Neuroscience 163 (2009) 1092–1101

PROBABILISTIC REVERSAL LEARNING IS IMPAIRED IN PARKINSON'S DISEASE

D. A. PETERSON,[a] C. ELLIOTT,[a] D. D. SONG,[b] S. MAKEIG,[c] T. J. SEJNOWSKI[a,d] AND H. POIZNER[a]*

[a]*Institute for Neural Computation, UCSD*

[b]*Department of Neurosciences, School of Medicine, UCSD*

[c]*Swartz Center for Computational Neuroscience, Institute for Neural Computation, UCSD*

[d]*Salk Institute for Biological Studies*

Abstract—In many everyday settings, the relationship between our choices and their potentially rewarding outcomes is probabilistic and dynamic. In addition, the difficulty of the choices can vary widely. Although a large body of theoretical and empirical evidence suggests that dopamine mediates rewarded learning, the influence of dopamine in probabilistic and dynamic rewarded learning remains unclear. We adapted a probabilistic rewarded learning task originally used to study firing rates of dopamine cells in primate substantia nigra pars compacta [Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9:1057–1063] for use as a reversal learning task with humans. We sought to investigate how the dopamine depletion in Parkinson's disease (PD) affects probabilistic reward learning and adaptation to a reversal in reward contingencies. Over the course of 256 trials subjects learned to choose the more favorable from among pairs of images with small or large differences in reward probabilities. During a subsequent otherwise identical reversal phase, the reward probability contingencies for the stimuli were reversed. Seventeen PD patients of mild to moderate severity were studied off of their dopaminergic medications and compared to 15 age-matched controls. Compared to controls, PD patients had distinct pre- and post-reversal deficiencies depending upon the difficulty of the choices they had to learn. The patients also exhibited compromised adaptability to the reversal. A computational model of the subjects' trial-by-trial choices demonstrated that the adaptability was sensitive to the gain with which patients weighted pre-reversal feedback. Collectively, the results implicate the nigral dopaminergic system in learning to make choices in environments with probabilistic and dynamic reward contingencies.

Key words: dopamine, basal ganglia, computational modeling, reinforcement learning.

It has become clear in recent years that Parkinson's disease (PD) affects not only the initiation and control of movements, but also motivational drive and reward-seeking behavior (Borek et al., 2006), which themselves are fundamental to the learning of new responses. A classic neuropathology of PD is the degeneration of dopamine cells in the substantia nigra pars compacta (Dauer and Przedborski, 2003). This not only produces a substantially reduced tonic level of dopaminergic activity in efferent targets, but also likely impairs phasic dopaminergic activity (Grace, 1991; Frank et al., 2004; Schultz, 2007). A broad body of theoretical and empirical evidence has accumulated suggesting that phasic activity of the midbrain dopamine system is critical to trial by trial feedback-based learning (Abler et al., 2006). The predominant concept is that the phasic dopamine activity signals actual versus expected reward values, or a reward "prediction error" (Montague et al., 1996; Schultz et al., 1997; Fiorillo et al., 2003). This prediction error, in turn, is thought to play a key role in rewarded learning and has gained widespread use in temporal difference models of learning that are driven by reinforcing rewards (Sutton and Barto, 1998).

Two key aspects of rewarded learning can make it particularly challenging. First, the relationship between choices and rewards can change over time. A common paradigm for investigating dynamic reward contingencies is reversal learning tasks. In such tasks, after learning associations between stimuli, choice, and reward, subjects have to adapt their internal representations to reflect a reversal in some aspect of the associations. Another source of challenge in rewarded learning is that the relationship between choices and rewards can be probabilistic. The relative merit of various options has to be inferred indirectly through protracted trial-and-error learning. If one option rarely rewards and an alternative frequently rewards, the choice is relatively easy. However, if the probabilities with which two alternatives reward are relatively similar, learning to make the favorable choice becomes more difficult. In an important extension of previous investigations of reversal learning in PD patients off medications, Robbins et al. (Swainson et al., 2000; Cools et al., 2007) have incorporated probabilistic reward contingencies. In these studies, however, subjects have been told ahead of time that the better of two choices would change and that they should modify their choice accordingly. Yet learning how or even whether the choice–reward contingencies will change is particularly challenging when one is not aware of these possibilities in advance. Thus how PD patients off dopaminergic medications respond to unexpected reversals in probabilistic reward structure remains unclear. In light of dopamine's role in effortful learning and decision making (Assadi et al., 2009), one would expect that choice difficulty may differentially affect probabilistic

*Corresponding author. Tel: +1-858-822-6765; fax: +1-858-534-2014.
E-mail address: hpoizner@ucsd.edu (H. Poizner).
Abbreviations: BDI, Beck Depression Inventory; MMSE, Mini-Mental State Exam; PD, Parkinson's disease; UPDRS, United Parkinson's Disease Rating Scale.

0306-4522/09 $ - see front matter
doi:10.1016/j.neuroscience.2009.07.033

reversal learning in PD patients compared with healthy controls.

The present study seeks to determine whether and how rewarded learning in the face of changing and variably difficult reward contingencies is impaired in PD. To investigate this issue, we combined a temporal difference reinforcement learning model and a rewarded learning task originally developed for use in midbrain single-unit recording in primates (Morris et al., 2006). As in the original experiment by Morris et al., we varied difficulty by having subjects choose between two visual stimuli the reward probabilities of which differed by either a small (e.g. 25%) or large (e.g. ≥50%) amount. However we added a test of reversal learning: midway through the task session and without forewarning the subjects, we reversed the reward probabilities of the visual stimuli. We hypothesized that, relative to healthy age-matched counterparts, PD subjects off dopaminergic therapy would show the greatest deficiency in learning to make favorable choices in the difficult case when stimuli differed by small reward probabilities. We further hypothesized that PD patients would be deficient in optimizing strategy and would show specific impairment in learning when reward probabilities are reversed. Because, to our knowledge, there are no prior reports on human behavior in this rewarded learning task, we analyzed each subject group's learning and reversal adaptation separately prior to directly comparing the two groups. We also applied the temporal differences reinforcement learning model to their trial-by-trial choices to identify mechanistic distinctions between how PD patients and controls adapt to the reward contingency reversal. Our results indicate that PD patients off dopaminergic medications exhibit learning and reversal adaptation deficiencies that are particularly sensitive to choice difficulty. Examination of differences in model parameters between normals and PD patients pointed to specific means through which dopamine deficiency may alter probabilistic reversal learning.

EXPERIMENTAL PROCEDURES

Subjects

Seventeen patients with mild to moderate idiopathic PD at Hoehn and Yahr Stages II and III of the disease (Hoehn and Yahr, 1967) participated. Patients were referred (D.D.S.) from the UCSD Movement Disorders Clinics, and from local PD support groups. We excluded any patients exhibiting additional deficits in other neural systems ("Parkinson plus" patients), dementia, major depression, psychosis or any neurological or psychiatric disease in addition to PD. After detailed explanation of the procedures, all subjects signed a consent form approved by the institutional review board of the University of California San Diego. Disease duration was calculated on the basis of patients' subjective estimate of the onset of first motor symptoms. Patients were evaluated OFF-medications in the morning at least 12 h (Defer et al., 1999) after discontinuing all anti-Parkinsonian medications. They were administered (D.A.P.) the Mini-Mental State Exam (MMSE (Folstein et al., 1975)) and Beck Depression Inventory (BDI, Psychological Corporation, Boston, MA, USA) to exclude subjects with dementia or major depression. In order to get a uniform measure of the clinical state of PD patients at the time of the experimental session, all PD patients were also rated (H.P.) on the motor scale of the United Parkinson's Disease Rating Scale (UPDRS (Goetz et al., 1995)) and staged on the Hoehn and Yahr scale (Hoehn and Yahr, 1967). Fifteen healthy controls were recruited through patient caregivers and the local community. All subjects had vision correctable to 20/40 with corrective lenses. All subjects were tested for hand dominance based on the Edinburgh Handedness Inventory (Oldfield, 1971). Eleven of the controls and 13 of the patients were right handed. Nine of the controls and five of the patients were female, reflecting the typical gender distribution of idiopathic PD. The groups were well matched by age, with similar ranges and mean ages differing by less than 1 year (patients: mean 66.1 (8.2), range 50−81; controls: mean 65.2 (7.2), range 52–77). Subject information and, for the PD patients, a basic clinical profile are given in Table 1.

Table 1. Subjects

	Controls	Patients
N	15	17
Age	65.2 (±7.2)	66.1 (±8.2)
Age range	[52–77]	[50–81]
Gender (M/F)	6/9	12/5
Handedness (L/R)	4/11	4/13
Disease duration (y)	n.a.	10.4 (±4.4)
UPDRS III	n.a.	37.6 (±9.2)
H & Y stage	n.a.	2.4 (±0.5)
BDI	n.a.	9.0 (±5.0)
MMSE	n.a.	28.4 (±1.8)

Experimental task

We adapted a task originally used to study firing rates of dopamine cells in primate substantia nigra pars compacta (Morris et al., 2006) for use as an instrumental reward-based learning task with humans. The task is a variant of the classic two-armed bandit (Robbins, 1952). Briefly, subjects were presented with a series of trials on which they chose abstract visual images with a possibility of accruing a small reward on each trial. Given the evidence that rewarded striatal-based learning is particularly sensitive to the use of real versus symbolic monetary rewards (Kunig et al., 2000; Martin-Soelch et al., 2001), we gave subjects actual cash for rewards. The images were chosen from among four possible images, each with a fixed probability of producing an identical reward value. In order to maximize their earnings, subjects had to learn through trial-and-error which images were more likely to pay off. Halfway through the experiment and without any cues from the experimenter, the reward probabilities of the four images were reversed, thereby testing subjects' ability to adapt to the reward contingency reversal.

Throughout the task, subjects were seated in front of a 19″ touch monitor (Elo Touchsystems, Menlo Park, CA, USA, model number et1925L-7uwa-1) in sufficiently close proximity to allow comfortable reaches to both upper corners. The touch monitor was placed on a table with the top approximately 45° back from vertical. As depicted in Fig. 1A, subjects initiated each trial by pressing the green "go button" square in the lower middle of the touch monitor. After 800–900 ms, a square visual image appeared in each of the two upper corners of the touch monitor. Subjects chose an image by pressing it. A short 50–100 ms after selecting an image, subjects were given simultaneous visual and auditory feedback signals. If they won money on that trial, their cumulative winnings were displayed above the chosen image for the remainder of the time that the images are displayed and they were presented with a 200 ms "high" tone (600 Hz). If they did not win money on that trial, "$0.00" was displayed and they were presented with a "low" tone (200 Hz). The two tones were provided

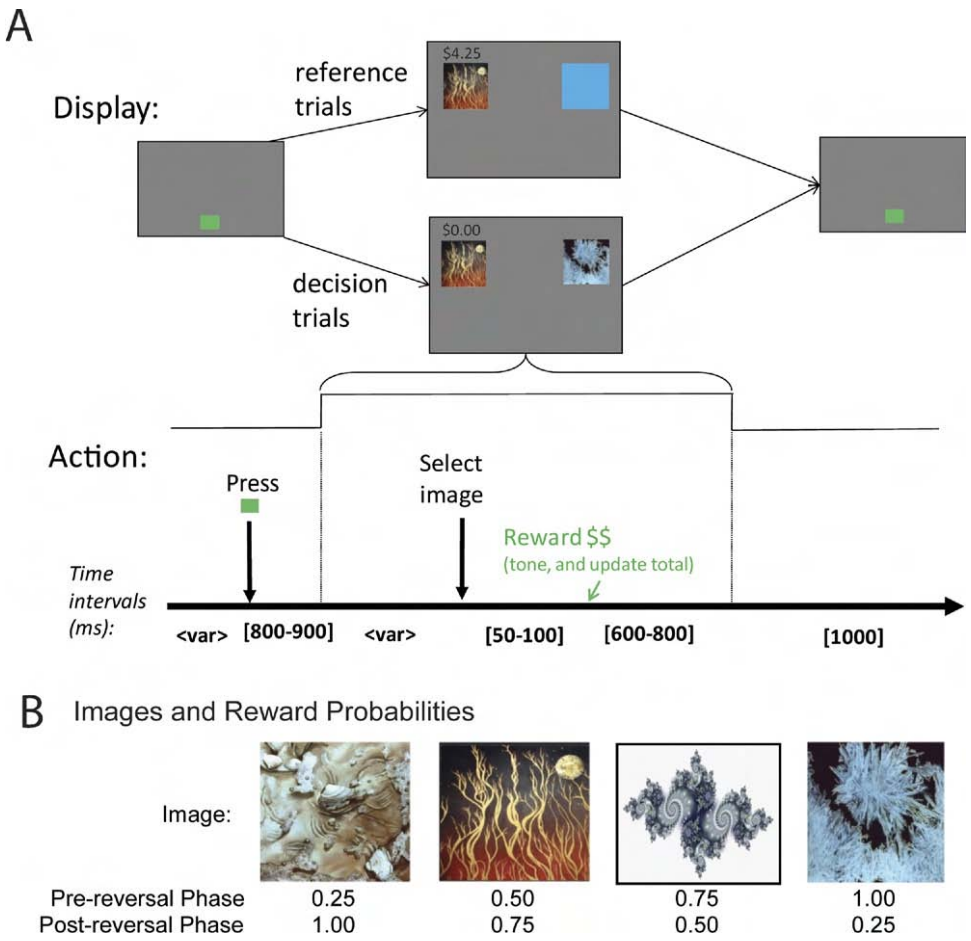

Fig. 1. Task design. (A) Per-trial timeline. Time intervals in square brackets represent durations drawn randomly from a uniform distribution over the specified range. The "<var>" denotes a variable length, subject-driven interval. (B) Visual images and their phase-contingent payoff probabilities.

free field by standard personal computer (Dell Corporation, Austin, TX, USA) speakers. The tones were identical in amplitude and linear ramp up/down (40 ms each). Prior to starting the experiment, subjects confirmed by verbal report that they could hear and distinguish the two tones. Approximately 600–800 ms after the feedback signal, the images disappeared and the go button reappeared in the lower center of the monitor, prompting the subject to begin the next trial. Subjects were required to wait until the two images appeared before releasing the go button. There were no other temporal constraints on their choice or the return to the go button. They were simply instructed to "move to touch the image as soon as you have decided which one to choose." Actual durations of each time interval specified above were chosen randomly from a uniform distribution on each trial. Total trial duration averaged about 4 s.

The task consisted of two phases of 256 trials each. Interleaved throughout the task were two trial types: reference and decision trials comprising 62.5% and 37.5% of the trials, respectively. On the reference trials, subjects were given an "instructed" choice. They were presented with a solid blue square and one of four abstract images. They were instructed to always choose the abstract image. On the decision trials, subjects faced a two-alternative forced choice. They were presented with two of the abstract images and were told to "choose the image that is more likely to pay off." If rewarded, they received $0.02 on reference trials and $0.15 on decision trials. The abstract images and the probability with which choosing them produced a reward [0.25, 0.50, 0.75 and 1.00] are shown in Fig. 1B. These reward contingencies were flipped in the otherwise identical post-reversal phase of the experiment. There were no decision trials on which the two images were identical. We fully counterbalanced the number of presentations of each image, the side on which they were presented, and the side on which rewards were available. Maximum run lengths were three decision trials, five reference trials, five trials with reward on the same side, three reference trials with the image on the same side, and five trials containing the same image on either side. Both 256-trial phases were divided into eight blocks of 32 trials each. At the end of each block, subjects were shown their cumulative winnings and the actual monetary amount placed on the table beside them was updated accordingly, rounded up to the nearest $.25.

Subjects were first given a brief practice session, with eight reference and four decision trials. The practice stimuli were four simple geometric shapes that were different from any of the stimuli used in the actual experiment. There were no feedback signals or rewards in this practice session in order to avoid teaching any associations prior to the actual experiment. Subjects were simply familiarized with the mechanics of the trials, and particularly the explicit instruction to not choose the solid blue square on reference trials. Prior to starting the primary experiment, subjects were given an explanation of the feedback signals and rewards. They were told that some images were more likely to pay off than others, and it did not matter which side they appeared on. They were also instructed that, on trials with two images, they should try to choose the image that is more likely to pay off. Finally, they were told that to maximize their winnings, they should try to figure out which images are more likely to pay off than others. During the

post-experiment debriefing, subjects were asked to provide one to two word descriptions of the four images, which ones they thought were most likely to pay off, and whether they noticed a change in the relative payoffs of the images. Subjects were paid their winnings from the game plus $20 per hour for the non-experimental portion of the session, including intake, BDI and MMSE, and UPDRS testing. The average duration of the overall session was approximately 2.0 h.

Reinforcement learning model

We implemented a computational reinforcement learning model to fit subjects' trial-by-trial behavior. Images $j \in \{1,2,3,4\}$ were assigned values $Q_t(j)$ at each trial t of the experiment. When image k is chosen, its value was incremented as a function of the reward $r_t \in \{0,1\}$ received upon choosing it:

$$Q_{t-1}(j) \leftarrow \begin{cases} Q_t(j) + \alpha\left[r_t - Q_t(j)\right] & \begin{cases} j = k \\ o.\ w. \end{cases} \\ Q_t(j) \end{cases}$$

The term $[r_t - Q_t(j)]$ was referred to as the prediction error. The amount by which the prediction error was used to increment the image's value was weighted by the learning rate, or "gain," α. On decision trials where subjects had to choose between two images m and n, we modeled their choice probabilistically with the softmax function:

$$p_t(m) = \frac{e^{\beta Q_t(m)}}{e^{\beta Q_t(m)} + e^{\beta Q_t(n)}}$$

where the parameter β quantified the bias between exploration (low β) and exploitation (high β). We investigated the role of gain and exploration/exploitation bias in the two phases separately, giving four parameters: $\alpha_{initial}$, $\alpha_{reversal}$, $\beta_{initial}$, $\beta_{reversal}$ evaluated over the ranges [0 0.70], [0 0.72], [0 10], and [0 11], at uniform intervals of 0.07, 0.08, 1, and 1, respectively. We used a simple grid search of the parameter space to evaluate the model's fit with each subject's actual behavior. The same grid of values for alpha and beta was explored for all subjects in each phase separately in order to determine which parameter value combination best fit each subject's decisions. The fit at each point in the parameter space was computed as the log likelihood that the model makes the same choices α_t that the subject makes on the decision trials:

$$LLE = \log \prod_{t \in 2AFC} p_t(\alpha_t)$$

Subjects for whom the "best" model did not fit better than chance were discarded from subsequent analyses. For all other subjects, the four parameter values that optimized their model fit defined their learning "profile."

Data analysis and statistics

Performance was measured as the percentage of decision trials on which subjects chose the favorable image (i.e. more likely to pay off) in each block of 12 decision trials. Learning magnitude in each phase was defined by the mean performance in the "late" (last two) blocks minus the mean performance in the "early" (first two) blocks. The decision trials were divided into two equal-sized mutually exclusive classes: the "large difference" trials and the "small difference" trials. On the relatively easy "large difference" trials, the payoff probabilities of the two presented images differed by 50% or more. Conversely, on the relatively difficult "small difference" trials, the payoff probabilities differed by only 25%. The mean reward probability of the two images presented, 62.5%, was identical for the easy and difficult choices. Performance was evaluated using a mixed-design four-factor ANOVA, with Group (control, PD) as a between subjects factor and Difficulty (easy, difficult), Phase (pre-reversal, post-reversal), and Block (early, late) as within subjects factors, and Geisser–Greenhouse corrections for non-spherical covariances.

Subjects' ability to adapt to the reversal in reward contingencies, which we refer to as their "adaptability," was measured as the increase in their learning from the pre-reversal phase to the post-reversal phase. These adaptability metrics were evaluated with two-tailed t-tests or non-parametric counterparts where the distributions were found to differ from normality based on Lilliefors's composite goodness-of-fit test (Lilliefors, 1967). We also investigated the extent to which learning in the pre-reversal phase predicted learning in the post-reversal phase by evaluating the correlation between performance in the two phases on a block-by-block basis. The proportion of subjects in each group reporting a change in image reward probability contingencies was compared using a chi-square test. For those subjects whose data could be fit by the model better than chance, their "best fit" model parameters were used to investigate the correlations, if any, between learning profiles and adaptability. Throughout the analysis, P values less than 0.05 were considered significant.

Table 2. 4-Way ANOVA summary

Factor(s)	$F(1,30)$	P
Block	38.27	<0.0001
Phase	15.84	<0.001
Phase×Difficulty	15.73	<0.001
Block×Difficulty	12.40	0.001
Phase×Block	9.96	0.004
Group×Phase×Block	5.90	0.021
Group×Block×Difficulty	4.33	0.046

RESULTS

Learning

Table 2 summarizes the results of the mixed-design four-factor ANOVA, with Group (control, PD) as a between subjects factor and Difficulty (easy, difficult), Phase (pre-reversal, post-reversal), and Block (early, late) as within subjects factors. For the purpose of brevity, only those main and interaction effects that were statistically significant are reported. As shown with the main effect of Block and depicted in Fig. 2, subjects demonstrated a learning effect, correctly choosing the more favorable image on average 67% of the time late in each learning phase, compared with 50% (chance level) early in each phase.

As expected, more difficult decisions, on which two images differed in payoff probability by only 25%, were harder to learn than the relatively easier decisions, as seen in the significant Block×Difficulty interaction and when comparing Fig. 2A and 2B. Disregarding the factor of Phase, there was a mean 22% improvement on easy decisions over trials compared with a 9% improvement on the difficult decisions. There was a significant Group×Block×Difficulty interaction indicating that controls and PD patients differed in how they learned to make the relatively easy versus more difficult decisions. On easy decisions controls chose the favorable image 55% of the time early in learning compared with 45% for patients, yet both groups performed almost equivalently by late in learning (73% and 71% favorable choices, respectively). The stronger performance early in learning is most evident in the pre-

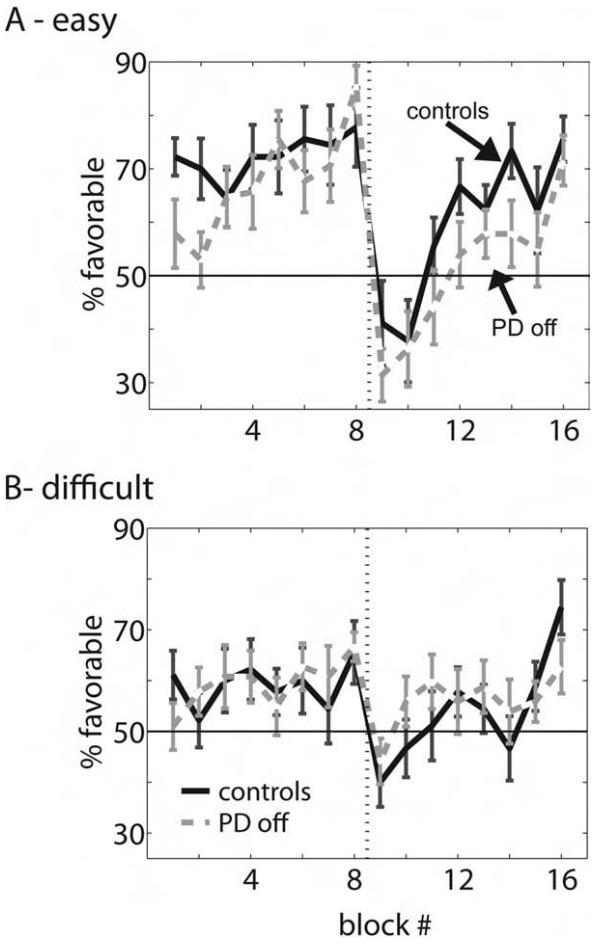

Fig. 2. Learning curves, group averages. Mean and ±standard error at each block of 12 decision trials. Chance performance is 50%. Dotted vertical line after block 8 denotes reward contingency reversal. (A) Easy ("large-difference") trial pairs only, on which the two images differed in their probability of payoff by 50% or greater and (B) difficult ("small-difference") trial pairs only, on which the images' payoff probabilities differed by 25%.

reversal phase (e.g. Fig. 2A, early blocks). In a post hoc analysis of only the pre-reversal phase, we found that controls chose the favorable image on easy trials an average of 71% of the time in the early blocks, whereas patients chose the favorable image only 55%, a statistically significant difference in a two-tailed *t*-test ($t(30)=2.434$, $P=0.02$). The majority of control subjects chose the more favorable image on all of the first block's easy decision trials, whereas the majority of the patients chose the more favorable image on only two of the first block's easy decision trials. The discrepancy was present even for the first easy decision trial, which came after four reference trials and on which 73% of the controls chose favorably, but only 35% of the patients did (chi-square=4.63, $P<0.05$).

In post hoc analyses of the post-reversal phase, although the two groups were statistically indistinguishable on the easy trials in terms of learning magnitudes, they had distinctly different learning magnitudes on the difficult trials. Specifically, the post-reversal phase "difficult choice" learning magnitude was 23% for controls but only 9% for patients, a statistically significant difference ($P=0.018$ in a rank-sum test). Although patients showed higher early post-reversal phase performance and lower late post-reversal phase performance than the controls, neither of these effects alone was significant ($t(30)=1.35$ and 1.55, respectively, n.s.) and neither could account for the between-group difference in post-reversal phase learning magnitudes. Despite this result, it should be noted that both groups had a hard time learning the "difficult" distinctions, on which two images differed in payoff probability by only 25% (see Fig. 2C). Over both phases of the experiment, there were only six out of the 16 blocks in which both groups performed at more than one standard error above chance on the difficult trials. In summary, because there was no statistically significant Group×Block interaction, the significant Group×Block×Difficulty interaction suggests that the patients exhibited a difficulty-dependent learning deficit in which they had a compromised ability to learn which of the more ambiguous "small difference" stimuli were more likely to pay off.

Adaptability

We analyzed the effect of the reward contingency reversal, which occurred after the completion of block 8, in several ways. As shown with the main effect of Phase, subjects demonstrated a reversal effect (see Fig. 2) where the reversal resulted in the immediately subsequent drop to below-chance performance in block 9. Within-phase learning depended on the phase, as evidenced by the significant Phase×Block interaction. Specifically, average performance during the pre-reversal phase increased from 59% favorable choices in early blocks to 70% in later blocks, compared to 42% and 64%, respectively, in the post-reversal phase. There was also a Group×Phase×Block interaction, in which patients exhibited weaker learning in the post-reversal phase (only increasing from 42% to 61% favorable choices) than controls (41% to 68%) despite stronger learning in the pre-reversal phase (54% to 71%) than controls (64% to 68%).

We also evaluated subjects' adaptations to the reversal in terms of the inter-phase dynamics of their learning. Fig. 3A and 3B depicts the relationship between pre- and post-reversal phase learning on a block-by-block basis for the easy and difficult cases, respectively. Group-average performance on each block is shown by the block numbers, the centers of which have x- and y-coordinates corresponding to post-reversal and pre-reversal performance, respectively. The dashed line on the diagonal divides each plot into halves: points in the left half are associated with lower performance on post-reversal relative to a corresponding block in the pre-reversal phase. Conversely, points in the right half are associated with higher performance on reversal relative to a corresponding block in the pre-reversal phase. Note that, for both groups, most of the data lie to the left of the diagonal. Thus, in most cases, post-reversal performance was lower than performance in the corresponding pre-reversal block.

In the case of the "easy" trials (Fig. 3A), the controls' adaptability could not be accounted for in a linear fashion ($R=0.57$, n.s.), but the patients' adaptability could

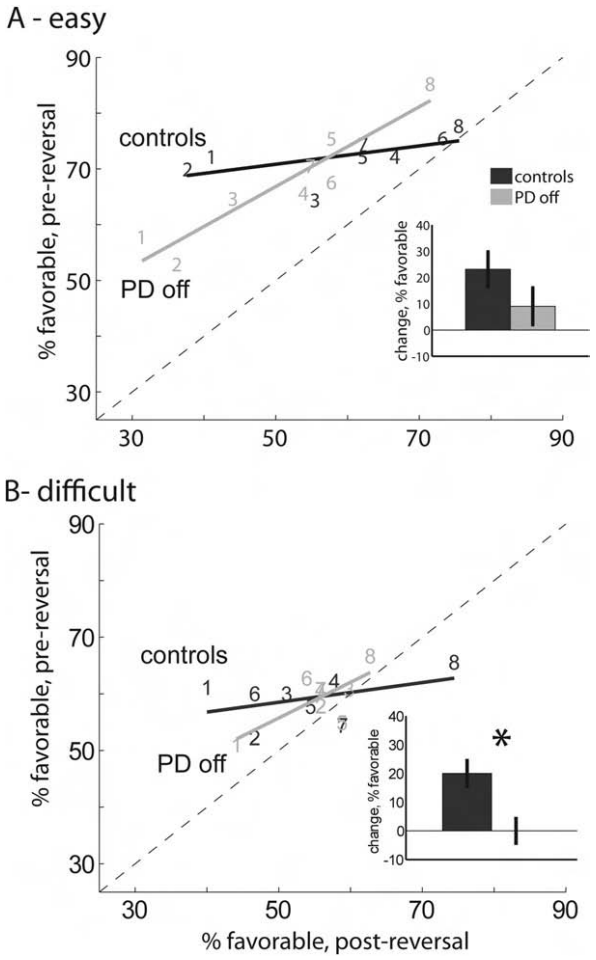

Fig. 3. Adaptability. Relationship between pre- and post-reversal phase block-by-block learning. Numbers denote block within each eight-block phase. Lines are best linear fits. Dashed line on the diagonal indicates where pre- and post-reversal phase learning are equivalent. (A) Easy ("large-difference") trial pairs only and (B) difficult ("small-difference") trials pairs only. Insets: mean change in learning from pre-to post-reversal phase. Bars denote standard error, * $P<0.05$.

($R=0.93$, $P<0.001$), with a slope of 0.72 (95% CI 0.43–1.00). In other words, controls parlay a given amount of pre-reversal learning into more post-reversal learning than do patients. On the "difficult" trials (Fig. 3B), again the controls did not show a statistically significant linear relationship ($R=0.43$, n.s.) but the patients did ($R=0.72$, $P<0.05$).

Insets for Fig. 3 depict a related but alternative metric of adaptability: the increase in learning from pre- to post-reversal phase. Lilliefors's test showed that the distribution of this metric was normal in both groups for each class of trial. In the easy case, patients had lower adaptability than controls, exhibiting a mean 9.0% (SE 28.9) improvement in learning compared to 23.1% (SE 25.6) for controls, a non-significant difference ($t(30)=1.45$, $P=0.157$). In the difficult case, the difference is more marked, with controls exhibiting a 20% (SE 17.2) improvement in learning, compared to zero improvement for patients (SE 17.7), a statistically significant difference ($t(30)=3.23$, $P<0.005$).

We also analyzed subjects' responses to the reward contingency reversal through debriefing. When asked if they noticed a change in the images' relative payoffs, 11 out of 14 control subjects and seven out of 17 patients said they did (the response data were missing for one of the controls). The proportion was significantly higher for controls than patients (chi-square(1)=4.41, $P<0.05$).

Learning profiles and adaptability

For the purposes of evaluating adaptation in the present study, the difficult trials were removed from analysis with the reinforcement learning models for two reasons. First, they exhibited a weaker effect at the reversal, with both groups showing weak and non-significant effects of phase in the difficult trials. Second, the behavioral choices on the difficult trials were deemed too noisy for fitting with the computational model. We expected that modeling such behavior would be more susceptible to overfitting even by models with very few free parameters. When only decisions on the easy trials were modeled, the model was able to fit 11 of the 15 (73%) of controls and eight of the 17 (47%) of patients' behavior better than chance. For both groups, the model fit was positively correlated with the endpoint performance in the pre-reversal learning phase ($R=0.80$, $P<0.001$). Thus, as a general rule, the subjects not fit by the model exhibited weak or nonexistent learning. As a result, all subsequent analyses investigating the relationship between learning profiles, as quantified by best-fit model parameters, are based solely on these 11 controls and eight patients.

Fig. 4A and 4C shows the learning profiles for each subject in each group, characterized by best-fit model parameters for the pre- and post-reversal phases of the experiment, respectively. The majority of subjects in both groups exhibited gain factors under 0.3. Both groups also exhibited a trend toward more exploration (lower beta) in the post-compared to the pre-reversal phase. Subjects' exploration/exploitation bias, as quantified for each phase by $\text{beta}_{\text{initial}}$ and $\text{beta}_{\text{reversal}}$, did not have a systematic influence on their inter-phase adaptability. However, in the case of the Parkinson's patients, their gain factor during the pre-reversal phase, quantified by $\text{alpha}_{\text{initial}}$, did have a systematic influence on their adaptability. Specifically, as shown in Fig. 4B, the patients with higher gain during the pre-reversal phase exhibited better adaptation than did patients with lower gain ($R=0.81$, $P<0.05$). In contrast, the pre-reversal phase gain did not seem to influence the control group's adaptation ($R=0.10$, n.s.). There was a trend for the opposite effect in the case of the post-reversal phase gain (Fig. 4D), although the correlation was not significant ($R=0.58$, $P=0.13$).

DISCUSSION

Basic findings

Parkinson's patients off medications initially exhibited weaker learning than their age-matched control subjects when facing relatively easy choices involving large differ-

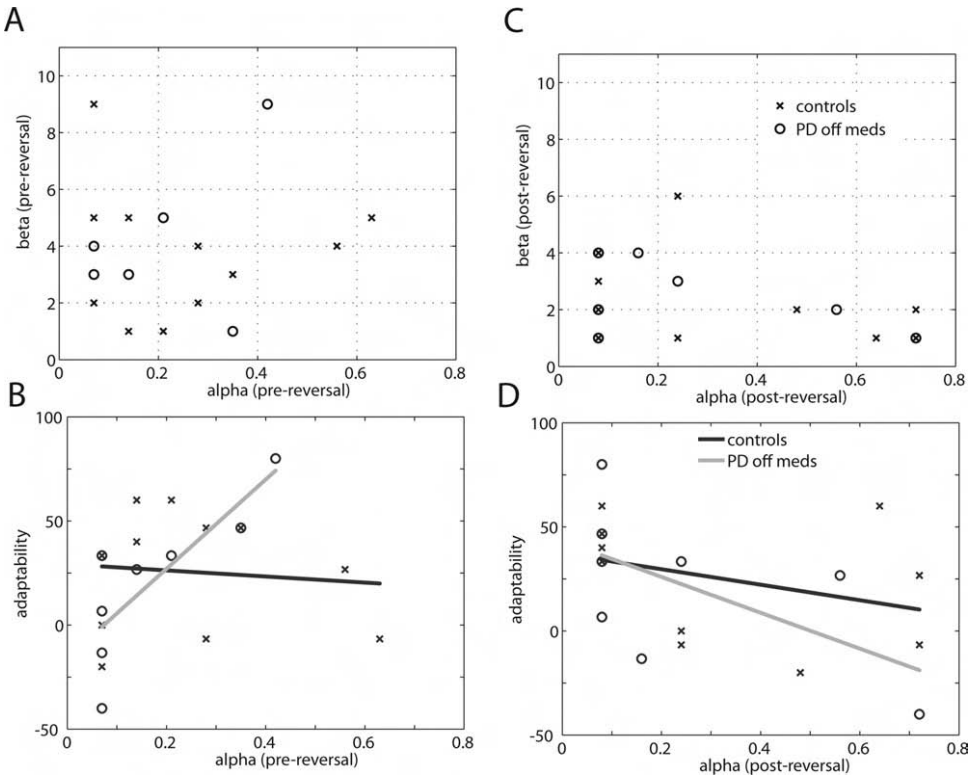

Fig. 4. Learning profile and adaptability. (A, C) Scatter plot of individual subjects' learning profiles on easy ("large-difference") decision trials (A, pre-reversal phase, C, post-reversal phase). Alpha is the learning rate, beta is the exploration/exploitation bias. (Although there are eight patients used in the analysis, only six unique points appear in (A) because the alpha=0.07, beta=3 point is a triplicate. Likewise in (C), the alpha=0.08, beta=1 point is a duplicate for both groups.) (B) Adaptability as a function of pre-reversal phase learning rate. Lines are best linear fits for the two groups. (D) As in (B), but for post-reversal phase learning rate. (The alpha=0.08 and adaptability=46.7 point is a duplicate.)

ences in payoff probabilities. However, by the end of the pre-reversal learning phase, patients caught up and their performance matched that of controls. In contrast, when faced with more difficult choices involving small differences in payoff probabilities, patients performed as well as controls initially, but faltered after the reward contingencies were reversed. The net result of these effects produced a compromised ability of patients to adapt to the reward contingency reversal, and this deficiency was associated with lower pre-reversal phase prediction error gains in a computational model of their behavior.

Learning

At a gross level when decision difficulty is disregarded, patients exhibited a learning profile similar to age-matched controls, indicating that mild to moderate PD patients off medications can still learn this type of task. However interesting differences emerge when decision difficulty is taken into account. First, patients were slower at learning the relative reward contingencies for image pairs that had a large difference in reward probabilities. This cannot be accounted for by overall learning ability on these trial types, because the patients' performance caught up to that of the controls by the end of the pre-reversal learning phase. There are at least three possible not mutually exclusive explanations for this. First, controls may give more weight to the reference trials, better leveraging that information during decision trials. This explanation is supported by the controls' much better performance than patients on even the first easy decision trial (which had been preceded by reference trials). Second, and relatedly, it may be that controls are better able to employ a declarative strategy that gives them an advantage early in learning. Third, it may be that early learning is particularly dependent upon the dopaminergic system (Horvitz et al., 2007). In contrast to the easy "large difference" trials, patients exhibit deficient post-reversal learning relative to controls on more difficult "small difference" trials. This result suggests that patients may be specifically impaired on more difficult reversal learning. Thus, reversal learning may be sensitive to a compromised dopamine system in a difficulty-dependent fashion, whereby more difficult dissociations are harder to re-learn than their otherwise equivalent easier counterparts.

Adaptability

We sought to examine how subjects would translate pre-reversal learning to learning capability in the post-reversal regimen. For both groups, block-wise performance in the pre-reversal phase predicted block-wise performance in the post-reversal phase. However, the groups differed in terms of how they parlayed pre-reversal phase learning

into post-reversal phase learning. Patients' post-reversal learning was associated with substantial pre-reversal learning. Controls' post-reversal learning, however, was associated with minimal pre-reversal learning. As a result, controls exhibited stronger adaptability, with a net increase in learning magnitude of 20% in response to the reward contingency reversal. In contrast, the patients showed no significant improvement in learning. Thus, relative to controls, PD patients were markedly deficient in their ability to adapt to the reward contingency reversal. This difference was not solely explained by the patients' deficient early learning in the pre-reversal phase for easy trials, because the effect was strongest when the difficult trials involving decisions between stimuli with small differences in reward probabilities were included. Thus patients exhibited reduced adaptability in the face of subtle changes in reward contingencies. Post-experiment debriefing corroborated this interpretation because a significantly lower percentage of the patients reported noticing the change in reward contingencies than did controls, consistent with earlier reports of reduced explicit knowledge of implicit learning in PD (Wilkinson and Jahanshahi, 2007; Wilkinson et al., 2008). A host of general factors that can influence performance in learning tasks is unlikely to account for the compromised learning and adaptation exhibited by the Parkinson's patients in this study. For example, it is unlikely that any group differences in understanding instructions, motivation, speed of choice execution, or fatigue played a differential role in reversal learning in this task, because both groups were able to learn the overall task, even after the reward contingency reversal.

Although PD patients off medications tend to exhibit reversal learning deficits in sensorimotor tasks (Krebs et al., 2001; Messier et al., 2007) they generally do not in cognitive reversal learning tasks (Swainson et al., 2000; Cools et al., 2006). Since cognitive forms of reversal learning have been linked to ventral striatum (Cools et al., 2002), and the dopamine depletion in mild PD is greater in dorsal than in ventral striatum (Kish et al., 1988), ventral striatal–mediated reversal learning would be relatively spared in mild PD off medications. Another not mutually exclusive possibility is that reversal learning's dependence upon tonic function of midbrain dopamine systems is sensitive to difficulty of the specific task (Shohamy et al., 2008). In the present study, relative pre- and post-reversal learning was sensitive to the difficulty of the probabilistically-rewarded choices.

Mechanisms for the compromised adaptability

We sought to determine whether aspects of the subjects' learning styles could account for their adaptability in this task. Individual subjects' learning styles were quantitatively characterized with learning "profiles," consisting of two model parameters inferred from their decisions on choice trials during each of the task's two phases. Reinforcement learning algorithms provide a powerful and increasingly prevalent means by which to estimate these internal variables that are otherwise not directly available from measurements of stimuli, rewards, and choices (Daw and Doya, 2006). One parameter, alpha, specified the gain that linearly weighted the prediction error on each trial in order to modify relative value for the chosen image. The other parameter, beta, was used to bias the tradeoff between exploration and exploitation in the course of translating image values into image choice probabilities. Thus, in the context of the framework recently put forth by Montague and colleagues (Rangel et al., 2008), alpha influences the learning, and beta influences the action selection.

We evaluated whether these characteristics of subjects' learning profiles could account for their relative adaptability. There was no clear pattern of association between adaptability and either the pre- or post-reversal beta. However, alpha during the pre-reversal phase was positively correlated with adaptability among PD patients. Higher gain factors during their pre-reversal phase led to positive adaptability, whereas lower gains led to lower (or in some cases negative) adaptability. This result is consistent with Berns and Sejnowski's (1998) proposition that set shifting deficits may be a natural consequence of slow learning. In the case of controls, however, the gain had no discernible effect on adaptability. This result raises two interesting points regarding adaptability in PD patients off medications. First, the learning profile prior to the reward contingency reversal can predict how the subject will subsequently adapt to it. Second, to the extent that a learning profile can be associated with a learning strategy, the result suggests that patients can compensate for deficient adaptability by modulating their learning strategy to use higher gains in early learning.

Whether an individual subject's behavior could be fit with the model corresponded to whether or not the subject successfully learned relative reward contingencies. That a much higher percentage of the patients had behavior that the model could not fit suggests a more general deficiency in learning in the patients. Although consistent with the patients' deficient learning from behavioral measures, making the same inference based on model fits needs to be treated with caution, because it is inherently reliant upon the specific computational model we chose. The model is relatively simple, with only two parameters for each of the two phases in the experiment. This should minimize the risk of overfitting one group more than the other. Nevertheless, the possibility remains that the patients' trial-by-trial choices reflect learning strategies unique to their group and that are less veridically captured by the specific model we employed. This raises the possibility that a further exploration of the space of potential models and associated parameters may help generate novel hypotheses about subjects' learning strategies.

Our results highlight the importance of considering individual differences in evaluating computational models of subjects' behavior in implicit learning tasks. This has also been demonstrated in relating information from models to activity levels in striatal and frontal cortical areas (Cohen, 2007; Brown and Braver, 2008). Learning rates in particular may be one of the key subject-specific model parameters, as they were a key predictor of adaptability in the present study and also predicted activity levels in an-

terior cingulate (Behrens et al., 2007) and broader frontal–striatal circuits (Cohen and Ranganath, 2005) in neuroimaging studies. Thus we expect that the growing use of computational models in conjunction with behavioral, neuroimaging, and electrophysiological approaches will lead to new insights and new hypothesis generation regarding the neural mechanisms supporting probabilistic reversal learning in humans.

CONCLUSION

In a reversal learning task not previously evaluated with humans, PD patients off medications achieved the same level of overall learning as their age-matched counterparts, but had distinct pre- and post-reversal deficiencies depending upon the difficulty of the choices they had to learn. The patients also exhibited compromised adaptability to the reversal. A computational model of the subjects' trial-by-trial choices demonstrated that the adaptability is sensitive to the gain with which patients weighted pre-reversal feedback. Collectively, the results suggest that the nigral dopaminergic system is involved in a difficulty-dependent fashion with multiple aspects of probabilistic reversal learning.

Acknowledgments—We are grateful to all our participants, especially the patients who voluntarily withdrew from their medication in order to participate. We thank Genela Morris, Hagai Bergman, Jonathan Nelson, Peter Dayan, Michael Mozer, and Pietro Mazzoni for their helpful discussions, Andrey Vankov and Slavik Bryksin for their assistance with experimental software, Jason Mclnerny for helpful discussions about our approach to the modeling, and Thomas Urbach for assistance with statistical software. We are also grateful to anonymous reviewers for helpful suggestions for improving the manuscript. This work was supported in part by the National Science Foundation grant SBE-0542013 to the Temporal Dynamics of Learning Center, an NSF Science of Learning Center and the National Institutes of Health grant 2 R01 NS036449.

REFERENCES

Abler B, Walter H, Erk S, Kammerer H, Spitzer M (2006) Prediction error as a linear function of reward probability is coded in human nucleus accumbens. Neuroimage 31:790–795.

Assadi SM, Yucel M, Pantelis C (2009) Dopamine modulates neural networks involved in effort-based decision-making. Neurosci Biobehav Rev 33:383–393.

Behrens TE, Woolrich MW, Walton ME, Rushworth MF (2007) Learning the value of information in an uncertain world. Nat Neurosci 10:1214–1221.

Berns GS, Sejnowski TJ (1998) A computational model of how the basal ganglia produce sequences. J Cogn Neurosci 10:108–121.

Borek LL, Amick MM, Friedman JH (2006) Non-motor aspects of Parkinson's disease. CNS Spectr 11:541–554.

Brown JW, Braver TS (2008) A computational model of risk, conflict, and individual difference effects in the anterior cingulate cortex. Brain Res 1202:99–108.

Cohen MX (2007) Individual differences and the neural representations of reward expectation and reward prediction error. Soc Cogn Affect Neurosci 2:20–30.

Cohen MX, Ranganath C (2005) Behavioral and neural predictors of upcoming decisions. Cogn Affect Behav Neurosci 5:117–126.

Cools R, Altamirano L, D'Esposito M (2006) Reversal learning in Parkinson's disease depends on medication status and outcome valence. Neuropsychologia 44:1663–1673.

Cools R, Clark L, Owen AM, Robbins TW (2002) Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. J Neurosci 22:4563–4567.

Cools R, Lewis SJ, Clark L, Barker RA, Robbins TW (2007) L-DOPA disrupts activity in the nucleus accumbens during reversal learning in Parkinson's disease. Neuropsychopharmacology 32:180–189.

Dauer W, Przedborski S (2003) Parkinson's disease: mechanisms and models. Neuron 39:889–909.

Daw ND, Doya K (2006) The computational neurobiology of learning and reward. Curr Opin Neurobiol 16:199–204.

Defer GL, Widner H, Marie RM, Remy P, Levivier M (1999) Core assessment program for surgical interventional therapies in Parkinson's disease (CAPSIT-PD). Mov Disord 14:572–584.

Fiorillo CD, Tobler PN, Schultz W (2003) Discrete coding of reward probability and uncertainty by dopamine neurons. Science 299: 1898–1902.

Folstein MF, Folstein SE, Mchugh PR (1975) Mini-Mental State—practical method for grading cognitive state of patients for clinician. J Psychiatr Res 12:189–198.

Frank MJ, Seeberger LC, O'Reilly RC (2004) By carrot or by stick: cognitive reinforcement learning in Parkinsonism. Science 306: 1940–1943.

Goetz CG, Stebbins GT, Chmura TA, Fahn S, Klawans HL, Marsden CD (1995) Teaching tape for the motor section of the Unified Parkinson's Disease Rating Scale. Mov Disord 10:263–266, tape.

Grace AA (1991) Phasic versus tonic dopamine release and the modulation of dopamine system responsivity—a hypothesis for the etiology of schizophrenia. Neuroscience 41:1–24.

Hoehn M, Yahr M (1967) Parkinsonism: onset, progression, and mortality. Neurology 17:427–442.

Horvitz JC, Choi WY, Morvan C, Eyny Y, Balsam PD (2007) A "good parent" function of dopamine: transient modulation of learning and performance during early stages of training. Ann N Y Acad Sci 1104:270–288.

Kish SJ, Shannak K, Hornykiewicz O (1988) Uneven pattern of dopamine loss in the striatum of patients with idiopathic Parkinson's disease—pathophysiologic and clinical implications. N Engl J Med 318:876–880.

Krebs HI, Hogan N, Hening W, Adamovich SV, Poizner H (2001) Procedural motor learning in Parkinson's disease. Exp Brain Res 141:425–437.

Kunig G, Leenders KL, Martin-Solch C, Missimer J, Magyar S, Schultz W (2000) Reduced reward processing in the brains of parkinsonian patients. Neuroreport 11:3681–3687.

Lilliefors HW (1967) On Kolmogorov-Smirnov test for normality with mean and variance unknown. J Am Stat Assoc 62:399–402.

Martin-Soelch C, Leenders KL, Chevalley AF, Missimer J, Kunig G, Magyar S, Mino A, Schultz W (2001) Reward mechanisms in the brain and their role in dependence: evidence from neurophysiological and neuroimaging studies. Brain Res Brain Res Rev 36:139–149.

Messier J, Adamovich S, Jack D, Hening W, Sage J, Poizner H (2007) Visuomotor learning in immersive 3D virtual reality in Parkinson's disease and in aging. Exp Brain Res 179:457–474.

Montague PR, Dayan P, Sejnowski TJ (1996) A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J Neurosci 16:1936–1947.

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006) Midbrain dopamine neurons encode decisions for future action. Nat Neurosci 9:1057–1063.

Oldfield RC (1971) Assessment and analysis of handedness—Edinburgh Inventory. Neuropsychologia 9:97–113.

Rangel A, Camerer C, Montague PR (2008) A framework for studying the neurobiology of value-based decision making. Nat Rev Neurosci 9:545–556.

Robbins H (1952) Some aspects of the sequential design of experiments. Bull Amer Math. Soc 58:527–535.
Schultz W (2007) Multiple dopamine functions at different time courses. Annu Rev Neurosci 30:259–288.
Schultz W, Dayan P, Montague PR (1997) A neural substrate of prediction and reward. Science 275:1593–1599.
Shohamy D, Myers CE, Kalanithi J, Gluck MA (2008) Basal ganglia and dopamine contributions to probabilistic category learning. Neurosci Biobehav Rev 32:219–236.
Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT Press.
Swainson R, Rogers RD, Sahakian BJ, Summers BA, Polkey CE, Robbins TW (2000) Probabilistic learning and reversal deficits in patients with Parkinson's disease or frontal or temporal lobe lesions: possible adverse effects of dopaminergic medication. Neuropsychologia 38:596–612.
Wilkinson L, Jahanshahi M (2007) The striatum and probabilistic implicit sequence learning. Brain Res 1137:117–130.
Wilkinson L, Lagnado DA, Quallo M, Jahanshahi M (2008) The effect of feedback on non-motor probabilistic classification learning in Parkinson's disease. Neuropsychologia 46:2683–2695.

(Accepted 16 July 2009)
(Available online 21 July 2009)