic dopamine signals in the striatum play a critical role in learning to update actions based on outcomes, often referred to as reward or reinforcement learning (Houk *et al.*, 1995; Schultz *et al.*, 1997). Patients with Parkinson's disease have a severe striatal dopamine deficit, offering a test of the necessary role of

Received April 20, 2015. Revised September 26, 2015. Accepted October 4, 2015.

striatal dopamine in learning in humans. Consistent with the role of striatal dopamine in reinforcement learning, patients with Parkinson's disease have been shown to be impaired at reinforcement learning tasks (Knowlton *et al.*, 1996; Frank *et al.*, 2004; Shohamy *et al.*, 2004). However, it is becoming increasingly apparent that learning from reward is not a unitary process, and in particular that some facets of it appear to be heavily dependent on prefrontal, executive processes such as working memory (Dickinson and Balleine, 2002; Balleine and O'Doherty, 2010; Collins and Frank, 2012; Otto *et al.*, 2013*a*, *b*, 2015). This raises the question: what specific mechanisms of reinforcement learning are impaired in Parkinson's disease? Moreover, do the learning impairments specifically relate to the striatal dopaminergic deficit (Braak *et al.*, 2003; Kordower *et al.*, 2013) or do extra-striatal effects of the disease (Braak *et al.*, 2003; O'Callaghan *et al.*, 2014; Pereira *et al.*, 2014) also (or instead) contribute? Addressing these questions has important implications both for understanding the effects of dopamine treatment, as well as for understanding to what extent early cognitive deficits of Parkinson's disease, such as planning and multitasking, may share a common substrate with reward learning.

A classic dichotomy exists between two dissociable sorts of instrumental behaviours, known as habits and goal-directed actions (Dickinson and Balleine, 2002). These are often distinguished using post-training reward-devaluation procedures, with habits (but not goal-directed actions) being characteristically insensitive to devaluation. Recent advances in computational neuroscience have led to more specific hypotheses about how these two classes of behaviours are acquired, connecting these psychological categories to the computational neuroscience of learning and permitting researchers to link these computational mechanisms (such as the reinforcing influence of reward) to measurable neural (e.g. dopamine neuron firing) or behavioural (e.g. animal and human choices) outcomes during learning (Dolan and Dayan, 2013). Altogether, it is now widely recognized that reinforcement learning can be driven by two separate but concurrent processes with distinct neural substrates, with each relying on different parts of frontal cortex and striatum (Balleine and O'Doherty, 2010; Daw *et al.*, 2011).

The first process relies on reward prediction errors carried by phasic dopamine responses, which stamp in stimulus-response associations in the striatum, consistent with the predominant hypothesis for a dopaminergic role in reinforcement learning (Schultz *et al.*, 1997). Computationally, this account is known as 'model-free' learning because preferences are formed through direct experience rather than through an understanding of a model of the environment. This form of learning is hypothesized to support the learning of habits (Sutton and Barto, 1998; Daw *et al.*, 2005). Model-free learning is accompanied by a second, dissociable mechanism for 'model-based' learning, which integrates feedback with knowledge of a model of the environment. The 'model' of the environment is assumed to comprise multiple action-outcome (or, in computational terms, state-action-state) associations that represent the often-complex map of cues and actions that ultimately lead to reward. In contrast to the reflexively reinforced stimulus-response associations of model-free learning, in model-based learning, the value of candidate actions is computed more constructively from the combination of these associations and the values of the resulting outcomes.

Model-based learning has been proposed to be the computational implementation for learning goal-directed actions. Its neural substrates are less well-defined than those of model-free learning but are believed to include parts of the prefrontal cortex as well as the striatum (Daw *et al.*, 2011; Smittenaar *et al.*, 2013; Deserno *et al.*, 2015). Although model-based learning is mechanistically quite distinct from model-free learning—in particular, it does not in its standard form make use of a dopaminergic reward prediction error—it may also be sensitive to dopaminergic function. For instance, dopamine is known to support working memory through innervation of the prefrontal cortex (Sawaguchi and Goldman-Rakic, 1991), which could in turn support the manipulation of action-outcome and outcome-value representations in computing action values. Indeed, model-based learning has been shown to be sensitive to core components of executive function, such as working memory and cognitive control (Otto *et al.*, 2013*a*, *b*, 2015).

It has generally been assumed that patients with Parkinson's disease are specifically impaired at habitual learning of stimulus-response associations (Knowlton *et al.*, 1996; Shohamy *et al.*, 2004), related to impaired reward signalling in the striatum. However, although this interpretation of reinforcement learning deficits in patients with Parkinson's disease is often appealed to as key evidence supporting the hypothesis that dopaminergic reward prediction errors drive (model-free) reinforcement learning (Shohamy and Daw, 2014), most existing data are actually ambiguous on this point. This is in large part because previous work has not effectively dissociated the contributions from model-free versus model-based learning. Only one recent study has addressed the (related) habit versus goal-directed learning dichotomy in Parkinson's disease. Using an instrumental conflict task with a devaluation procedure (de Wit *et al.*, 2011), they found that, contrary to predictions, patients had preserved habitual stimulus-response learning. Based on these surprising findings, and the fact that Parkinson's disease also affects extra-striatal executive functions, which may contribute to model-based learning, such as working memory (Lange *et al.*, 1992; Owen *et al.*, 1997; Lewis *et al.*, 2005; Beato *et al.*, 2008), we applied computational methods to more closely examine whether patients with Parkinson's disease have a model-based deficit and, critically, whether this deficit is dopamine-mediated.

Table 1 Demographic and clinical characteristics of participants

	Parkinson's patients	Healthy controls	*P*-value
Age	61.1 (6.5)	62.8 (6.8)	0.4
Sex (male)	13/22	11/21	0.7
Education	17 (2)	16 (3)	0.3
MoCA	28.5 (1.3)	28.9 (0.5)	0.1
F-A-S fluency	49 (17)	52 (13)	0.5
Trails B	86 (43)	66 (24)	0.08
Stroop[a]	61 (19)	64 (17)	0.6
Digit Span total[b]	12.9 (2.4)	13.5 (2.0)	0.4
Geriatric Depression Scale	3.0 (2.5)	1.1 (1.4)	0.01
Starkstein Apathy scale	24 (6)	22 (5)	0.2
BIS-11	61 (9)	54 (7)	0.001
UPDRS OFF	18.6 (6.0)	–	–
UPDRS ON[c]	13.3 (6.0)	–	–
Disease duration	6.8 (2.9)	–	–
Daily levodopa dose (mg)	522 (235)	–	–
LEED (mg)[d]	715 (273)	–	–

Table shows mean (SD). MoCA = Montreal Cognitive Assessment; BIS-11 = Barratt Impulsiveness Scale; UPDRS = Unified Parkinson's Disease Rating Scale–Part III; LEED = Levodopa equivalent dose. *P*-values are based on *t*-tests.
[a]Stroop score calculated as difference between colour and interference stages.
[b]Digit span total = sum of forward and backward span.
[c]UPDRS ON was significantly lower than UPDRS OFF ($P < 0.001$).
[d]LEED includes levodopa, dopamine agonists, amantadine, monoamine oxidase inhibitors and catechol-*O*-methyl transferase inhibitors.

To address these questions we used a task previously shown to successfully distinguish model-free from model-based learning in both healthy and patient populations (Daw *et al.*, 2011; Eppinger *et al.*, 2013; Voon *et al.*, 2015) and tested patients with Parkinson's disease in a within-subject design, ON versus OFF dopaminergic medication.

Materials and methods

Participants

Twenty-two patients with idiopathic Parkinson's disease (13 males, mean age 61 ± 7 years) (diagnosed as per UK Brain Bank criteria) were recruited either from the Center for Parkinson's Disease and other Movement Disorders at the Columbia University Medical Center or from the Michael J Fox Foundation Trial Finder website. Twenty-one healthy control participants (11 males, mean age 63 ± 7 years) were recruited from the local community. All participants provided informed consent and were paid $12/h for their participation. The study was approved by the Institutional Review Board of Columbia University.

Patients were in the mild-to-moderate stage of disease [mean Unified Parkinson's Disease Rating Scale (UPDRS) OFF 19 ± 6, as examined by a movement disorders neurologist, disease duration: 2–14 years; Table 1]. All patients had been receiving levodopa treatment for at least 6 months (mean total daily levodopa dose 522 ± 235 mg, Supplementary Table 2). Nine patients were additionally taking dopamine agonists.

Participants completed a battery of neuropsychological tests focusing on executive function [Montreal Cognitive Assessment (MoCA), Trails A and B, Stroop, Digit Span and Phonemic word fluency] and psychiatric domains [Geriatric Depression Scale, Starkstein Apathy Scale, Barratt Impulsiveness Scale-11 (BIS-11)] (Table 1). Participants had no history of other major neurological or psychiatric disease. Participants with dementia (based on MoCA < 26) were excluded.

The patients with Parkinson's disease and controls did not differ in age, sex distribution, education, and general measures of cognition (Table 1). As expected (Weintraub *et al.*, 2006), patients scored higher on the Geriatric Depression Scale ($P = 0.01$) and on the BIS-11 ($P = 0.001$).

Procedure

Participants were tested in two sessions, in counterbalanced order: OFF, after an overnight withdrawal (>16 h, which is at least 10 half-lives for the carbidopa-levodopa and two half-lives for the dopamine agonists) of all their Parkinson's medications; and ON, 1–1.5 h after their usual dose of levodopa (mean 151 ± 61 mg). Patients who were also on regular doses of dopamine agonists did not receive them for the ON session because we wanted to isolate the effects of levodopa, which most closely mimics normal synaptic release (Pothos *et al.*, 1996). As expected, UPDRS-III scores were significantly lower when measured ON than OFF ($P < 0.0001$). Healthy control participants were also tested twice to control for practice effects. The interval between sessions ranged from 1 to 3 weeks.

Task

We used a reward learning task with a design feature that allowed for dissociation of model-free and model-based

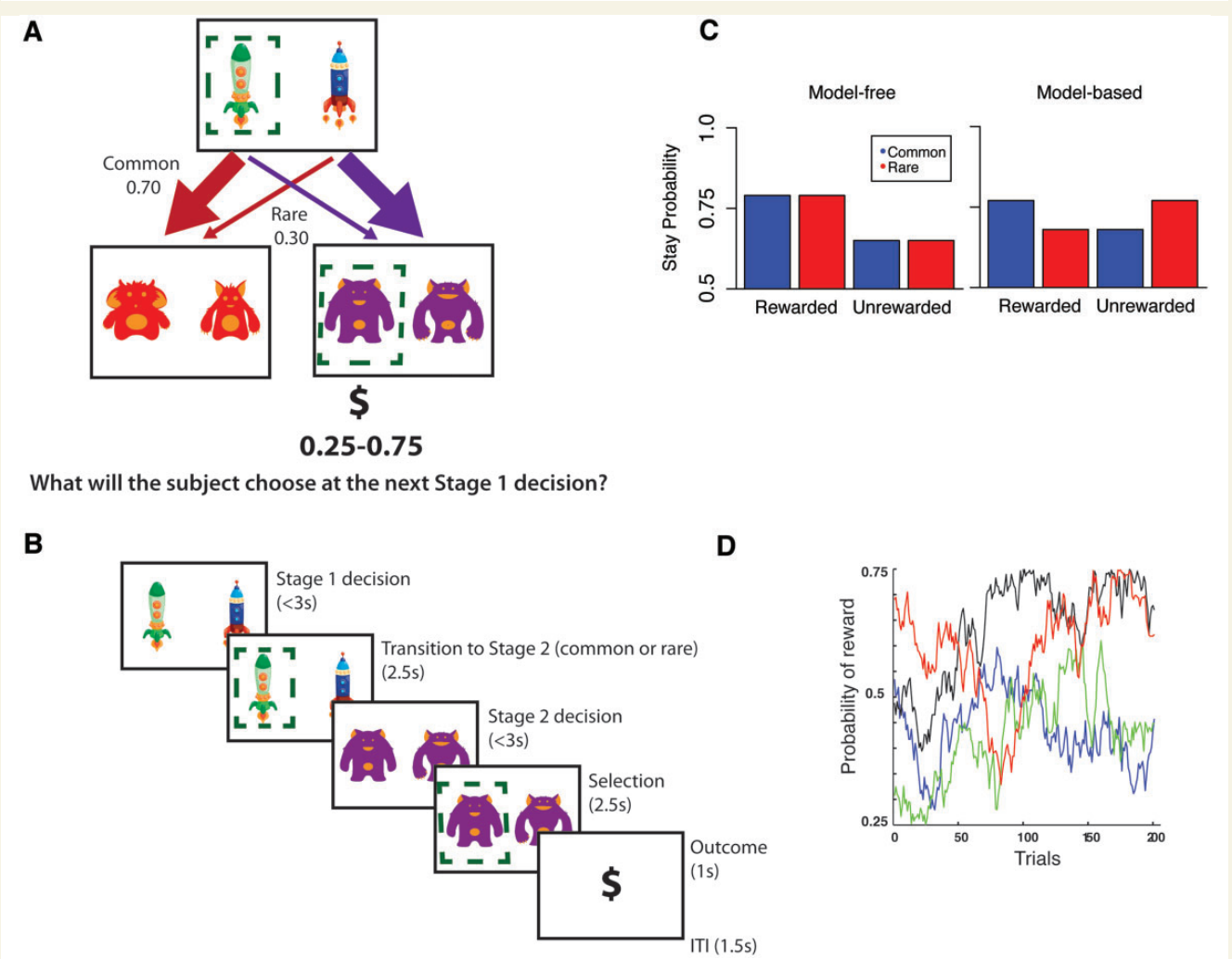


Figure 1 Trial structure, logic and simulated data for the task. (**A**) Task design. At Stage 1, participants chose between two spaceships. This choice determined the transition to the next stage according to a fixed probability scheme: each spaceship was predominantly associated with one or the other Stage 2 states (i.e. planets) and led there 70% of the time. At Stage 2, participants chose one of two aliens to discover if they would be rewarded. Each alien was associated with a probabilistic reward (range: 0.25–0.75) that changed gradually over time as determined by a Gaussian random walk. (**B**) Timing of stages within a single trial. (**C**) Simulated data. Model-based and model-free strategies predict different patterns of behaviour. A pure model-free learner (*left*) shows only a main effect of reward, choosing to stay with the same Stage 1 choice after a reward irrespective of the transition type (common versus rare) associated with that reward on the last trial. Conversely, a pure model-based learner (*right*) is influenced by the interaction between reward and transition type on the last trial. For instance, after winning on the planet less commonly visited (rare transition) a participant will be less likely to stay with the same spaceship and will instead pick the alternate spaceship in order to maximize chances of returning to that same planet. Previous studies show that healthy participants use a mixture of both model-free and model-based processes (Daw *et al.*, 2011; Eppinger *et al.*, 2013; Otto *et al.*, 2013*b*). (**D**) Example of one of the sets of Gaussian random walks that determined the probability of reward at each of the four Stage 2 options.

contributions to behaviour (Fig. 1; see also Daw *et al.*, 2011). Specifically, each trial proceeded in two stages, each of which required a decision. The first stage decision was between two spaceships destined for two different planets; this choice determined which options were presented at the second stage. The second stage represented which planet the participant visited and required a decision between two aliens on the planet. The choice of alien at this second stage led to a possible 'space gold' reward, with each piece of gold worth $0.10. The winnings were dispensed at the end of the experiment. Because the transition from the first stage choice to the second stage planet was stochastic (Fig. 1), first stage choices could be dissociated from the second stage choices that determined their ultimate reward. This allowed dissociating two learning strategies, either model-free (in which second stage rewards are associated directly with the preceding first stage decision), or model-based (in which second stage rewards and knowledge of the probabilistic structure of the task are used to infer the value of the first stage outcomes in terms of the second stage planets to which they predominantly lead). Participants were pressed to learn continually because the reward probabilities associated with each of the four second stage options were determined by independently drifting Gaussian random walks [standard deviation (SD) = 0.025] with a lower

boundary of 0.25 probability of reward and an upper boundary of 0.75, such that probability of reward from any particular second stage option changed very slowly from trial to trial (Fig. 1D).

Analysis

Stay-switch behaviour

Model-free and model-based learning can be dissociated by examining how participants adjust their first stage choices in response to feedback. Consider what happens following a trial in which a first stage choice is followed by a rare transition (to the planet less likely to follow that choice) and then reward. On the subsequent trials, repeating the same first stage choice (i.e. 'staying' with the same spaceship) indicates model-free learning, guided only by reward, and insensitivity to the probability structure (rare versus common transition, i.e. the 'model'). In contrast, making the alternate first stage choice (i.e. switching to the other spaceship commonly associated with the planet that produced the reward) indicates model-based learning because the effect of reward is mediated by the probability structure (Fig. 1B).

Accordingly, to quantify the contributions of model-free and model-based learning and the effects of disease and medication state on these contributions, we conducted a mixed effects logistic regression where the dependent variable was the probability of repeating the same first stage choice (staying = 1, switching = 0) on each trial. [Such a regression represents a simplified limiting case of fitting a more elaborate reinforcement learning model incorporating model-free and model-based components (Daw *et al.*, 2011); and is more robustly estimated for the purpose of studying within- and between-subject individual differences.] The basic within-subject (random-effect) binary explanatory variables were reward at the preceding trial (reward = 1, no reward = −1: the model-free effect, MF), transition at preceding trial (common = 1, rare = −1), and the interaction of reward and transition (reward × transition: the model-based effect, MB). We also included two binary covariates, each fully interacted with all the preceding effects: the between-subject effect of disease [disease: Parkinson's disease (PD) = 0, control = 1], and the within-subject effect of medication state (med: ON = 1, OFF = 0, where all controls were considered OFF).

This model defines the PD-OFF as the baseline group allowing comparison to PD-ON (such that the MB × med interaction reflects the effect of medications on model-based learning among disease, and the reward × med interaction reflects the effect of medication on model-free learning) and to Controls (where similarly, the MB × disease and MF × disease terms reflect the difference in model-based and model-free learning, respectively between controls and PD-OFF) (see Supplementary material for additional details). Statistical analyses were performed in R (R Development Core and Team, 2015), using the lme4 package (Bates *et al.*, 2013).

Finally, we were also interested in the association between model-based contribution to learning and certain clinical variables. Because of previous findings indicating a relationship between model-based learning and executive function (Otto *et al.*, 2013*a*)—working memory in particular (Eppinger *et al.*, 2013; Otto *et al.*, 2013*b*; Smittenaar *et al.*, 2013)—we tested for this association among our participants using total digit span score as a measure of working memory. We tested working memory only once in this study, and in the case of the patients with Parkinson's disease, it was tested during the ON session in 17/22 patients, and during the OFF session in the remaining 5/22. Because we were interested in memory capacity as a predictor of model-based learning, we included only the 17/22 patients whose memory was tested while ON in these models. We ran logistic regressions as detailed above but for each group separately (controls, patients with Parkinson's disease ON, and patients OFF) and included mean-centred digit span score as a fully interacted covariate. Running all the previous analyses with this subgroup of 17 patients yielded the same results, with no new differences in demographic or cognitive measures (Supplementary Table 3). Similarly, because it has been suggested that motor symptom severity could reflect striatal dopamine deficiency (Broussolle *et al.*, 1999), we tested for an association between UPDRS-III score and model-based contribution (we used the UPDRS-III score measured while OFF medications to model this association because it best reflects motor symptom severity). We ran a separate logistic regression in the patients with Parkinson's disease ON and OFF as above and included the mean-centred UPDRS-III score as a fully interacted covariate.

Results

The patterns of switching as a function of reward are shown, for each group and medication condition, in Fig. 2. The results for both healthy controls and patients with Parkinson's disease ON appear similar to those from healthy populations in previous studies, in that effects of both reward and transition are visible. However, patients tested OFF medication show a qualitatively different pattern of results. Quantifying these effects statistically, we found that patients with Parkinson's disease OFF showed a significant effect of reward on their behaviour indicating model-free (MF) learning ($P < 0.0001$) but no significant effect of reward × transition (i.e. MB, $P = 0.2$) indicating no detectable contribution of model-based learning. In comparison, the patients with Parkinson's disease ON showed a similar model-free contribution (MF × med: $P = 0.3$ indicating they are not different than patients OFF) but additionally showed a significantly greater contribution of model-based learning (MB × med: $P = 0.04$). The controls were also not different than patients with Parkinson's disease OFF with respect to model-free learning (MF × disease: $P = 1.0$) but, similarly to the patients ON, showed a significantly greater contribution of model-based learning (MB × disease: $P = 0.02$) than the patients with Parkinson's disease OFF. The full regression specification and coefficient estimates are reported in Table 2, and the distribution and group-level variability of the subject-specific model-free and model-based coefficients are reported in Supplementary Fig. 1 and Supplementary Table 4, respectively. In a second analysis, we fit the participants'

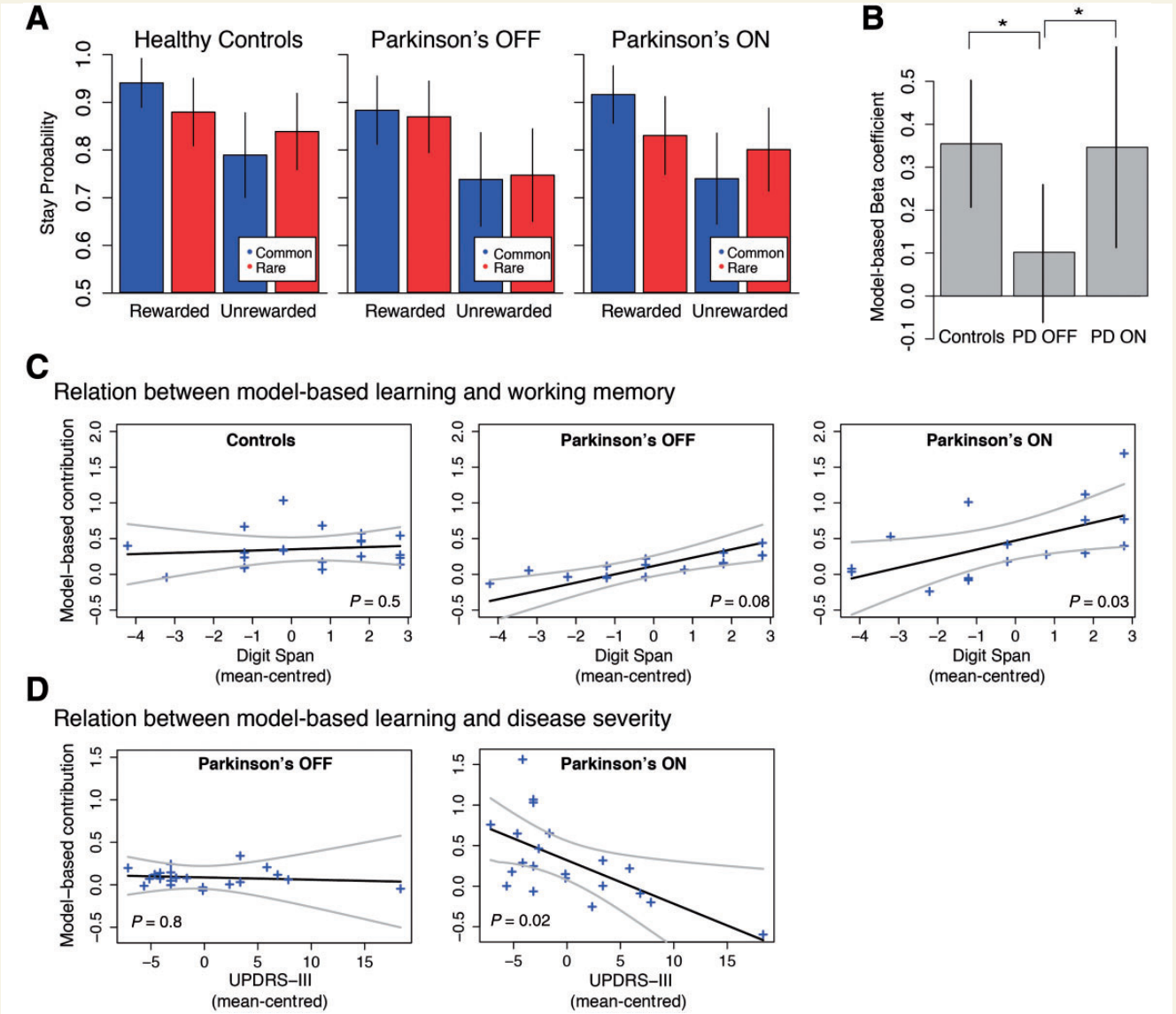


Figure 2 Effect of Parkinson's disease and dopaminergic medications on model-based learning. (**A**) Probability of repeating the same first stage choice as a function of the transition on the previous trial (common versus rare) and the outcome (rewarded versus unrewarded). Both controls and patients ON showed an effect of the interaction between reward and transition, the signature of model-based learning. In contrast, the patients OFF showed no effect of transition probability, they showed only model-free learning. (**B**) Comparison of model-based contribution to behaviour across groups as measured by the reward × transition interaction coefficient in the logistic regression (the model-based beta coefficient). Controls and patients ON showed comparable levels of model-based behaviour, whereas patients OFF showed no significant model-based contribution to their behaviour. (**C**) Relation between model-based learning and working memory: individual participants' model-based effect sizes (arbitrary units) are plotted separately for control participants and for patients OFF and ON levodopa against mean-centred total digit span score. The regression line is computed from the group-level effect of working memory. There was a significant positive effect of working memory on expression of model-based learning in the Parkinson's patients ON levodopa only. (**D**) Relation between model-based learning and disease severity—as indexed by UPDRS-III measured when OFF medications—on model-based learning: the regression line is computed from the group-level effect of UPDRS-III, where a higher score indicates more severe symptoms. There was a significant negative effect of motor symptom severity on expression of model-based learning in the patients ON medication. Error bars represent (**A**) SEM and (**B**) 95% confidence intervals. Grey lines indicate two standard errors, estimated from the group-level mixed effects regression (**C** and **D**). *P*-values are for the regression coefficient of the group-level interaction between the degree of model-based contribution to behaviour and either working memory or UPDRS-III in the respective participant groups (**C** and **D**).

choices with a more elaborate computational model of model-based and model-free learning, producing largely consistent results (Supplementary material and Supplementary Table 1).

Analysis of covariates revealed that better working memory was associated with a higher model-based contribution in patients with Parkinson's disease ON ($P = 0.03$), but did not reach significance in patients OFF ($P = 0.08$) or

Table 2 Results of the logistic regression analysis with the between participants factor Disease and the within-participant factor Medication

Predictor	Estimate (SE)	*P*-value
Intercept	1.97 (0.23)	<0.001
Reward (i.e. MF)	0.61 (0.13)	<0.001
Transition type	0.04 (0.07)	0.6
Reward × Transition (i.e. MB)[a]	0.10 (0.08)	0.2
Disease	0.42 (0.33)	0.2
Med	0.05 (0.20)	0.8
MF × Disease	0.01 (0.19)	0.97
MF × Med	−0.16 (0.14)	0.3
Transition × Disease	0.15 (0.09)	0.1
Transition × Med	0.07 (0.09)	0.5
MB × Disease[b]	0.25 (0.11)	0.02
MB × Med[c]	0.24 (0.12)	0.04

[a]The Reward × Transition interaction reflects the model-based behaviour in the PD OFF. Here, the PD OFF show no significant contribution of model-based learning.
[b]The Reward × Transition × Disease interaction reflects the model-based behaviour in the healthy controls who show a significant contribution of model-based learning compared to the PD OFF.
[c]The Reward × Transition × Med interaction reflects the model-based behaviour in the PD ON who show a significant contribution of model-based learning compared to the PD OFF.
MF = model-free learning; MB = model-based learning; PD = Parkinson's disease.

controls ($P = 0.5$) (Fig. 2C). There was no effect of working memory on model-free learning in either the patients or controls. We also found that worse motor function (i.e. a higher UPDRS-III measured OFF, reflecting worse disease severity) was associated with less model-based learning in patients with Parkinson's disease ON ($P = 0.02$) but not in patients OFF ($P = 0.8$) (Fig. 2D). The measures of mood and personality that differed between patients and controls (Geriatric Depression Scale and BIS-11) were not associated with model-based contribution to learning.

Discussion

Habit and goal-directed learning depend on neighbouring dopamine-rich striatal regions (Yin and Knowlton, 2006), but the influential prediction error theory of dopamine is thought to account only for the role of dopamine signals in habitual, model-free learning (Schultz *et al.*, 1997). Motivated by these observations, we sought to directly test whether dopamine also plays a role in model-based learning, and whether Parkinson's disease is primarily characterized by a deficit in model-free learning—as has often been assumed, or by a model-based learning deficit. We found that dopamine-deficient Parkinson's disease patients had a model-based learning deficit that was fully restored by dopamine replacement and was associated with poor working memory performance. Surprisingly, we also found that model-free learning was intact in patients and unaffected by medication state, which is consistent with one previous study showing that patients were able to form stimulus-response associations (de Wit *et al.*, 2011).

Model-based learning has been shown to rely on frontostriatal networks. Neuroimaging studies in humans have shown that the caudate, medial orbitofrontal and dorsomedial prefrontal cortex are associated with model-based learning (Daw *et al.*, 2011; Doll *et al.*, 2012; Voon *et al.*, 2015). These regions are either directly or indirectly modulated by dopamine, which could account for the strong association between model-based learning and dopamine that we observed.

Our findings are in line with previous studies showing that levodopa administered to healthy, young adults increased model-based learning (Wunderlich *et al.*, 2012). However, because the neurobiological effect of levodopa in healthy brains is unknown, it has been unclear how or where dopamine exerts this positive effect on model-based learning. One possibility is that the modulatory effect occurs in the striatum, and that model-based learning relies on a dopaminergic reward prediction error signal shared with model-free learning. Although such prediction errors are not used in standard model-based algorithms, there are variants that might explain model-based behaviours while sharing this stage (Doll *et al.*, 2012; Daw and Dayan, 2014). However, this interpretation is difficult to reconcile with the fact that model-free learning was preserved in patients withdrawn from dopaminergic medication.

Another possibility is that levodopa restored model-based learning through direct effects in prefrontal cortex. Indeed, prefrontal cortex atrophy, evident even in the early stages of Parkinson's disease (Tessa *et al.*, 2014), correlates with learning deficits (O'Callaghan *et al.*, 2013); and frontal-based executive dysfunction is also well described in the early stages of Parkinson's disease (Lange *et al.*, 1992; Cools *et al.*, 2002; Ko *et al.*, 2013; Pereira *et al.*, 2014). Furthermore, in patients, levodopa has positive effects on prefrontal cognitive functions such as working memory (Lewis *et al.*, 2005), planning (Lange *et al.*, 1992), generalization of learning (Shiner *et al.*, 2012), and task-switching (Cools *et al.*, 2001; Rutledge *et al.*, 2009). Finally, improved model-based learning on levodopa could be due to an effect on motivation through modulation of tonic rather than phasic dopamine signals (Niv, 2007; Beierholm *et al.*, 2013).

In contrast to the clear effect on model-based learning, we did not find a model-free impairment in patients with Parkinson's disease nor any effect of dopaminergic medications on model-free learning. This is surprising given the wealth of studies examining learning in patients with Parkinson's disease, where deficits are interpreted for the most part as impairments in habitual, stimulus-response learning (Knowlton *et al.*, 1996; Frank *et al.*, 2004; Shohamy *et al.*, 2004). However, the learning tasks typically used in these studies might additionally (or instead) be measuring contributions from model-based learning. For instance, participants may develop a rule-based approach and rely on explicit processes such as working memory

(Foerde *et al.*, 2006; Collins and Frank, 2012). A more recent study of learning in patients with Parkinson's disease used an instrumental conflict task modelled on those used to study habit learning in rodents, and also showed that patients with Parkinson's disease had intact stimulus-response habit learning, which relates to model-free learning (de Wit *et al.*, 2011). Though they additionally interrogated goal-directed learning with an added devaluation procedure, they did not find a significant effect of dopaminergic medications or disease on goal-directed learning.

Of course, as with any negative result, the lack of effects reported here on model-free learning must be interpreted with caution. The current study cannot rule out the possibility that an impairment of model-free learning also exists in the patients with Parkinson's disease but is not adequately operationalized by our task. For instance, although we detect substantial model-free learning with a level of individual variability comparable to that for model-based (Supplementary Fig. 1), it is possible that this sort of learning might itself be heterogeneous in the brain, with the sort hypothesized to involve dopaminergic action in striatum not dominating our measure. Indeed, the link between the computational mechanism of model-free learning and the psychological category of habits is also more controversial than that between model-based and goal-directed learning (Dezfouli and Balleine, 2012), though on that account (according to which habitual behaviour and seemingly model-free behaviour on the sequential decision task used here are proposed to arise from a common choice mechanism which is ultimately model-based) it is not clear why we see differential effects of disease and medication on model-based and model-free choices. It is also possible that model-free learning is a very robust cognitive function, in keeping with its role as a faster but less accurate system (Keramati *et al.*, 2011), and thus that model-free learning, but not model-based, can withstand a certain degree of dopamine deficiency. Future work will be required to more finely probe the relationship, if any, between model-free learning, habits, and striatal dopaminergic function.

Our finding, of an association between model-based learning and working memory capacity in patients with Parkinson's disease, helps bridge two seemingly independent constructs of impaired cognition in Parkinson's disease—reward learning and executive function. Previous studies have focused on dissociating these domains on the basis of hypothesized separate neural substrates for each—striatal dopamine for reward learning (Schultz *et al.*, 1997; Frank *et al.*, 2004; Schonberg *et al.*, 2010), and frontostriatal (Cools *et al.*, 2002; Lewis *et al.*, 2003, 2005; Monchi *et al.*, 2007; Ko *et al.*, 2013; Nagano-Saito *et al.*, 2014) and extrastriatal non-dopaminergic networks for executive function (Hilker *et al.*, 2005; Weintraub *et al.*, 2010; Ye *et al.*, 2014, 2015). Our finding that increased working memory capacity was associated with a greater contribution of model-based learning in patients with Parkinson's disease but not in healthy controls could suggest that working memory is recruited as a compensatory mechanism to support model-based learning. Similarly, in healthy younger adults, better working memory predicts model-based performance (Eppinger *et al.*, 2013; Otto *et al.*, 2013*b*; Smittenaar *et al.*, 2013). However, similar to our results here, the effect of working memory most often became apparent only under challenge; in particular, better working memory predicted robustness to stress or transcranial magnetic stimulation. In a study of ageing in healthy participants, the relationship between working memory and model-based performance was age-related: while present in younger adults, working memory was not related to model-based performance among healthy older adults (Eppinger *et al.*, 2013) (again, consistent with our results). More generally, normal prefrontal cortex function (Smittenaar *et al.*, 2013) and executive functioning (Otto *et al.*, 2013*a*, 2015) are necessary for model-based learning. These findings support the idea that the neural substrates underlying reward learning and executive function are co-dependent rather than independent. A main limitation of this study is that we cannot directly comment on the neural substrates of the model-based learning deficit or on the likely crucial role that dopamine plays in modulating striatal-prefrontal cortex connections to support model-based learning.

In conclusion, our findings demonstrate that learning deficits in Parkinson's are not merely a matter of reduced reward prediction but, rather, are related to an inability to pursue reward based on more complete representations of the environment. These findings emerged from a computational characterization of these two forms of learning and are generally consistent with previous results from a different task (de Wit *et al.*, 2011). Together, these findings challenge the general assumption that patients with Parkinson's disease cannot form stimulus-response associations and therefore cannot form habits (Knowlton *et al.*, 1996). This is congruent with observations of patients with Parkinson's disease who function well in simple environments where they maintain an ability to respond to informative cues, but have marked difficulties with executive functions such as multi-tasking and planning, which depend on the building and maintenance of a model of the environment (Brown and Marsden, 1991). An important goal for future research, especially in the context of treatment of cognitive symptoms in Parkinson's disease, is to further understand the interdependence of striatal learning processes and executive function, and to identify where exogenous dopamine is exerting its positive effect on model-based learning so we can target treatments to these regions while avoiding possible detrimental effects of medications on others (Cools *et al.*, 2001).

Acknowledgements

We thank Cate Hartley for the task and stimuli; Ross Otto, Bradley Doll, Katherine Duncan for assistance with analyses; Caroline Marvin, Kendall Braun and Peter Myers for help with participant testing and Stanley Fahn, Roy

Alcalay and Cheryl Waters for help with patient recruitment. We also thank the patients and controls who participated in the study and the Clinical Trials Transportation Program for their assistance.

Funding

This work was supported by National Institutes of Health R01DA038891 (D.S. and N.D.D.) and a University of British Columbia Clinician Investigator Award (M.E.S.).

Supplementary material

Supplementary material is available at *Brain* online.

References

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

Bates D, Maechler M, Bolker B, Walker S, Christensen RHB, Singmann H, et al. Linear mixed-effects models using 'Eigen' and S4. Version 1.0. 2013.

Beato R, Levy R, Pillon B, Vidal C, du Montcel ST, Deweer B, et al. Working memory in Parkinson's disease patients: clinical features and response to levodopa. Arq Neuropsiquiatr 2008; 66: 147–51.

Beierholm U, Guitart-Masip M, Economides M, Chowdhury R, Duzel E, Dolan R, et al. Dopamine modulates reward-related vigor. Neuropsychopharmacology 2013; 38: 1495–503.

Braak H, Del Tredici K, Rüb U, de Vos RAI, Jansen Steur ENH, Braak E. Staging of brain pathology related to sporadic Parkinson's disease. Neurobiol Aging 2003; 24: 197–211.

Broussolle E, Dentresangle C, Landais P, Garcia-Larrea L, Pollak P, Croisile B, et al. The relation of putamen and caudate nucleus 18F-Dopa uptake to motor and cognitive performances in Parkinson's disease. J Neurol Sci 1999; 166: 141–51.

Brown RG, Marsden CD. Dual task performance and processing resources in normal subjects and patients with Parkinson's disease. Brain 1991; 114 (Pt 1A): 215–31.

Collins AG, Frank MJ. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. Eur J Neurosci 2012; 35: 1024–35.

Cools R, Barker RA, Sahakian BJ, Robbins TW. Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. Cereb Cortex 2001; 11: 1136–43.

Cools R, Stefanova E, Barker RA, Robbins TW, Owen AM. Dopaminergic modulation of high-level cognition in Parkinson's disease: the role of the prefrontal cortex revealed by PET. Brain 2002; 125: 584–94.

Daw ND, Dayan P. The algorithmic anatomy of model-based evaluation. Philos Trans R Soc Lond B Biol Sci 2014; 369.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron 2011; 69: 1204–15.

Daw ND, Niv Y, Dayan P. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. Nat Neurosci 2005; 8: 1704–11.

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.

Deserno L, Huys QJM, Boehme R, Buchert R, Heinze H-J, Grace AA, et al. Ventral striatal dopamine reflects behavioral and neural signatures of model-based control during sequential decision making. Proc Natl Acad Sci USA 2015; 112: 1595–600.

Dezfouli A, Balleine BW. Habits, action sequences and reinforcement learning. Eur J Neurosci 2012; 35: 1036–51.

Dickinson A, Balleine B. The role of learning in the operation of motivational systems. In: Gallistel CR, editor. Steven's handbook of experimental psychology: learning, motivation and emotion. 3rd ed. New York: John Wiley & Sons; 2002. p. 497–534.

Dolan RJ, Dayan P. Goals and Habits in the Brain. Neuron 2013; 80: 312–25.

Doll BB, Simon DA, Daw ND. The ubiquity of model-based reinforcement learning. Curr Opin Neurobiol 2012; 22: 1075–81.

Eppinger B, Walter M, Heekeren HR, Li S-C. Of goals and habits: age-related and individual differences in goal-directed decision-making. Front Neurosci 2013; 7: 253.

Foerde K, Knowlton BJ, Poldrack RA. Modulation of competing memory systems by distraction. Proc Natl Acad Sci USA 2006; 103: 11778–83.

Frank MJ, Seeberger LC, O'reilly RC. By carrot or by stick: cognitive reinforcement learning in parkinsonism. Science 2004; 306: 1940–3.

Hilker R, Thomas AV, Klein JC, Weisenbach S, Kalbe E, Burghaus L, et al. Dementia in Parkinson disease: functional imaging of cholinergic and dopaminergic pathways. Neurology 2005; 65: 1716–22.

Houk J, Adams J, Barto A. A model of how the basal ganglia generates and uses neural signals that predict reinforcement. In: Houk J, Davis J, Beiser D, editors. Models of information processing in the basal ganglia. Cambridge: MIT Press; 1995. p. 249–70.

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

Knowlton BJ, Mangels JA, Squire LR. A neostriatal habit learning system in humans. Science 1996; 273: 1399–402.

Ko JH, Antonelli F, Monchi O, Ray N, Rusjan P, Houle S, et al. Prefrontal dopaminergic receptor abnormalities and executive functions in Parkinson's disease. Hum Brain Mapp 2013; 34: 1591–604.

Kordower JH, Olanow CW, Dodiya HB, Chu Y, Beach TG, Adler CH, et al. Disease duration and the integrity of the nigrostriatal system in Parkinson's disease. Brain 2013; 136: 2419–31.

Lange KW, Robbins TW, Marsden CD, James M, Owen AM, Paul GM. L-dopa withdrawal in Parkinson's disease selectively impairs cognitive performance in tests sensitive to frontal lobe dysfunction. Psychopharmacology 1992; 107: 394–404.

Lewis SJ, Slabosz A, Robbins TW, Barker RA, Owen AM. Dopaminergic basis for deficits in working memory but not attentional set-shifting in Parkinson's disease. Neuropsychologia 2005; 43: 823–32.

Lewis SJG, Dove A, Robbins TW, Barker RA, Owen AM. Cognitive impairments in early Parkinson's disease are accompanied by reductions in activity in frontostriatal neural circuitry. J Neurosci 2003; 23: 6351–6.

Monchi O, Petrides M, Mejia-Constain B, Strafella AP. Cortical activity in Parkinson's disease during executive processing depends on striatal involvement. Brain 2007; 130: 233–44.

Nagano-Saito A, Habak C, Mejia-Constain B, Degroot C, Monetta L, Jubault T, et al. Effect of mild cognitive impairment on the patterns of neural activity in early Parkinson's disease. Neurobiol Aging 2014; 35: 223–31.

Niv Y. Cost, benefit, tonic, phasic: what do response rates tell us about dopamine and motivation? Ann N Y Acad Sci 2007; 1104: 357–76.

O'Callaghan C, Moustafa AA, de Wit S, Shine JM, Robbins TW, Lewis SJG, et al. Fronto-striatal gray matter contributions to discrimination learning in Parkinson's disease. Front Comput Neurosci 2013; 7.

O'Callaghan C, Shine JM, Lewis SJG, Hornberger M. Neuropsychiatric symptoms in Parkinson's disease: fronto-striatal atrophy contributions. Parkinsonism Relat Disord 2014; 20: 867–72.

Otto AR, Gershman SJ, Markman AB, Daw ND. The curse of planning: dissecting multiple reinforcement-learning systems by taxing the central executive. Psychol Sci 2013a; 24: 751–61.

Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND. Working-memory capacity protects model-based learning from stress. Proc Natl Acad Sci USA 2013b; 110: 20941–6.

Otto AR, Skatova A, Madlon-Kay S, Daw ND. Cognitive control predicts use of model-based reinforcement learning. J Cogn Neurosci 2015; 27: 319–33.

Owen AM, Iddon JL, Hodges JR, Summers BA, Robbins TW. Spatial and non-spatial working memory at different stages of Parkinson's disease. Neuropsychologia 1997; 35: 519–32.

Pereira JB, Svenningsson P, Weintraub D, Bronnick K, Lebedev A, Westman E, et al. Initial cognitive decline is associated with cortical thinning in early Parkinson disease. Neurology 2014; 82: 2017–25.

Pothos E, Desmond M, Sulzer D. L-3,4-dihydroxyphenylalanine increases the quantal size of exocytotic dopamine release in vitro. J Neurochem 1996; 66: 629–36.

R Development Core and Team. R: A language and environment for statistical computing. Version 3.1.3. 2015.

Rutledge RB, Lazzaro SC, Lau B, Myers CE, Gluck MA, Glimcher PW. Dopaminergic drugs modulate learning rates and perseveration in Parkinson's patients in a dynamic foraging task. J Neurosci 2009; 29: 15104–14.

Sawaguchi T, Goldman-Rakic PS. D1 dopamine receptors in prefrontal cortex: involvement in working memory. Science (New York, NY) 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, Dayan P, Montague PR. A neural substrate of prediction and reward. Science 1997; 275: 1593–9.

Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P, et al. Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. Brain 2012; 135: 1871–83.

Shohamy D, Daw N. Habits and reinforcement learning. In: Gazzaniga M, Mangun G, editors. Cognitive neurosciences. 5th ed. Cambridge: MIT Press; 2014. p. 591–604.

Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA. Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain 2004; 127: 851–9.

Smittenaar P, FitzGerald THB, Romei V, Wright ND, Dolan RJ. Disruption of dorsolateral prefrontal cortex decreases model-based in favor of model-free control in humans. Neuron 2013; 80: 914–9.

Sutton RS, Barto AG. Reinforcement learning: an introduction. IEEE Trans Neural Netw 1998; 9: 1054.

Tessa C, Lucetti C, Giannelli M, Diciotti S, Poletti M, Danti S, et al. Progression of brain atrophy in the early stages of Parkinson's disease: a longitudinal tensor-based morphometry study in de novo patients without cognitive impairment. Hum Brain Mapp 2014; 35: 3932–44.

Voon V, Derbyshire K, Ruck C, Irvine MA, Worbe Y, Enander J, et al. Disorders of compulsivity: a common bias towards learning habits. Mol Psychiatry 2015; 20: 345–52.

Weintraub D, Mavandadi S, Mamikonyan E, Siderowf AD, Duda JE, Hurtig HI, et al. Atomoxetine for depression and other neuropsychiatric symptoms in Parkinson disease. Neurology 2010; 75: 448–55.

Weintraub D, Oehlberg KA, Katz IR, Stern MB. Test characteristics of the 15-item geriatric depression scale and Hamilton depression rating scale in Parkinson disease. Am J Geriatr Psychiatry 2006; 14: 169–75.

Wunderlich K, Smittenaar P, Dolan RJ. Dopamine enhances model-based over model-free choice behavior. Neuron 2012; 75: 418–24.

Ye Z, Altena E, Nombela C, Housden CR, Maxwell H, Rittman T, et al. Improving response inhibition in Parkinson's disease with atomoxetine. Biol Psychiatry 2015; 77: 740–8.

Ye Z, Altena E, Nombela C, Housden CR, Maxwell H, Rittman T, et al. Selective serotonin reuptake inhibition modulates response inhibition in Parkinson's disease. Brain 2014; 137: 1145–55.

Yin HH, Knowlton BJ. The role of the basal ganglia in habit formation. Nat Rev Neurosci 2006; 7: 464–76.

Supplementary materials

Impaired goal-directed learning in Parkinson's disease

Madeleine E Sharp[1], Karin Foerde[2], Nathaniel D Daw[3], Daphna Shohamy[4]

[1] Department of Neurology, Columbia University Medical Centre, New York NY
[2] Department of Psychology, New York University, New York NY
[3] Center for Neural Science, New York University, New York NY
[4] Department of Psychology & Kavli Center for Brain Sciences, Columbia University, New York NY

Regression model specification

We used a single mixed-effects logistic regression model to capture the differences between Parkinson's patients and controls and between patients ON and OFF medications. This approach excels at handling unbalanced designs (see, e.g., Gelman & Hill, 2006) by explicitly characterizing the structure of between- and within-subjects factors. As we describe below, this models the behavior of each of the three group/condition combinations separately and estimates parameters testing the two comparisons of interest.

Predictor coding:
We used effect coding (1/-1) for the behavioural terms of interest. This was necessary so that we could interpret the interaction term Reward*Transition.
Reward: 1=reward, -1=no reward
Transition: 1=Common, -1=Rare

We used dummy coding for the group variables Disease and Medication.
Disease: 0=Parkinson's, 1=Controls
Med: 0=OFF, 1=ON

This arrangement of dummy coding ensured that when running the full model that included all three group/condition combinations (controls OFF, PD ON and PD OFF) the PD OFF would be the baseline group and the main effects of the predictors Disease and Med would represent the difference between the PD OFF and Controls (in the case of Disease term) and the difference between the PD OFF and PD ON (in the case of Med term).

Regression model:
We will go through the model simplification for each group to demonstrate the logic behind our predictor coding:
(Note that the Reward*Transition term (i.e. Model-based) is simplified to MB in the following equations)

Full model:
P(Stay)= Reward + Transition + MB + Dis + Med + Reward*Dis + Reward*Med + Transition*Dis + Transition*Med + MB*Dis + MB*Med

PD OFF: (Dis=0, Med=0)
P(Stay)= Reward + Transition + MB + ~~Dis~~ + ~~Med~~ + ~~Reward*Dis~~ + ~~Reward*Med~~ + ~~Transition*Dis~~ + ~~Transition*Med~~ + ~~MB*Dis~~ + ~~MB*Med~~
P(Stay)= Won + Rare +MB
So the 'main effect' terms for Reward and MB are really the average Model-free and Model-based contributions *among the PD-OFF only.*

PD ON: (Dis=0, Med=1)
P(Stay)= Reward + Transition + MB + ~~Dis~~ + Med + ~~Reward*Dis~~ + **Reward*Med** + ~~Transition*Dis~~ + Transition*Med + ~~MB*Dis~~ + **MB*Med**
So the difference in Model-Based contribution between PD-ON and PD-OFF is captured by MB*Med term, and the difference in Model-Free between these 2 groups is captured by the Reward*Med term since these are the only new terms when compared to the equation for the baseline group (PD-OFF).
The main effect of Med is not of interest since it indicates only the effect of Med on 'staying'.

Controls: (Dis=1, Med=0)
P(Stay)= Reward + Transition + MB + Dis + ~~Med~~ + **Reward*Dis** + ~~Reward*Med~~ + Transition*Dis + ~~Transition*Med~~ + **MB*Dis** + ~~MB*Med~~
Similarly, the difference in Model-Based contribution between Controls and PD-OFF is captured by MB*Dis term, and the difference in Model-Free between these 2 groups is captured by the Reward*Dis term.
The main effect of Dis is also of no interest since all it indicates is the effect of Dis on 'staying'.

Computational model fitting analysis

Full model-based and model-free learning algorithms learn the value of each option incrementally over multiple trials. The logistic regressions reported so far simplify this to examine only the effects of the most recent trial's experience on each choice, which model simulations show are characteristic of these two sorts of learning and which results in a robust, factorial analysis (Daw *et al.*, 2011). To verify that the effects we report are robust to this simplification, we also examined the effects of patient status and medication in the fits to participant behavior of a full computational model in which model-based and model-free learning take into account the effects of experience over all previous trials (Daw *et al.*, 2011).

The model is a version of the one from Daw et al. (2011), modified slightly for better estimation in the individual differences setting (similar to (Otto *et al.*, 2013)). We estimated the model using Markov Chain Monte Carlo (MCMC) inference, which allows drawing samples from the posterior distribution of parameters in a generative model of data, conditional on the data. We first describe the generative model.

Subject-level model

For a subject choosing option $c_{1,t}$ (= 1 or 2) in first stage state 1 at trial t, transitioning to second-stage state s_t (= 2 or 3), where they choose option $c_{2,t}$ and receive reward r_t, the model free learner learns a function $Q^{MF}(s, c)$ measuring the value of each option in each state. This is updated after each trial's experience as:

$$Q_{t+1}^{MF}(1, c_{1,t}) = (1-\alpha)Q_t^{MF}(1, c_{1,t}) + r_t$$
$$Q_{t+1}^{MF}(s_t, c_{2,t}) = (1-\alpha)Q_t^{MF}(s_t, c_{2,t}) + r_t$$

where $\alpha \in [0,1]$ is a free learning rate parameter.

The model-based learner computes the value $Q^{MB}(1, c)$ of the first level options from the value of the resulting second-level states as

$$Q^{MB}(1, c_i) = \mathrm{argmax_a} Q^{MF}(s_i, a)$$

for both actions i (= 1,2), where s_i is the state most often visited previously after choice of c_i at the first stage.

We model the subjects' first and second level choice probabilities using a logistic softmax in the various estimated values. At the first stage this is a weighted sum of model-based and model free values (weighted according to free parameters β^{MF} and β^{MB}):

$$P(c_{1,t} = c_i) \propto \exp\left(\beta^{MF} * Q_t^{MF}(1, c_i) + \beta^{MB} * Q_t^{MB}(1, c_i) + \beta^p I(c_{1,t-1} = c_i)\right)$$

where the third term encodes a bias with weight β^p toward sticking with or switching from the option chosen on the previous trial. (Here $I(c_{1,t-1} = c_i)$ is a binary indicator for the previous choice.)

Since model-based and model-free learning coincide at the second stage, choice depends only on Q_t^{MF}, with free weight β^2.

$$P(c_{2,t} = c_i) \propto \exp(\beta^2 * Q_t^{MF}(s_t, c_i))$$

Thus, for each subject s (and medication condition m), this model has five free parameters: four softmax weights ($\beta^{MF}, \beta^{MB}, \beta^p, \beta^2$) and a learning rate α. This model is a slightly simplified version of that from Daw et al. (2011), after Otto et al. (2013). Notably, the relative model-based vs model-free weighting parameter w and overall first-stage softmax weight β have been equivalently re-parameterized as two weights (with $\beta^{MF} = \beta \cdot (1-w)$ and $\beta^{MB} = \beta \cdot w$); the eligibility trace parameter λ (which proved to be near 1 and difficult to estimate in this dataset) has been fixed to 1; and the update rules for Q^{MF} and Q^{MB} have been rescaled by dividing the r_t term by the learning rate α (which simply rescales the Qs and the softmax weight parameters

to values that are more robust to variation in α over subjects; (Camerer and Ho, 1999, Otto *et al.*, 2013)).

Group-level model

We nested the subject-level model inside a group-level model to capture population variation and test for differences across conditions and groups. Each of the five free parameters was taken as instantiated for each subject and each medication condition from a group-level distribution over the population, whose parameters were themselves estimated. In particular, the four softmax weight parameters β (e.g., β^{MB} were taken as Gaussian over the subjects, each with a group-level mean (e.g., μ^{MB}) and standard deviation (σ^{MB}). The group level distribution on the learning rate α (which is bounded by [0, 1]) was taken as a beta distribution $\alpha_s \sim Beta(a, b)$, whose two parameters were reparameterized using the change of variables $\mu_\alpha = \frac{a}{a+b}$ and $s_\alpha = \frac{1}{\sqrt{a+b}}$, (which capture the more interpretable characteristics of central tendency and variability, respectively).

As in the logistic regression model, we tested the possibility of disease- or medication- related variation in the two parameters of interest, β^{MB} and β^{MF}, by including additional terms (k^{MB-PD} and k^{MB-med} and similarly for MF) coding the effect of disease and medication. Altogether, for subject s and medication condition m this entailed:

$$\beta_{s,m}^{MB} \sim N(\mu^{MB} + k^{MB-PD} I(s, \text{'healthy'}) + k_s^{MB-med} I(m, \text{'on'}), \sigma^{MB})$$

and similarly for $\beta_{s,m}^{MF}$. Here the terms $I(\cdot)$ are binary indicators for healthy subjects and on-medication sessions. The medication effect (because it is within-subject) is itself a subject-specific random variable with its own population-level mean and variance, which are themselves inferred ($k_s^{MB-med} \sim N\left(k^{MB-med}, \sigma^{MB-med}\right)$.

We did not include these additional terms for the remaining betas (β^2 and β^p) nor for the learning rate, since our hypotheses concerned MB and MF learning (though note that intersubject variation in these parameters is still captured by the model, just not systematically tested by group or condition). Thus, for instance, $\beta_{s,m}^{p} \sim N(\mu^p, \sigma^p)$.

Altogether the model included two parameters for the learning rates (μ^α, s^α), and, for the logistic softmax parameters, four means $\mu^{MB}, \mu^{MF}, \mu^p, \mu^2$, four group or medication effects $k^{MB-PD}, k^{MF-PD}, k^{MB-med}, k^{MF-med}$, and six variances $\sigma^{MB}, \sigma^{MF}, \sigma^p, \sigma^2, \sigma^{MB-med}, \sigma^{MF-med}$. Finally, we specified hyperpriors on these parameters. Hyperpriors on the softmax means μ and effects k were each taken as as $Cauchy(0,2)$, which are uninformative within the range of parameters estimated in previous studies, and heavy-tailed to allow for further variation. Hyperpriors on variances σ were taken as half-Cauchy (i.e., the same distribution, truncated to exclude negative variances). Finally, the hyperprior for the learning rate mean μ^α was $Uniform(0,1)$ and its inverse sample size s^α was given an improper infinite uniform prior over positive values (as recommended for this parameterization by (Gelman *et al.*, 2003)).

Estimation

We implemented the model using the Stan programming language (Stan Development Team, 2015a,b). We ran four chains of 1250 samples each (discarding the first 250 for burn-in), and verified convergence using manual examination of the traceplots and of the potential scale reduction factor $\hat{r}$ (Gelman and Rubin, 1992), which was below 1.03 for all parameters.

Results

Table S1 reports medians and symmetric 95% posterior intervals (known in Bayesian inference as credible intervals and similar to confidence intervals), drawn from the quantiles of the samples for each of the group-level parameters. The main question is whether the estimated disease and medication effects k on model-based learning are nonzero. (Positive values for these effects indicate that model based learning is larger for controls than patients and larger for patients on relative to off medication.) The table reports P (twice the posterior probability that $k < 0$, or equivalently 1 minus the size of the largest symmetric confidence interval excluding 0, roughly comparable to a two-tailed P-value) for each of these effects. The results are broadly consistent with those reported for the logistic regression analysis. In particular, we can with more than 95% confidence exclude zero for the effect of disease on model-based learning. In this analysis, the effect of medication is nonzero with 94. 7% confidence (in classical terms, a trend). We find no significant effects of either disease or medication on model-free learning.

Table S1. Median and 95% credible intervals for the group-level parameters from the computational model, drawn from quantiles of the samples of each parameter from MCMC estimation.

	Median	Lower CI	Upper CI	*P*
$\boldsymbol{\mu^{MB}}$	.09	-.04	.21	.17
$\boldsymbol{\mu^{MF}}$	.49	.30	.68	.00
$\boldsymbol{\mu^{p}}$	1.66	1.39	1.93	.00
$\boldsymbol{\mu^{2}}$	.78	.67	.90	.00
$\boldsymbol{k^{MB-PD}}$	.18	.01	.35	.04
$\boldsymbol{k^{MF-PD}}$	-.01	-.28	.25	.96
$\boldsymbol{k^{MB-med}}$	.13	.00	.27	.05
$\boldsymbol{k^{MF-med}}$	-.11	-.31	.07	.22
$\boldsymbol{\sigma^{MB}}$	.24	.17	.33	
$\boldsymbol{\sigma^{MF}}$	.40	.31	.53	
$\boldsymbol{\sigma^{p}}$	.91	.72	1.16	
$\boldsymbol{\sigma^{2}}$	.36	.28	.46	
$\boldsymbol{\sigma^{MB-med}}$	.21	.07	.39	
$\boldsymbol{\sigma^{MF-med}}$	.34	.22	.55	
$\boldsymbol{\mu^{\alpha}}$	.59	.52	.66	
$\boldsymbol{s^{\alpha}}$	.52	.40	.68	

Table S2. Dopaminergic medications taken in addition to levodopa

	Patients # (%)
Dopamine agonist[1]	9 (0.41)
MAOI[2]	9 (0.41)
COMT[3]	4 (0.18)
Amantadine	6 (0.27)

[1]Dopamine agonists included pramipexole (6/9), ropinirole (3/9)
[2]MAOI, monoamine oxidase inhibitor included rasagiline (7/9) and selegiline (2/9)
[3]COMT, catechol-*O*-methyl transferase inhibitors included entacapone either as stand-alone or as part of Stalevo

Table S3. Demographic and clinical characteristics of participants who were included in the analyses of working memory

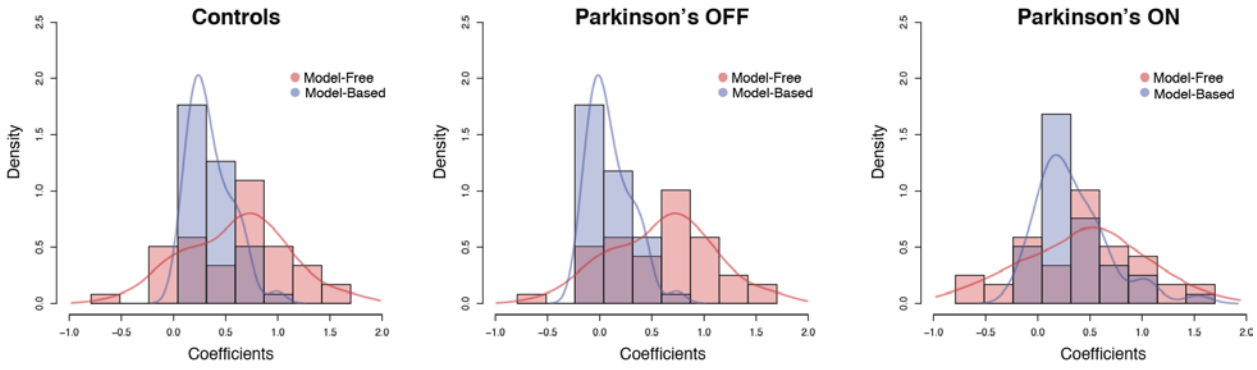
	Parkinson's patients n=17	**Healthy Controls n=21**	***p*-value**
Age	62.2 (6.7)	62.8 (6.8)	0.8
Sex (Male)	10/17	11/21	0.7
Education	17 (2)	16 (3)	0.6
MoCA	28.5 (1.3)	28.9 (0.5)	0.4
F-A-S fluency	47 (17)	52 (13)	0.3
Trails B	91 (47)	66 (24)	0.04
Stroop[1]	58 (20)	64 (17)	0.3
Digit Span total[2]	13.0 (2.3)	13.5 (2.0)	0.5
Geriatric Depression Scale	3.1 (2.6)	1.1 (1.4)	0.01
Starkstein Apathy scale	24 (6)	22 (5)	0.3
BIS-11	62 (9)	54 (7)	0.004
UPDRS OFF	18.6 (6.4)	--	--
UPDRS ON[3]	13.3 (6.3)	--	--
Disease duration	7.2 (2.9)	--	--
Daily levodopa dose (mg)	499 (219)	--	--
LEED (mg)[4]	728 (277)	--	--

Table shows mean (SD). MoCA, Montreal Cognitive Assessment; BIS-11, Barratt Impulsiveness Scale; UPDRS, Unified Parkinson's Disease Rating Scale–Part III; LEED, Levodopa equivalent dose. *p*-values are based on t-tests.
[1]Stroop score calculated as difference between color and interference stages
[2] Digit span total = sum of forward and backward span
[3] UPDRS ON was significantly lower than UPDRS OFF ($p<0.001$)
[4]LEED includes levodopa, dopamine agonists, amantadine, monoamine oxidase inhibitors and catechol-*O*-methyl transferase inhibitors

Table S4. Standard deviations of subject-level estimates for model-free and model-based learning

Group	Model-Free	Model-Based
Controls	0.57	0.23
Parkinson's OFF	0.42	0.19
Parkinson's ON	0.60	0.45

Figure S1. Distribution of subject-specific model coefficients for model-free and model-based learning



References:

Camerer C, Ho TH. Experience-weighted Attraction Learning in Normal Form Games. Econometrica. 1999; 67: 827-74.

Daw ND, Gershman SJ, Seymour B, Dayan P, Dolan RJ. Model-based influences on humans' choices and striatal prediction errors. Neuron. 2011; 69: 1204-15.

Gelman A, Carlin JB, Stern HS, Rubin DB. Bayesian data analysis. 2nd ed. London: Chapman and Hall; 2003.

Gelman A, Rubin DB. Inference from iterative simulation using multiple

sequences. Stat Sci. 1992; 7: 457-72.

Otto AR, Raio CM, Chiang A, Phelps EA, Daw ND. Working-memory capacity protects model-based learning from stress. Proc Natl Acad Sci USA. 2013; 110: 20941-6.

Stan Development Team. Stan: A C++ Library for Probability and Sampling, Version 2.7.0.; 2015.

Stan Development Team. Stan Modeling Language Users Guide and Reference Manual, Version 2.7.0.; 2015.