Behavioral/Systems/Cognitive

Dopaminergic Drugs Modulate Learning Rates and Perseveration in Parkinson's Patients in a Dynamic Foraging Task

Robb B. Rutledge,[1] **Stephanie C. Lazzaro,**[1] **Brian Lau,**[2] **Catherine E. Myers,**[3] **Mark A. Gluck,**[4] **and Paul W. Glimcher**[1]

[1]Center for Neural Science, New York University, New York, New York 10003, [2]Department of Neuroscience, Columbia University, New York, New York 10027, and [3]Department of Psychology and [4]Center for Molecular and Behavioral Neuroscience, Rutgers University, Newark, New Jersey 07102

Making appropriate choices often requires the ability to learn the value of available options from experience. Parkinson's disease is characterized by a loss of dopamine neurons in the substantia nigra, neurons hypothesized to play a role in reinforcement learning. Although previous studies have shown that Parkinson's patients are impaired in tasks involving learning from feedback, they have not directly tested the widely held hypothesis that dopamine neuron activity specifically encodes the reward prediction error signal used in reinforcement learning models. To test a key prediction of this hypothesis, we fit choice behavior from a dynamic foraging task with reinforcement learning models and show that treatment with dopaminergic drugs alters choice behavior in a manner consistent with the theory. More specifically, we found that dopaminergic drugs selectively modulate learning from positive outcomes. We observed no effect of dopaminergic drugs on learning from negative outcomes. We also found a novel dopamine-dependent effect on decision making that is not accounted for by reinforcement learning models: perseveration in choice, independent of reward history, increases with Parkinson's disease and decreases with dopamine therapy.

Introduction

Midbrain dopamine neurons are thought to play a critical role in reinforcement learning. Electrophysiological recordings from the ventral tegmental area and substantia nigra in animals suggest that dopamine neurons encode the reward prediction error (RPE) signal hypothesized to guide action value learning in contemporary reinforcement learning (RL) models (Schultz et al., 1997; Hollerman and Schultz, 1998; Nakahara et al., 2004; Bayer and Glimcher, 2005). According to standard RL models, action values are updated on a trial-by-trial basis using a RPE term, the difference between the experienced and predicted reward (Rescorla and Wagner, 1972; Sutton and Barto, 1998). The phasic activity of midbrain dopamine neurons is widely hypothesized to carry this error term, possibly after multiplication by a learning rate term (for review of the dopaminergic RPE hypothesis, see Niv and Montague, 2009).

Parkinson's disease is characterized by a dramatic loss of dopamine neurons in the substantia nigra (Dauer and Przedborski, 2003). Several studies have found general learning deficits accompanying this loss in Parkinson's patients (Knowlton et al., 1996; Swainson et al., 2000; Czernecki et al., 2002; Shohamy et al., 2004, 2009), and dopaminergic medication has been found to affect performance in many tasks (Cools et al., 2001, 2006; Frank et al., 2004, 2007b; Shohamy et al., 2006; Bódi et al., 2009). However, no study has ever fit standard RL models to choice behavior in Parkinson's patients on and off dopaminergic medications to quantitatively test predictions the dopaminergic RPE hypothesis makes about learning rates.

Parkinson's disease is typically treated with levodopa (L-DOPA), the biosynthetic precursor to dopamine (Hornykiewicz, 1974), which is thought to increase phasic dopamine release (Keller et al., 1988; Wightman et al., 1988; Harden and Grace, 1995). If phasic dopamine activity encodes a RPE signal of some kind, then L-DOPA should modulate RPE magnitude. According to theory, this effect should manifest itself as a change in the learning rate estimated in standard RL models. This prediction stems from the fact that, in standard models, values placed on actions are updated by the product of the RPE and the learning rate. Thus, if dopamine carries either this product or simply the RPE signal, changes in learning rates will capture the effects of dopaminergic manipulations.

To test this prediction, we studied decision making in Parkinson's patients using a dynamic foraging task in which subjects could be expected to learn the value of actions using a reinforcement learning mechanism. We also tested whether dopaminergic drugs including L-DOPA (combined in some patients with dopamine receptor agonists) differentially affect learning from positive and negative outcomes as some previous studies suggest (Daw et al., 2002; Frank et al., 2004). We tested Parkinson's pa-

A Example trial sequence
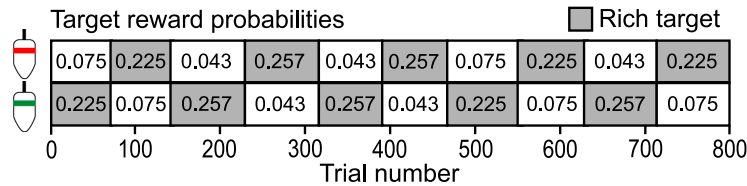

B Example block sequence

Figure 1. Experimental task design. **A**, Subjects selected one of two crab traps marked by red and green buoys and earned 10 cents for each crab caught. Example unrewarded and rewarded trials are shown. Earnings for the past 40 trials were displayed on the screen. **B**, Example block sequence. Relative reward rates (6:1, 3:1, 1:3, 1:6) changed in blocks of 70–90 trials separated by unsignaled transitions. The identity of the rich (higher reward probability) option alternated between consecutive blocks. Subjects completed 800 trials in 10 blocks.

tients both on and off dopaminergic medication, in addition to testing both healthy young and elderly control subjects, and fit choice behavior with standard RL models to test quantitative predictions of the dopaminergic RPE hypothesis.

Materials and Methods

Subjects. Seventy-eight paid volunteers participated in the experiment: 26 patients with Parkinson's disease (12 females; mean age, 65.7 years), 26 matched healthy elderly control subjects (12 females; mean age, 67.3 years), and 26 healthy young subjects (14 females; mean age, 22.8 years). Parkinson's patients were diagnosed with idiopathic Parkinson's disease and recruited by neurologists. Patients were at the mild to moderate stages of the disease, with scores on the Hoehn–Yahr scale of motor function (Hoehn and Yahr, 1967) of 2 or 2.5. We used the motor exam (section III) of the Unified Parkinson's Disease Rating Scale (UPDRS) (Lang and Fahn, 1989) to quantify symptom severity at the time of testing. All patients participated in two sessions, one on and one off dopaminergic medication, usually (94%; 49 of 52) in the morning. For the "on" session, patients were tested an average of 1.6 h after a dose of dopaminergic medication. For the "off" session, patients refrained from taking all dopaminergic medications for a minimum of 10 h (mean, 14.4 h). Session order was randomized across patients (11 completed the off session first). All patients were receiving treatment with L-DOPA, the precursor for dopamine, and the majority were also taking a D_2 receptor agonist ($n = 17$). Some patients were also taking serotonergic ($n = 9$) or cholinergic ($n = 4$) medications (for medication information, see supplemental data, available at www.jneurosci.org as supplemental material). Care was taken to minimize potential serotonergic and cholinergic medication effects by having subjects take those drugs at the same time before both on and off sessions. However, it is possible that the chronically administered medications taken by our patients, and not by any elderly subjects, might affect serotonergic, cholinergic, adrenergic, or noradrenergic transmission and potentially influence choice behavior. This limitation is common to studies of decision making in Parkinson's patients.

Parkinson's patients and elderly subjects were screened for the presence of any neurological disorder in addition to Parkinson's disease and any history of psychiatric illness including depression. The Parkinson's and elderly subject groups did not differ significantly in age, years of education, verbal intelligence quotient, or several other neuropsychological measures (for details, see supplemental Table 1, available at www.

jneurosci.org as supplemental material). All Parkinson's patients and elderly subjects gave informed consent in accordance with the procedures of the Rutgers Institutional Review Board for the Protection of Human Subjects of Rutgers University (Newark, NJ). All healthy young participants gave informed consent in accordance with the procedures of the University Committee on Activities Involving Human Subjects of New York University.

Behavioral task training. After reading task instructions, subjects answered five multiple-choice questions to ensure that they had a basic understanding of the task. They then completed 200 trials of training in five separate blocks of 40 trials in which they were informed that the reward probabilities were fixed within a block and would not change. On each trial, subjects could choose either the red or the green option, animated crab traps attached to red and green buoys (Fig. 1A). When a reward had been scheduled for their chosen option, the chosen trap was raised from the ocean to reveal a crab inside. Otherwise, the chosen trap was revealed to be empty. The probability ratio specifying the relative values of the two traps in any block was either 6:1 or 1:6, with actual probabilities summing to 0.3 within each block. Subjects were asked to verbally identify the rich (higher reward probability) option after each training block. They were given feedback as to whether or not they were correct. Subjects were not paid according to performance in the training blocks, and all subjects earned $5 for completing the instructions and training.

Behavioral task experiment. Subjects then completed 800 trials in the dynamic task environment as they tried to maximize earnings (Fig. 1B). The precise mathematical structure of the task replicated the critical features of the concurrent variable-interval tasks used by Herrnstein (1961) to formulate the matching law, which describes how animals make choices among options that differ in expected value. Monkeys performing this type of task allocate their choices according to reward probabilities and dynamically track changing probabilities (Platt and Glimcher, 1999; Sugrue et al., 2004; Corrado et al., 2005; Lau and Glimcher, 2005). Measuring choice under these conditions thus reveals the subject's expectations about the relative value of possible actions. Subjects completed 10 blocks of 70–90 trials with four possible reward probability ratios (6:1, 3:1, 1:3, 1:6). Blocks were separated by unsignaled transitions in which the identity of the higher reward probability option reversed, but with the reward probability otherwise unpredictable. Once a reward was scheduled for a trap, it remained available until the associated trap was chosen. This meant that the longer a trap remained unchosen, the greater the probability that a reward would be earned by choosing it. We used this reward schedule because monkeys performing a similar task have been shown to make choices consistent with reinforcement learning (Lau and Glimcher, 2005). The display indicated the number of crabs caught over the past 40 trials (or since the start of the experiment for the first 40 trials). Subjects were paid according to performance, earning 10 cents for each crab caught. At the end of the session, the total catch was revealed and subjects were paid accordingly. Subjects had unlimited time to make each choice but typically completed 800 trials in <30 min. Finally, subjects answered six multiple-choice questions about the experiment (for preexperiment and postexperiment questionnaires, see supplemental data, available at www.jneurosci.org as supplemental material). We computed multiple measures of task performance and fit choice data with multiple models. We used both paired and unpaired two-tailed t tests to compare behavioral measures. We used Wald tests to compare model parameter estimates across subject groups.

Matching law analysis. We fit steady-state choice behavior using the logarithmic form of the generalized matching law (Baum, 1974):

$$\log_2\left(\frac{C_R}{C_G}\right) = a \log_2\left(\frac{R_R}{R_G}\right) + \log_2 c. \quad (1)$$

Here, C_R and C_G are the number of choices to the red and green options, respectively, and R_R and R_G are the number of rewards received from choices to the red and green options, respectively. In each block, we allowed 20 trials for choice behavior to stabilize and then included 50 trials in this analysis, fitting a line by least-squares regression. By the generalized matching law (Baum, 1974), the slope of this line (a) is the reward sensitivity, a measure of the sensitivity of choice allocation to reward frequency.

Single α reinforcement learning model. We fit choice data from all subject groups with a standard RL model (Sutton and Barto, 1998). The model uses the sequence of choices and outcomes to estimate the expected value of each option for every trial. The expected values are set to zero at the beginning of the experiment, and after each trial, the value of the chosen option [for example, $V_R(t)$ for the red option at trial t] was updated according to the following rule:

$$V_R(t + 1) = V_R(t) + \alpha\delta(t) \quad (2)$$

$$\delta(t) = R_R(t) - V_R(t). \quad (3)$$

Here, $\delta(t)$ is the RPE, the difference between the experienced and expected reward. $R_R(t)$ represents the outcome received from the red option on trial t with a value of 1 for a reward and 0 otherwise. The learning rate α determines how rapidly the estimate of expected value is updated. If the learning rate is high, recent outcomes have a relatively greater influence on the expected value than less recent outcomes. Given expected values for both options, the probability of choosing the red option $P_R(t)$ is computed using the following softmax rule:

$$P_R(t) = \frac{1}{1 + \exp(-[\beta(V_R(t) - V_G(t)) + c(C_R(t - 1) - C_G(t - 1))])}. \quad (4)$$

Here, β is a noise parameter and $C_R(t - 1)$ and $C_G(t - 1)$ represent the choice of the red or green option on the previous trial $t - 1$ with a value of 1 for the chosen option and 0 otherwise [$C_R(t) = 1 - C_G(t)$]. The choice perseveration parameter c captures tendencies to perseverate or alternate (when positive or negative, respectively) that are independent of reward history (Lau and Glimcher, 2005; Schönberg et al., 2007). This parameter is similar to b_1 in the linear regression model below. The constants α (learning rate), β (noise parameter), and c (choice perseveration parameter) were estimated by maximum likelihood (Burnham and Anderson, 2002).

Dual α reinforcement learning model. We also fit a second learning model closely related to the standard RL model, which has been proposed previously (Daw et al., 2002; Frank et al., 2007a) and for which there is growing evidence from Parkinson's studies (Frank et al., 2004, 2007b; Cools et al., 2006; Bódi et al., 2009). This model is identical with the one specified above except that it uses separate learning rates for positive and negative outcomes. In this model, the value of the chosen option (for example, the red option) was updated according to the following rule:

$$\begin{array}{ll} \text{If } \delta(t) \geq 0, & V_R(t + 1) = V_R(t) + \alpha_{positive}\delta(t) \\ \text{If } \delta(t) < 0, & V_R(t + 1) = V_R(t) + \alpha_{negative}\delta(t). \end{array} \quad (5)$$

Parameter estimation for reinforcement learning models. For single and dual α RL models, we fit choice data across Parkinson's and elderly subject groups (excluding young subjects) with a single shared noise parameter and separate learning rate and choice perseveration parameters for each subject group. We estimated all parameters simultaneously by maximum likelihood (Burnham and Anderson, 2002). To determine whether learning rates and choice perseveration parameters are affected by age, we also fit the single α RL model separately for each subject in the Parkinson's and elderly subject groups with a single shared noise parameter across all subjects. We used a shared noise parameter because the learning rate and noise parameter estimates are not fully independent, and Schönberg et al. (2007) have shown that

leaving both parameters fully unconstrained can lead to interpretability problems. One solution is to fix one of the parameters to examine specific hypotheses regarding the other parameter, which remains unconstrained. The noise parameter can be fixed, for example, to test for dopaminergic influences on the learning rate. Alternatively, the learning rate can be fixed to test for dopaminergic influences on subject randomness (stochasticity). Since we wanted to test whether dopaminergic manipulations influence behavior in a way that can be captured by changes in the learning rate, we fit a single noise parameter across Parkinson's and elderly subject groups.

Hypothesis testing. We used these RL models to compare our Parkinson's and elderly subject groups as well as to compare the more and less severely affected of our Parkinson's patients. Knowlton et al. (1996) found that the one-half of their Parkinson's patients with the most severe symptoms had worse performance on a probabilistic classification task than all patients combined. We therefore split our Parkinson's patients into two groups of equal sizes according to symptom severity measured in the off medication session by the UPDRS motor exam to determine whether disease progression affects choice behavior.

The dopaminergic RPE hypothesis predicts that learning rates should be higher in Parkinson's patients on than off dopaminergic medication and lower in patients off medication than elderly control subjects. It also predicts that learning rates should be lower in the more than in the less severely affected Parkinson's subgroup. The hypothesis makes no specific predictions for how choice perseveration parameters might be modulated by dopamine levels nor does it explicitly predict how learning rates for positive and negative outcomes might differ across subject groups. For all comparisons for which we have specific predictions, we report uncorrected p values for these ex ante hypotheses. For analyses about which we had no ex ante hypotheses, we also report p values Bonferroni-corrected for multiple comparisons.

L-DOPA is thought to increase phasic dopamine release (Keller et al., 1988; Wightman et al., 1988; Harden and Grace, 1995). Value estimates in the standard RL model (Eq. 2) are updated by the product $\alpha\delta(t)$, which we refer to as the error correction term. If phasic dopamine activity encodes a RPE signal, then treatment with L-DOPA should affect value learning by amplifying the error correction term, and this manipulation will be reflected in choice behavior by a change in the learning rate. Despite the fact that the dopaminergic RPE hypothesis is silent about the role of tonic dopamine activity in behavior, it is important to note that both phasic and tonic dopamine signaling may well be affected by the dopamine depletion associated with Parkinson's disease and by the dopaminergic drugs taken to treat the disorder. There is now growing evidence that phasic and tonic dopamine activity may play distinct roles in reinforcement learning (Niv et al., 2007), and as those phasic/tonic models evolve it may become possible to test additional hypotheses with these data.

A final note concerns the interpretation of our hypothesis tests with regard to the dopaminergic RPE hypothesis. Value estimates are updated by the product $\alpha\delta(t)$. This implies that a difference in observed learning rates between subject groups might reflect the role of dopamine in encoding α, $\delta(t)$, or the product $\alpha\delta(t)$. In electrophysiological studies, phasic dopaminergic firing rates are observed to vary under conditions in which α is believed to be constant (Schultz et al., 1997; Bayer and Glimcher, 2005), implying that the phasic activity of dopamine neurons does not encode α alone. These data make it unlikely that L-DOPA, by increasing the dopamine release associated with action potentials, could uniquely influence α without affecting $\delta(t)$. It is possible that other dopaminergic manipulations, including dopamine receptor agonists taken by many of our subjects, could uniquely affect α. Regardless, even the electrophysiological data give no guidance for determining whether L-DOPA should influence $\delta(t)$ or $\alpha\delta(t)$. Model fits would reveal a change in learning rate in either case. For this reason, changes in α estimated from behavior by RL models should be interpreted as evidence that either $\delta(t)$ or $\alpha\delta(t)$ is influenced by L-DOPA administration.

Linear regression model. In an effort to test the robustness of any conclusions drawn from the fits of our RL models, we also used a linear regression approach to fit our choice data [as in Lau and Glimcher (2005)]. To perform this robustness check, we assumed that influences of past rewards and choices were linearly combined to determine choice on each trial, with choice probability computed using the softmax rule, as in the RL model. We used logistic regression to estimate weights for rewards received and choices
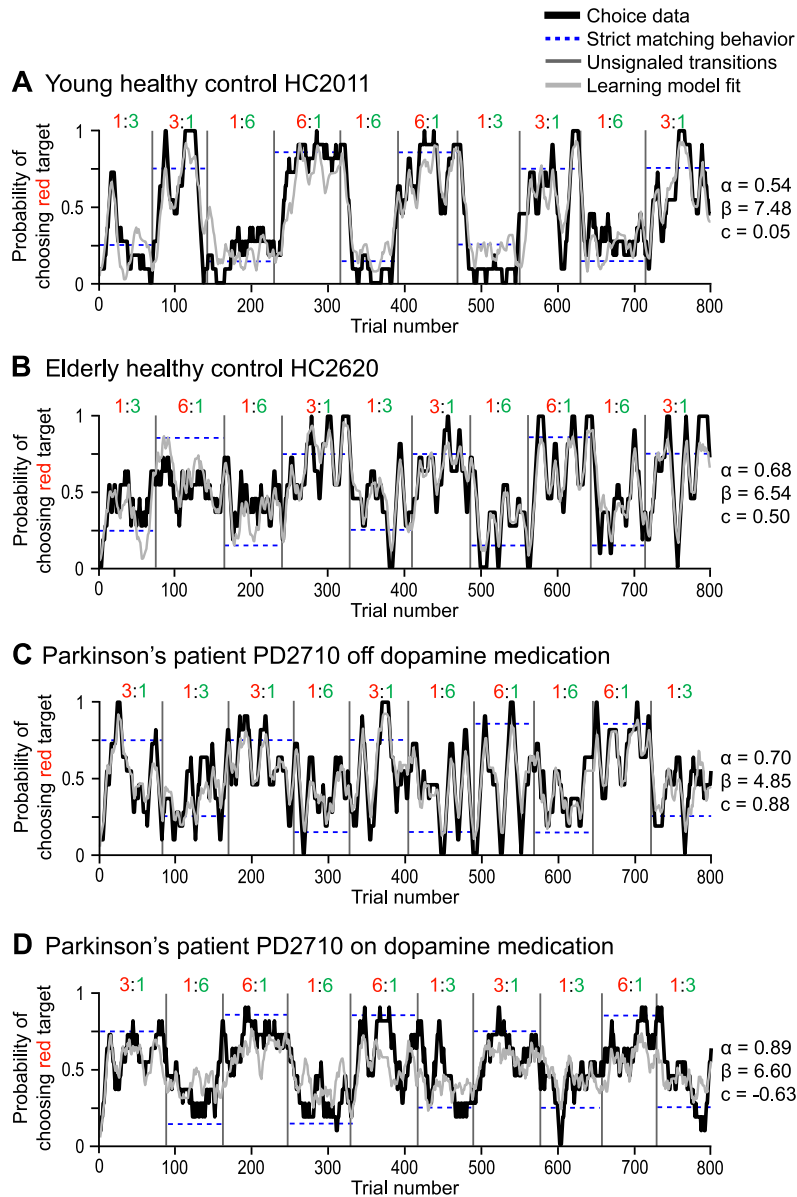
Figure 2. Observed and predicted choice for example subjects. **A**, Healthy young subject HC2011. **B**, Healthy elderly subject HC2620. **C, D**, Parkinson's patient PD2710 tested both on and off dopaminergic medication. Data were fit with a learning model with three parameters (learning rate α, noise parameter β, and choice perseveration parameter c). Choice data (black) and model predictions (gray) were smoothed with an 11-trial moving average. The vertical lines indicate unsignaled transitions between blocks. The horizontal lines indicate strict matching behavior (Herrnstein, 1961). Individual parameter estimates are indicated. The block sequence for young example subject HC2011 is displayed in Figure 1*B*.

made on previous trials, so the noise parameter of the softmax rule is effectively incorporated into reward weights by the regression. The goal of the regression was to estimate the probabilities $P_R(t)$ and $P_G(t)$ of choosing the red and green options, respectively. Since there are only two options and assuming symmetric weights for the two options, the model for 10 previous trials reduces to the following:

$$\log\left(\frac{P_R(t)}{P_G(t)}\right) = \sum_{i=1}^{10} a_i (R_R(t-i) - R_G(t-i))$$

$$+ \sum_{i=1}^{10} b_i (C_R(t-i) - C_G(t-i)). \quad (6)$$

Here, a and b coefficients represent changes in the log odds of choosing the red or green options with a_i the weight for a reward received i trials ago and b_i the weight for a choice made i trials ago. Negative weights indicate de-

creases in the log odds of choice as a function of previous rewards (a_i) or choices (b_i). The log odds of the subject making a given choice on a specific trial is obtained by linearly combining previous choices and outcomes (rewards), weighted by the coefficients extracted by the regression. The logistic regression is linear in log odds but nonlinear in choice probability. Choice probability is recovered from the log odds by exponentiating both sides of Equation 6 and solving for $P_R(t)$. This formulation has been used to study choice and striatal function in monkeys (Lau and Glimcher, 2005, 2008). If all b_i are set to zero, this is similar to the linear reward model that has been used in several previous studies of choice in monkeys (Sugrue et al., 2004; Corrado et al., 2005; Kennerley et al., 2006).

If the RL models discussed above perfectly described behavior, then the weights a_i would decline exponentially with the decay rate specified by the learning rate term in the RL models. The linear regression model thus relaxes the constraint, imposed by the RL models presented above, that weights must decline exponentially. The inclusion of multiple b_i terms relaxes the constraint in the specific RL models we used that, independent of rewards, only the previous choice may influence future choice. The set of b_i allows long-term trends in choice to be identified. Comparing these less constrained linear regression analyses to the standard RL models thus allows us to examine the robustness of the conclusions drawn with the more structured RL model-based approach.

Model comparison. To evaluate model fits, we computed a pseudo-R^2 statistic (Camerer and Ho, 1999) using the following equation:

$$\text{pseudo-}R^2 = (R - L)/R. \quad (7)$$

Here, L is the maximum log likelihood for the estimated model given the data and R is the log likelihood of the model under random choice. To compare the RL model and linear regression approaches, we penalized model fits for complexity using the Bayesian information criterion (BIC) (Schwarz, 1978). We computed BIC using the following equation:

$$\text{BIC} = -2 \log L + k \log n. \quad (8)$$

Here, L is the maximum log likelihood for the estimated model given the data, k is the number of free parameters in the model, and n is the number of trials. The model with the lower BIC is preferred.

Results
Behavioral results
Choice data are shown for example subjects from each subject group including on and off medication sessions for one Parkinson's patient (Fig. 2). We found that subjects from all four groups chose the richer option more frequently in the training session (Fig. 3*A*). Subjects correctly identified the richer option in all five training blocks in most (88%; 84 of 96) sessions. In the dynamic task environment, choice behavior after transitions to different reward probability ratios quickly stabilized at the group level, suggesting that subjects in all four groups adjusted choice behavior according to option reward rates (Fig. 3*B*).

Parkinson's patients earned similar amounts on (mean ± SEM, $19.50 ± 0.34 excluding one subject who did not complete all 800 trials) and off ($19.45 ± 0.26) dopaminergic medication (paired t test, $t_{(24)} = 0.31, p = 0.76$). Young control subjects earned more ($20.54 ± 0.23) than Parkinson's patients either on (unpaired t test, $t_{(49)} = 2.55, p = 0.014$) or off medication ($t_{(50)} = 3.17, p = 0.003$). Elderly subjects also earned more ($20.13 ± 0.34), although not significantly, than Parkinson's patients either on ($t_{(49)} = 1.32, p = 0.19$) or off medication ($t_{(50)} = 1.60, p = 0.11$). All four subject groups allocated more choices to the richer option in high-probability (6:1, 1:6) than low-probability (3:1, 1:3) blocks (young, paired t test, $t_{(25)} = 4.50, p = 0.0001$; elderly, $t_{(25)} = 1.96, p = 0.061$; Parkinson's on, $t_{(25)} = 2.29, p = 0.031$; Parkinson's off, $t_{(25)} = 2.80, p = 0.01$) (Fig. 3C–F). We fit a line to determine how the log ratio of choices to the two options at steady state related to the log ratio of rewards from the two options, which provides a measure of reward sensitivity (Fig. 4) (for fits, see supplemental Table 2, available at www.jneurosci.org as supplemental material). Linear fits accounted for 42–65% of the variance within each subject group. Choice was lawfully related to reward rates, suggesting that subjects in all four subject groups made their choices based on the relative values of the two options.

Reward sensitivity in the young group was comparable with measures from young adult monkeys performing a similar task (Lau and Glimcher, 2005) (see supplemental data, available at www.jneurosci.org as supplemental material). Reward sensitivity for elderly subjects was intermediate but not significantly different from Parkinson's patients either on (unpaired t test, $t_{(472)} = 1.11, p = 0.27$) or off medication ($t_{(473)} = 1.36, p = 0.17$). Reward sensitivity was higher in Parkinson's patients on than off medication ($t_{(476)} = 2.34, p = 0.02$), and this increase in choice allocation to the richer option while on dopaminergic medication is consistent with dopamine playing a role in reinforcement learning.

Reinforcement learning models
We analyzed individual trial-by-trial choice data using a standard RL model with a learning rate α and a noise parameter β. The average individual pseudo-R^2 was 0.18 (SD = 0.12; $n = 104$), significantly better than a model predicting random choice for most sessions (88%, 92 of 104; likelihood ratio test, $p < 0.05$), suggesting that a standard RL model can account for choice behavior in our task. Because choice behavior might also depend on previous choices independent of previous rewards, we also included a choice perseveration parameter c that captures short-term tendencies to alternate or perseverate (when the value of c is negative or positive, respectively) that are independent of reward history (Lau and Glimcher, 2005; Schönberg et al., 2007). The average individual pseudo-R^2 for the RL model with this additional parameter was 0.27 (SD = 0.16; $n = 104$), significantly

A Training

B Session

C Young healthy controls

D Elderly healthy controls

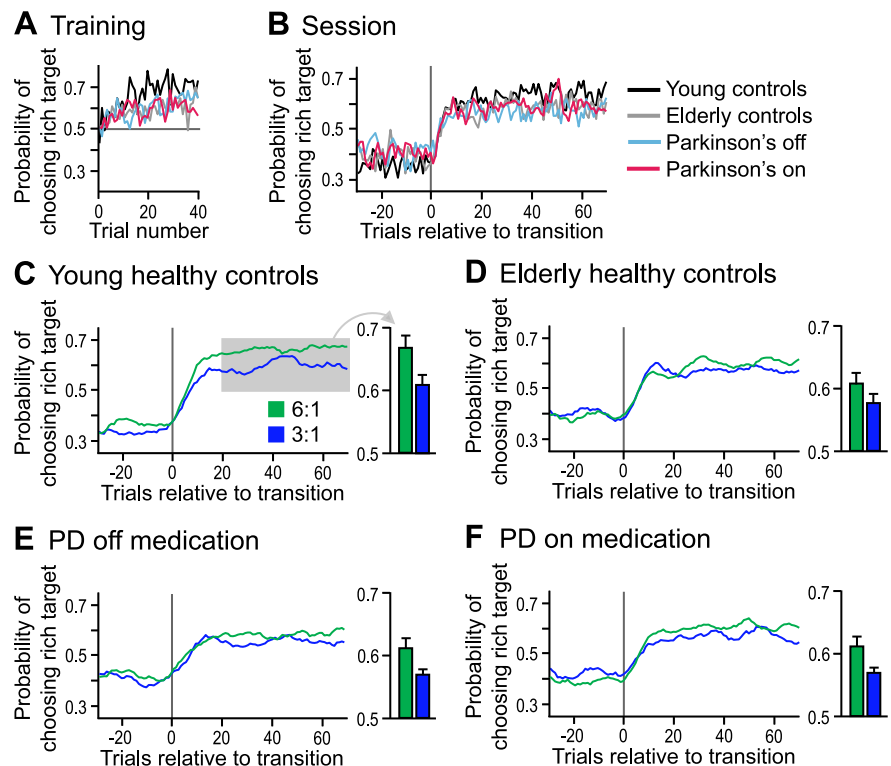**E** PD off medication

F PD on medication

Figure 3. Behavioral results. **A**, Average choice behavior in the training session for young and elderly control subjects (both $n = 26$) and Parkinson's patients on and off medication (both $n = 22$). Subjects instructed to identify the rich (higher reward probability) option chose the rich option more than expected by chance (indicated by a horizontal line). Data are unsmoothed. **B**, Average choice behavior in the experimental session for all four subject groups ($n = 26$). Subjects experienced multiple unsignaled transitions between blocks and all groups reacted quickly to transitions. The probability of choosing the option that is richer after the unsignaled transition is plotted with choices aligned on the trial on which a transition occurred (indicated by a vertical line). This convention is used in subsequent panels. Data are unsmoothed. **C–F**, Average choice behavior in all four subject groups for high-probability (6:1) and low-probability (3:1) ratio blocks. All groups allocated choices according to option reward rates, choosing the rich option more in the high-probability than low-probability ratio blocks. Data are smoothed using an 11-trial moving average. After allowing 20 trials from the beginning of each block for choice behavior to stabilize, choices are averaged separately for high-probability and low-probability ratio blocks for the next 50 trials and plotted at the right of each panel. Error bars reflect ±SEM across subjects.

better than a model predicting random choice for all but one session (99%, 103 of 104; likelihood ratio test, $p < 0.05$) and not significantly different between the two Parkinson's groups (paired t test, $t_{(25)} = 0.63, p = 0.54$) or between elderly subjects and either Parkinson's group (unpaired t test, both $t_{(50)} < 0.91$, $p > 0.3$). We plot individual fits for this model for example subjects in Figure 2 (pseudo-R^2, 0.30–0.40).

To characterize choice behavior across subject groups, we fit choice data with a single shared noise parameter across Parkinson's and elderly groups ($\beta = 1.73 ± 0.03$; SEs are indicated for all parameter estimates) and separate learning rate and choice parameters for each group. The group parameter estimates are plotted in Figure 5 relative to parameter estimates for elderly subjects ($\alpha = 0.60 ± 0.02$; $c = 0.39 ± 0.02$; $n = 26$). Learning rates for Parkinson's patients off medication were similar to learning rates in elderly control subjects (Wald test, $p = 0.44$).

We checked whether model parameters were affected by disease progression by splitting the Parkinson's patients according to disease severity, into more and less affected subgroups of equal sizes. Learning rates were lower in the more than in the less affected Parkinson's subgroup off medication (Wald test, $p < 0.0001$; both $n = 13$; noise parameter fixed at $\beta = 1.73$). This difference cannot be accounted for by increased motor symptoms impairing task performance since reaction times, measured from stimulus onset to response completion, were similar off
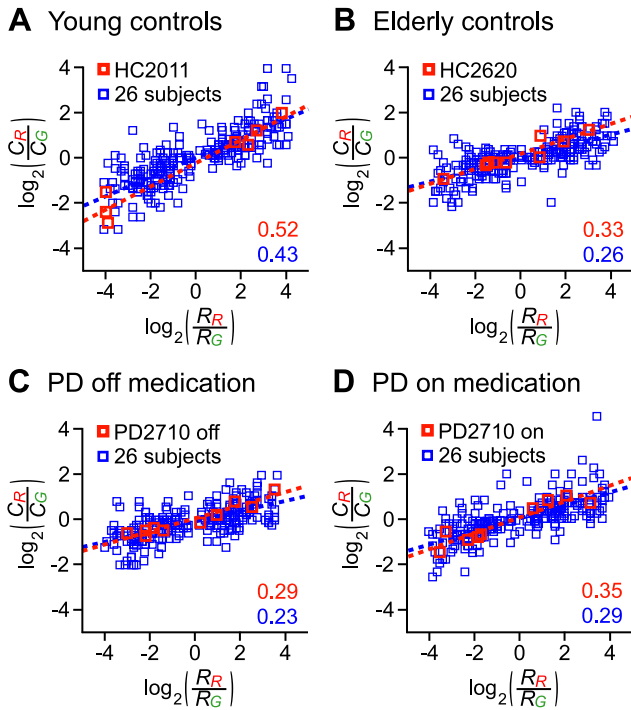
A Young controls

B Elderly controls

C PD off medication

D PD on medication

Figure 4. Steady-state choice behavior obeys the matching law. *A–D*, Log choice ratios (ratio of red to green choices) are plotted as a function of log reward ratios for 50 trials of each block after allowing 20 trials for choice behavior to stabilize (the same period used in Fig. 3*C–F*). Data are plotted for all subjects in a group and the example subjects (in red) from Figure 2*A–D*. All four groups obey the matching law (all $n = 26$), allocating choices as a function of reward ratios. Blocks in which one option was never rewarded in the 50-trial period are excluded. The lines represent least-squares fits of the generalized matching law with the slope, corresponding to reward sensitivity, noted on each plot for subject group (blue) and example subject (red). Fit parameters are listed in supplemental Table 2 (available at www.jneurosci.org as supplemental material).

A Reward effects

B Choice effects

Figure 5. Reward and choice effects for Parkinson's and elderly subject groups. *A*, *B*, Choice data for Parkinson's and elderly groups (all $n = 26$) were fit with a single shared noise parameter and group parameter estimates are plotted relative to elderly parameter estimates. The insets show fits for the more affected one-half of the Parkinson's patients with moderate symptoms ($n = 13$). All differences between Parkinson's patients on and off dopaminergic medication are highly significant (all $p < 0.001$, indicated by stars). Parkinson's patients have higher learning rates on than off dopaminergic medication. Parkinson's patients perseverated in their choices more, independent of reward history, off than on dopaminergic medication. These differences were maintained in the more affected one-half of the patients with moderate symptoms. Error bars indicate 95% confidence intervals.

A Positive outcomes

B Negative outcomes

Figure 6. Learning rates for positive and negative outcomes. *A*, *B*, Choice data for Parkinson's and elderly subject groups were fit with a shared noise parameter and group parameter estimates are plotted relative to elderly parameter estimates. Parkinson's patients have higher learning rates for positive outcomes on than off dopaminergic medication ($p = 0.003$, indicated by a star). Parkinson's patients had similar learning rates for negative outcomes on and off dopaminergic medication ($p = 0.44$). Error bars indicate 95% confidence intervals.

medication in the more affected (mean $\pm$ SEM of median reaction times, 364 ± 56 ms; $n = 13$) and the less affected Parkinson's patients (368 ± 41 ms, $n = 13$; paired t test, $t_{(24)} = 0.063$, $p = 0.95$). Learning rates were also lower in the more affected Parkinson's patients off medication than in elderly control subjects (Wald test, $p = 0.006$) (Fig. 5*A*). Surprisingly, the less affected Parkinson's subgroup had higher learning rates off medication than the elderly control subjects ($p < 0.0001$).

Most importantly, as predicted by the dopaminergic RPE hypothesis, learning rates were higher in Parkinson's patients on than off dopaminergic medication. This difference was significant at $p = 0.0003$ (Wald test) (Fig. 5*A*). We also found an additional dopamine-dependent effect on decision making. Parkinson's patients off dopaminergic medication perseverated more, independent of reward history, than elderly control subjects (Wald test, $p < 0.0001$; corrected $p < 0.0001$), and dopaminergic medication reduced this tendency ($p < 0.0001$; corrected $p < 0.0001$) (Fig. 5*B*), making them more like control subjects. The same differences in learning rate and perseveration between Parkinson's patients on and off medication were maintained in the more affected Parkinson's subgroup (both $p < 0.0001$) (Fig. 5).

Because block lengths varied within a limited range (70–90 trials), subjects might have learned to predict block transitions and to adjust their choice behavior accordingly. For example, learning rates might be higher immediately after block transitions than in later block phases and our analysis might have obscured this fact. However, we found that parameter estimates were similar in early and late block phases, and there was no indication that sub-
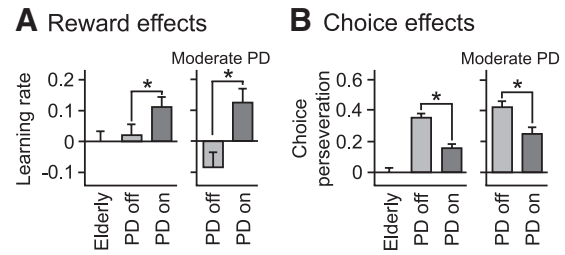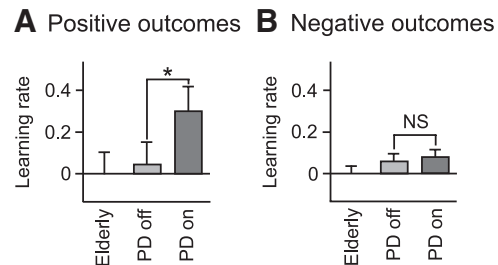
jects were able to predict block transitions (see supplemental data, available at www.jneurosci.org as supplemental material).

Previous studies have suggested that dopamine neurons might be differentially involved in learning from positive and negative outcomes (Daw et al., 2002; Frank et al., 2004, 2007b; Bayer and Glimcher, 2005; Cools et al., 2006, 2009; Frank and O'Reilly, 2006; Bayer et al., 2007; Bódi et al., 2009). To test for this possibility, we fit separate learning rates for positive and negative outcomes, as in a recent study of reinforcement learning in healthy young subjects (Frank et al., 2007a), again with a single shared noise parameter ($\beta = 1.12 \pm 0.03$) and separate learning rate and choice perseveration parameters for each Parkinson's and elderly subject group. This model fit the data significantly better than the single learning rate model after accounting for number of parameters (likelihood ratio test, $p < 0.0001$). Group choice parameters were similar to those estimated using the single learning rate model (Wald test, all $p > 0.19$) and differences between group choice parameters remained significant at $p < 0.0001$. Group parameter estimates are plotted in Figure 6 relative to parameter estimates for elderly subjects ($\alpha_{\text{positive}} = 1.20 \pm 0.05$; $\alpha_{\text{negative}} = 0.47 \pm 0.02$). We found that learning rates for positive outcomes were significantly higher in Parkinson's patients on than off dopaminergic medication (Wald test, $p = 0.003$; corrected $p = 0.016$), but dopaminergic medication did not affect learning rates for negative outcomes ($p = 0.44$; corrected $p = 1.0$). This finding supports a role for dopamine neu-

rons in learning from positive but not negative outcomes. Learning rates in Parkinson's patients off medication were similar to elderly control subjects for positive outcomes ($p = 0.58$; corrected $p = 1.0$) but were marginally higher for negative outcomes ($p = 0.032$; corrected $p = 0.19$), indicating a possible increase in sensitivity to negative outcomes in Parkinson's patients, consistent with a previous study (Frank et al., 2004).

Since the number of dopamine neurons in the substantia nigra (although not in the ventral tegmental area) declines with normal aging (Stark and Pakkenberg, 2004), we examined whether learning rate or choice perseveration parameters correlated with age. Excluding young subjects from this analysis, we fit choice data across Parkinson's and elderly subject groups with a single noise parameter ($\beta = 2.61 \pm 0.16$) and separate learning rate and choice parameters for all individuals (Fig. 7). We found no significant correlation between learning rate and age for Parkinson's or elderly groups (all $r < 0.18$; $p > 0.3$) (Fig. 7A). We did not find a correlation between choice perseveration and age in either Parkinson's group (both $r < 0.16$; $p > 0.3$), but we did find a positive correlation between choice perseveration and age in the elderly subjects ($r = 0.40$; $p = 0.043$) (Fig. 7B).

We considered whether other variables might also be correlated with perseveration in elderly subjects. Choice perseveration was significantly correlated with scores on the two memory tests (both $r > 0.51$; $p < 0.007$), but not with each of the other six demographic and neuropsychological variables (all $r < 0.22$; $p > 0.29$). We note that age and scores on the two memory tests were highly correlated (both $r > 0.72$; $p < 0.0001$ for both memory scores) and that multiple regression across demographic and neuropsychological variables revealed no significant correlations at $p < 0.05$, even without correcting for multiple comparisons. Our results thus indicate that, although perseveration might result from normal aging, perseveration might also be related to memory decline. Our sample size is simply too small to determine to what extent each contributes differentially to perseveration.

Linear regression model of reward influence on choice

Not all recent studies of value-based decision making have used a standard RL model. Several studies in monkeys have used an alternate approach that does not assume that the influence of rewards received on previous trials decays in an exponential manner (Sugrue et al., 2004; Corrado et al., 2005; Lau and Glimcher, 2005, 2008; Kennerley et al., 2006), an assumption effectively embedded in standard RL models by the learning rate α. To examine the robustness of our RL model-based findings, we fit trial-by-trial choice dynamics using a more general linear regression model (Fig. 8). Our results using this approach were broadly consistent with the results obtained using the more constrained RL models.

To perform this analysis, we assumed only that past rewards and choices were weighted and linearly combined, with no restriction on the structure of weights, to determine what choice a subject would make. Best fitting weights for past rewards and choices obtained in this way quantify changes in the probability of a subject choosing a particular option exerted by past events. A positive reward weight for
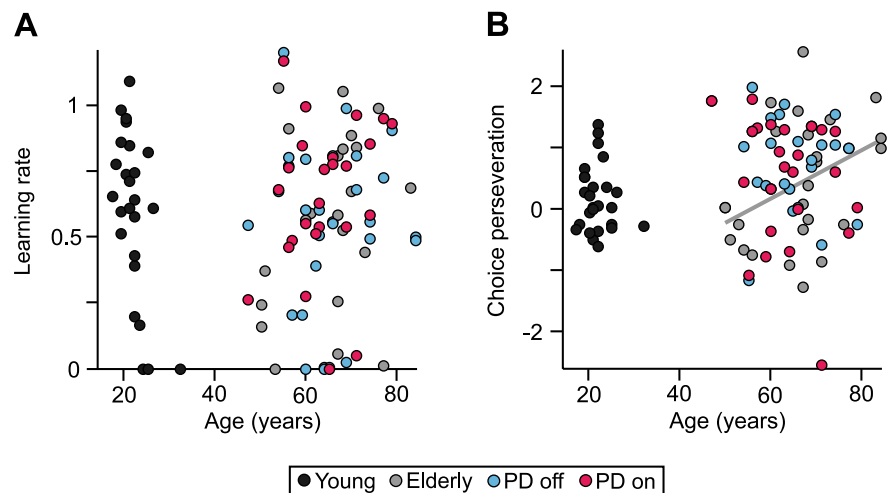
Figure 7. Effect of age on individual parameter estimates. **A, B**, Parameter estimates for individual subjects are plotted against subject age. Choice data for individual Parkinson's patients and elderly subjects were fit with a shared noise parameter. There was no significant relationship between individual learning rates and age in elderly or Parkinson's subject groups (all $p > 0.3$). Choice perseveration increased with age for elderly subjects ($p = 0.043$; the line represents the least-squares fit) but not for either Parkinson's group (both $p > 0.3$). Young subjects were excluded from these analyses, but we used the noise parameter obtained for Parkinson's and elderly groups to fit their data and plot the results for comparison purposes.

the most recent trial indicates that a reward received from the red trap (Fig. 8) had the effect of increasing the probability of choosing the red trap on the next trial. A negative choice weight on the most recent trial indicates that a choice to the red trap had the effect of decreasing the probability of choosing the red trap on the next trial, independent of past rewards. A linear regression computed in this manner effectively identifies the weights that, when summed, best describe the influence of previous rewards and choices on future choice. If these influences sum to zero, either option is equally likely to be chosen.

Model weights in log odds for rewards received and choices made for 10 previous trials were estimated by logistic regression for each of the four subject groups (Fig. 9). (A similar estimation was not made for the two Parkinson's subgroups because of the small size of those groups relative to the number of parameters being estimated.) Reward effects for all four groups (Fig. 9A,B) decayed in an approximately exponential manner, consistent with subjects using a reinforcement learning mechanism to learn action values. A negative weight for the most recent choice for the young subjects indicates a tendency to alternate independent of reward history (Fig. 9C). Reward and choice weights for young adult subjects are comparable with those for young adult monkeys performing a similar task (Lau and Glimcher, 2005).

Critically, reward weights decayed more quickly in Parkinson's patients on than off dopaminergic medication (Fig. 9B) as expected from the RL model-based analysis. As an additional check of robustness, we fit an exponential function to the reward weights and compared that time constant in Parkinson's patients on and off medication. This time constant was smaller for Parkinson's patients on than off medication (Parkinson's on, 0.96 ± 0.22; Parkinson's off, 1.67 ± 0.26; Wald test, $p = 0.037$). This result is again consistent with the higher learning rates we found in Parkinson's patients on than off medication. If subjects learn values according to standard RL models, this iterative computation can be equivalently described as an exponentially weighted average of previous rewards with the rate of decay determined by the learning rate α (Bayer and Glimcher, 2005). Thus, the fact that reward weights decay exponentially, although they are not con-
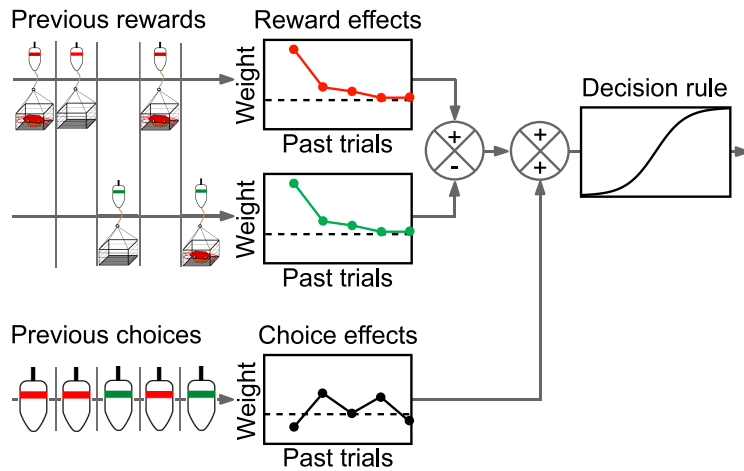
Figure 8. Two-stage description of choice. At the first stage, past rewards and choices are used to value the two options. An example sequence of five trials is shown with the most recent trial being a rewarded choice to the red option. We fit choice data with a linear model and, as an example, display reward and choice effects for young subject HC2011 from Figure 2 *A*. The relative weight of rewards received on past trials typically decays with time. Choice effects capture tendencies to alternate or perseverate that are independent of reward history. The negative weight for the most recent choice corresponds to a reward-independent tendency to alternate. A positive weight reflects a tendency to perseverate. At the second stage, the value of the options is compared, incorporating choice effects, and an option is selected using a decision rule.
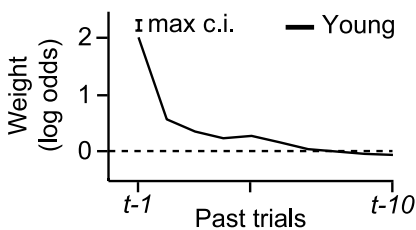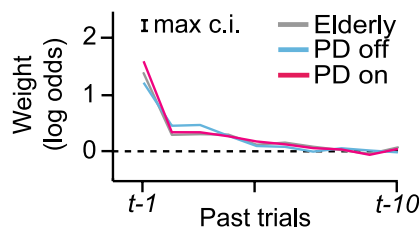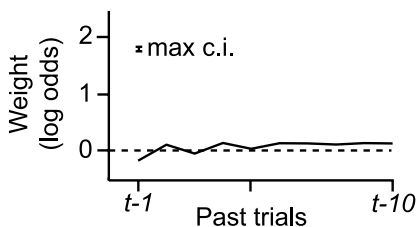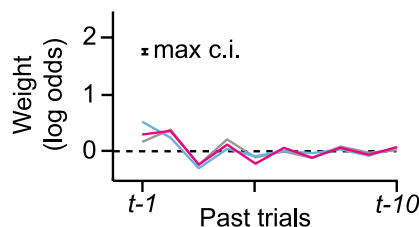

Figure 9. Reward and choice effects estimated by a linear regression model. *A*, Reward effects for young subjects. *B*, Reward effects for Parkinson's and elderly subject groups. The linear model described in Figure 8 was fit to choice data for each group (all *n* = 26 and >20,000 trials) using rewards and choices for 10 past trials. Coefficients are plotted as a function of trials in the past relative to the current trial. The decay in reward weights resembles an exponential function. Maximum 95% confidence intervals for all parameter estimates in each plot are indicated. *C*, Choice effects for young subjects. *D*, Choice effects for Parkinson's and elderly groups. The choice weight for the most recent trial captures the greater perseveration of Parkinson's patients off than on dopaminergic medication, as in Figure 5 *B*.

tion thus demonstrates that a significant part of the dopaminergic medication effect on learning must be attributable to a change in learning rate and not a change in the noise parameter.

For Parkinson's and elderly groups, a positive weight for the most recent choice captures the tendency, observed in the RL model-based analysis, to repeat the choice just made (Fig. 9 *D*). This tendency to perseverate was higher in Parkinson's patients off dopaminergic medication than elderly control subjects, but was reduced by dopaminergic medication, as indicated by our RL model analysis, with all differences between subject groups significant at $p < 0.0001$. We fit models with 10 reward weights and up to 20 choice weights and used BIC (Schwarz, 1978) to compare fits penalizing for model complexity. The most preferred model for all four subject groups had at least five choice weights, demonstrating that choice weights for previous trials capture additional variance in choice data. This is an observation made previously for monkeys performing a similar task (Lau and Glimcher, 2005).

Reward effects estimated by the linear regression and RL model approaches are plotted for example subjects (Fig. 10 *A*–*D*). To verify the appropriateness of the RL model over the more general linear regression model, we penalized model fits for complexity using BIC. We compared individual fits for the three-parameter RL model (α, β, c) to an 11-parameter linear regression model (10 reward weights, 1 choice weight) and found the (more constrained) RL model preferred for the majority (82%; 85 of 104) of sessions (Fig. 10 *E*). To verify that this finding was not attributable to the choice of reward weight number, we fit linear regression models with 1–20 weights and found the learning model always preferred for the majority (at least 59%; 61 of 104) of sessions, showing that the RL model containing an error correction term explains the data better (in the BIC sense) than the more general linear regression model. The three-parameter RL model was also preferred to a two-parameter RL model that omitted the choice parameter in the majority (83%; 86 of 104) of sessions (Fig. 10 *F*), supporting inclusion of the choice perseveration parameter and reiterating the importance of considering reward-independent choice effects on decision making in reinforcement-learning tasks.

Discussion

We used a dynamic foraging task and fit choice behavior with RL models to test the quantitative hypothesis that dopaminergic drugs affect the error correction term of the RL mechanism in humans. We found that (1) dopaminergic drugs increased learning rates in Par-

strained in any way to do so by the linear regression approach, is consistent with subjects using a RL model with an error correction term to estimate option values. Since the learning rate and noise parameter are not fully independent (Schönberg et al., 2007), it is possible that, because the noise parameter is shared across groups, changes in the noise parameter will be reflected by changes in the learning rate. Importantly, the decay rate of reward weights in the linear regression analysis provides an estimate of the rate of learning that is independent of the noise parameter. A difference in the time constants for this decay in Parkinson's patients on and off medica-
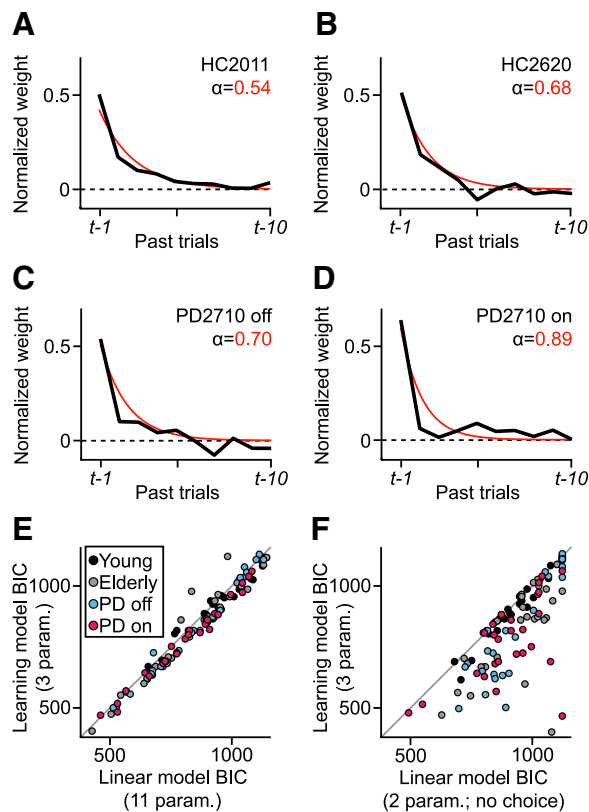
Figure 10. Comparison of RL and linear regression models. ***A–D***, In RL models, the value of an option is computed using an exponentially weighted average with a learning rate α. The linear model in Figure 8 does not constrain reward weights to decay exponentially. For the example subjects in Figure 2, linear model weights are plotted alongside the exponential function with the learning rate estimated by a three-parameter RL model, like that used in Figure 5, as its time constant. ***E***, ***F***, BIC measures indicated that the three-parameter RL model we used is preferred (lower BIC) to both an 11-parameter linear regression model with 10 reward weights and one choice weight and a two-parameter RL model without a choice parameter.

kinson's patients; (2) learning rates were similar in Parkinson's patients off dopamine medication and elderly control subjects; (3) learning rates were lower in more affected Parkinson's patients than either less affected patients or elderly control subjects; (4) dopaminergic drugs selectively increased learning rates for positive but not negative outcomes; and (5) perseveration in choice, independent of reward history, increased with normal aging and Parkinson's disease and decreased with dopamine therapy.

Human reinforcement learning and dopamine

Although dynamic foraging tasks have only recently been adapted for use in humans (Daw et al., 2006; Serences, 2008), they are commonly used to study value-based decision making in monkeys (Platt and Glimcher, 1999; Sugrue et al., 2004; Lau and Glimcher, 2005, 2008; Samejima et al., 2005). Activity in striatal neurons, which receive dense dopaminergic inputs, is correlated with trial-by-trial action values estimated from choice behavior with both approaches we used: RL models (Samejima et al., 2005) and linear regression models (Lau and Glimcher, 2008). In our task, all four subject groups adjusted to unsignaled changes in reward probabilities and allocated choices according to option reward rates, allocating more choices to options with higher reward probabilities. Steady-state choice behavior was well described by the matching law, and we found greater reward sensitivity in Parkinson's patients on than off dopaminergic medication, consistent with previous studies finding that dopaminer-

gic drugs affect performance in learning tasks (Cools et al., 2001, 2006, 2009; Frank et al., 2004, 2007b; Frank and O'Reilly, 2006; Pessiglione et al., 2006; Shohamy et al., 2006; Bódi et al., 2009).

To test quantitative predictions of the dopaminergic RPE hypothesis, we fit choice behavior using a standard RL model. We found higher learning rates in Parkinson's patients on than off dopaminergic medication. If the phasic activity of dopamine neurons encodes RPEs (Schultz et al., 1997; Hollerman and Schultz, 1998; Nakahara et al., 2004; Bayer and Glimcher, 2005), then L-DOPA, by increasing phasic dopamine release (Keller et al., 1988; Wightman et al., 1988; Harden and Grace, 1995), should affect RPE magnitude. In standard models, action values are updated by the product of RPE and the learning rate, so our observation that learning rates increase with dopaminergic medication is predicted by standard RL models. These results suggest that dopamine neurons are involved in encoding the error correction term in RL models and are consistent with the widely held hypothesis that dopamine neurons encode a RPE signal.

Surprisingly, learning rates were similar in Parkinson's patients off medication and elderly control subjects. If the RL mechanism remains relatively intact despite significant loss of dopamine neurons in the substantia nigra during the early phase of the disease, as these data suggest, then dopamine therapy might have the effect of "overdosing" this mechanism with regard to learning behavior (Gotham et al., 1988; Swainson et al., 2000; Cools et al., 2001). This might be reflected by higher learning rates in medicated Parkinson's patients than in control subjects. This is exactly what we observed. This finding may highlight the role of the ventral tegmental area in learning, an area that is relatively intact early in Parkinson's disease (Kish et al., 1988).

To test whether disease progression influences reinforcement learning, we split our Parkinson's patients into more and less affected halves. We found that the more affected subgroup had lower learning rates than both the less affected subgroup and elderly control subjects. This suggests that, as the disease progresses, learning rates decline as dopamine depletion worsens. We also found that learning rates were higher in the less affected patients than control subjects. This may reflect higher motivation in patients than control subjects because the patients knew we were studying the effects of their disorder on decision making (the Hawthorne effect) (Frank et al., 2004). In a similar way, the effects of dopaminergic medication we observed might be explained by nonspecific effects of motivation or arousal that manifest as specific changes in learning rates.

Separating positive and negative reward prediction errors

We also found that the dopamine-dependent effect on learning was selective for learning from positive RPEs and does not appear to affect learning from negative RPEs. That dopaminergic drugs affect learning from positive RPEs (or outcomes) and asymmetrically affect learning from positive and negative RPEs (or outcomes) is consistent with a growing body of evidence. Theoretical (Daw et al., 2002; Dayan and Huys, 2009) and electrophysiological (Bayer and Glimcher, 2005; Bayer et al., 2007) studies suggest that midbrain dopamine neurons may encode positive RPEs by increases in spiking activity and that a nondopaminergic system might encode negative RPEs. Pharmacological studies (Frank et al., 2004, 2007b; Cools et al., 2006, 2009; Frank and O'Reilly, 2006; Bódi et al., 2009) suggest that positive and negative outcomes have differential effects on learned values. Our results are compatible with all of these previous findings. However, our finding that dopaminergic drugs did not affect learning from negative RPEs may be at odds with pharmacological studies find-

ing evidence for greater learning from negative outcomes off than on medication (Frank et al., 2004, 2007b; Cools et al., 2006, 2009; Frank and O'Reilly, 2006; Bódi et al., 2009). The explanation for this inconsistency may lie in the subtle distinction between RPEs and outcomes, or in the precise reward patterns our subjects experienced. Future research will have to resolve this ambiguity. One interesting implication of the asymmetry we found is a possible explanation for the increased prevalence of pathological gambling in Parkinson's patients taking some dopaminergic drugs (Molina et al., 2000; Dodd et al., 2005; Voon et al., 2006; Dagher and Robbins, 2009). When learning rates for positive and negative outcomes are balanced, the probability of selecting an action accurately reflects the true value of that action. In contrast, an RL mechanism that overweights positive relative to negative outcomes would overvalue some options in gambling tasks because gains would effectively loom larger than losses.

Bursts, pauses, and tonic activity
Electrophysiological findings suggest that dopamine neurons encode positive RPEs by phasic bursts of activity (Bayer and Glimcher, 2005) and L-DOPA, by increasing phasic dopamine release, could amplify positive RPEs. However, it is unknown what effect dopaminergic drugs have on the duration of the pauses in dopamine neuron activity that have been correlated with negative RPEs (Bayer et al., 2007), consistent with a computational model of basal ganglia function (Cohen and Frank, 2009). If negative RPEs are, in fact, encoded in pauses and these drugs do not significantly affect pause durations, learning rates for negative outcomes could be similar in Parkinson's patients on and off medication, as we found. Alternatively, dopaminergic drugs might reduce pause durations, decreasing learning rates for negative outcomes, consistent with results of some studies (Frank et al., 2004, 2007b; Cools et al., 2006; Frank and O'Reilly, 2006).

It should also be noted that both Parkinson's disease and dopaminergic drugs, including D_2 receptor agonists taken by the majority ($n = 17$) of our patients, are likely to affect both phasic and tonic dopamine activity. Phasic and tonic dopamine signaling may play distinct roles in reinforcement learning (Niv et al., 2007), and it is possible that tonic dopamine effects also contribute to the differences we observe between subject groups. It is also possible that dopaminergic drugs might have effects outside of the neural pathways implicated in reinforcement learning or might have nondopaminergic effects that could alter choice behavior in the same way as predicted by RL models. Future experiments might address the role of dopamine neuron pauses in reinforcement learning and the relative contributions of phasic and tonic dopamine activity and of nondopaminergic activity to reinforcement learning and decision making.

Dopamine-dependent effects on perseveration
We also found that perseveration in choice, independent of reward history, increased with normal aging and was higher in Parkinson's patients off dopaminergic medication than in elderly subjects. Dopaminergic drugs reversed this effect, reducing perseveration in Parkinson's patients. Several studies in Parkinson's patients have found deficits in switching attention from one stimulus or task to another (Lees and Smith, 1983; Cooper et al., 1991; Owen et al., 1993; Cools et al., 2001; Lewis et al., 2005; Slabosz et al., 2006), consistent with our finding of perseveration in Parkinson's patients. Parkinson's patients may be better able to switch to a new cue when it is novel (Shohamy et al., 2009). In some tasks, dopaminergic medication has been shown to improve set switching (Owen et al., 1993; Cools et al., 2001), consistent with our finding of reduced

perseveration on dopaminergic medication. However, perseveration after reward contingency changes might be accounted for in some tasks by RL models (Suri and Schultz, 1999). In contrast, the perseverative effects we observed cannot be accounted for by existing RL models. This finding emphasizes the importance of considering dopamine effects on decision making that are independent of reward history. This finding may highlight the role of the substantia nigra, which is affected by both normal aging (Stark and Pakkenberg, 2004) and Parkinson's disease (Dauer and Przedborski, 2003), in this non-RL process.

Assessing the robustness of model findings
To examine the robustness of our findings, we used a more conservative linear regression approach to explain choice behavior and found results consistent with our learning model analysis. This approach is often used to study value-based decision making in animals (Sugrue et al., 2004; Corrado et al., 2005; Lau and Glimcher, 2005; Kennerley et al., 2006). We also confirmed that behavioral results were comparable for young adult subjects and young adult monkeys performing a similar task (Lau and Glimcher, 2005) (see supplemental data, available at www.jneurosci.org as supplemental material). If subjects learn values according to standard RL models, reward weights estimated by linear regression, although not constrained to do so, will decay exponentially. Reward weights across our groups bear a striking resemblance to the exponentially weighted average of reward history in RL models. Reward weights for Parkinson's patients decayed significantly more quickly on than off dopaminergic medication, consistent with higher learning rates in patients on than off dopaminergic medication. Finally, we used model comparison methods to show that the RL model explains the data better than the more general linear regression approach, although results using the two different modeling approaches were consistent.

Conclusion
Our results are consistent with the hypothesis that dopamine neurons encode a RPE signal for reinforcement learning. The dopaminergic RPE hypothesis predicts that learning rates estimated from choice behavior will differ in Parkinson's patients on and off dopaminergic medication and that is what we found. More specifically, we found that the increase in learning rates we observed with dopaminergic drugs is selective for learning from positive outcomes, consistent with the hypothesis that dopamine neurons might differentially encode positive and negative RPEs. We found reinforcement learning to remain relatively intact with aging and Parkinson's disease, but that reward-independent perseveration in choice increased with both. Dopaminergic medication reversed this effect. This novel dopamine-dependent effect is not predicted by standard RL models and highlights the importance of considering additional roles for dopamine in decision making that are independent of reward history.

References
Baum WM (1974) On two types of deviation from the matching law: bias and undermatching. J Exp Anal Behav 22:231–242.

Bayer HM, Glimcher PW (2005) Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47:129–141.

Bayer HM, Lau B, Glimcher PW (2007) Statistics of midbrain dopamine neuron spike trains in the awake primate. J Neurophysiol 98:1428–1439.

Bódi N, Kéri S, Nagy H, Moustafa A, Myers CE, Daw N, Dibó G, Takáts A, Bereczki D, Gluck MA (2009) Reward-learning and the novelty-seeking personality: a between- and within-subjects study of the effects of dopamine agonists on young Parkinson's patients. Brain 132:2385–2395.

Burnham KP, Anderson DR (2002) Model selection and multimodel inference. New York: Springer.

Camerer C, Ho TH (1999) Experience-weighted attraction learning in normal form games. Econometrica 67:827–874.

Cohen MX, Frank MJ (2009) Neurocomputational models of basal ganglia function in learning, memory and choice. Behav Brain Res 199:141–156.

Cools R, Barker RA, Sahakian BJ, Robbins TW (2001) Enhanced or impaired cognitive function in Parkinson's disease as a function of dopaminergic medication and task demands. Cereb Cortex 11:1136–1143.

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, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M (2009) Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J Neurosci 29:1538–1543.

Cooper JA, Sagar HJ, Jordan N, Harvey NS, Sullivan EV (1991) Cognitive impairment in early, untreated Parkinson's disease and its relationship to motor disability. Brain 114:2095–2122.

Corrado GS, Sugrue LP, Seung HS, Newsome WT (2005) Linear-nonlinear-Poisson models of primate choice dynamics. J Exp Anal Behav 84:581–617.

Czernecki V, Pillon B, Houeto JL, Pochon JB, Levy R, Dubois B (2002) Motivation, reward, and Parkinson's disease: influence of dopatherapy. Neuropsychologia 40:2257–2267.

Dagher A, Robbins TW (2009) Personality, addiction, dopamine: insights from Parkinson's disease. Neuron 61:502–510.

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

Daw ND, Kakade S, Dayan P (2002) Opponent interactions between serotonin and dopamine. Neural Netw 15:603–616.

Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006) Cortical substrates for exploratory decisions in humans. Nature 441:876–879.

Dayan P, Huys QJM (2009) Serotonin in affective control. Annu Rev Neurosci 32:95–126.

Dodd ML, Klos KJ, Bower JH, Geda YE, Josephs KA, Ahlskog JE (2005) Pathological gambling caused by drugs used to treat Parkinson's disease. Arch Neurol 62:1377–1381.

Frank MJ, O'Reilly RC (2006) A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behav Neurosci 120:497–517.

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

Frank MJ, Moustafa AA, Haughey HM, Curran T, Hutchison KE (2007a) Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc Natl Acad Sci U S A 104:16311–16316.

Frank MJ, Samanta J, Moustafa AA, Sherman SJ (2007b) Hold your horses: impulsivity, deep brain stimulation, and medication in Parkinsonism. Science 318:1309–1312.

Gotham AM, Brown RG, Marsden CD (1988) "Frontal" cognitive function in patients with Parkinson's disease "on" and "off" levodopa. Brain 111:299–321.

Harden DG, Grace AA (1995) Activation of dopamine cell firing by repeated L-DOPA administration to dopamine-depleted rats: its potential role in mediating the therapeutic response to L-DOPA treatment. J Neurosci 15:6157–6166.

Herrnstein RJ (1961) Relative and absolute strength of response as a function of frequency of reinforcement. J Exp Anal Behav 4:267–272.

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

Hollerman JR, Schultz W (1998) Dopamine neurons report an error in the temporal prediction of reward during learning. Nat Neurosci 1:304–309.

Hornykiewicz O (1974) Mechanisms of action of L-DOPA in Parkinson's disease. Life Sci 15:1249–1259.

Keller RW Jr, Kuhr WG, Wightman RM, Zigmond MJ (1988) The effect of L-DOPA on in vivo dopamine release from nigrostriatal bundle neurons. Brain Res 447:191–194.

Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, Rushworth MFS (2006) Optimal decision making and the anterior cingulate cortex. Nat Neurosci 9:940–947.

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

Knowlton BJ, Mangels JA, Squire LR (1996) A neostriatal habit learning system in humans. Science 273:1399–1402.

Lang AE, Fahn S (1989) Assessment of Parkinson's disease. In: Quantification of neurologic deficit (Munsat TL, ed), pp 285–309. Boston: Butterworth-Heinemann.

Lau B, Glimcher PW (2005) Dynamic response-by-response models of matching behavior in rhesus monkeys. J Exp Anal Behav 84:555–579.

Lau B, Glimcher PW (2008) Value representations in the primate striatum during matching behavior. Neuron 58:451–463.

Lees AJ, Smith E (1983) Cognitive deficits in the early stages of Parkinson's disease. Brain 106:257–270.

Lewis SJ, Slabosz A, Robbins TW, Barker RA, Owen AM (2005) Dopaminergic basis for deficits in working memory but not attentional set-shifting in Parkinson's disease. Neuropsychologia 43:823–832.

Molina JA, Sáinz-Artiga MJ, Fraile A, Jiménez-Jiménez FJ, Villanueva C, Ortí-Pareja M, Bermejo F (2000) Pathologic gambling in Parkinson's disease: a behavioral manifestation of pharmacologic treatment? Mov Disord 15:869–872.

Nakahara H, Itoh H, Kawagoe R, Takikawa Y, Hikosaka O (2004) Dopamine neurons can represent context-dependent prediction error. Neuron 41:269–280.

Niv Y, Montague PR (2009) Theoretical and empirical studies of learning. In: Neuroeconomics: decision making and the brain (Glimcher PW, Camerer CF, Fehr E, Poldrack RA, eds), pp 331–351. New York: Academic.

Niv Y, Daw ND, Joel D, Dayan P (2007) Tonic dopamine: opportunity costs and the control of response vigor. Psychopharmacology (Berl) 191:507–520.

Owen AM, Roberts AC, Hodges JR, Summers BA, Polkey CE, Robbins TW (1993) Contrasting mechanisms of impaired attentional set-shifting in patients with frontal lobe damage or Parkinson's disease. Brain 116:1159–1175.

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

Platt ML, Glimcher PW (1999) Neural correlates of decision variables in parietal cortex. Nature 400:233–238.

Rescorla RA, Wagner AR (1972) A theory of pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. In: Classical conditioning II: current research and theory (Black AH, Prokasy WF, eds), pp 64–99. New York: Appleton.

Samejima K, Ueda Y, Doya K, Kimura M (2005) Representation of action-specific reward values in the striatum. Science 310:1337–1340.

Schönberg T, Daw ND, Joel D, O'Doherty JP (2007) Reinforcement learning signals in the human striatum distinguish nonlearners from learners during reward-based decision making. J Neurosci 27:12860–12867.

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

Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464.

Serences JT (2008) Value-based modulations in human visual cortex. Neuron 60:1169–1181.

Shohamy D, Myers CE, Grossman S, Sage J, Gluck MA, Poldrack RA (2004) Cortico-striatal contributions to feedback-based learning: converging data from neuroimaging and neuropsychology. Brain 127:851–859.

Shohamy D, Myers CE, Geghman KD, Sage J, Gluck MA (2006) L-DOPA impairs learning, but spares generalization, in Parkinson's disease. Neuropsychologia 44:774–784.

Shohamy D, Myers CE, Hopkins RO, Sage J, Gluck MA (2009) Distinct hippocampal and basal ganglia contributions to probabilistic learning and reversal. J Cogn Neurosci 21:1820–1832.

Slabosz A, Lewis SJ, Smigasiewicz K, Szymura B, Barker RA, Owen AM (2006) The role of learned irrelevance in attentional set-shifting impairments in Parkinson's disease. Neuropsychology 20:578–588.

Stark AK, Pakkenberg B (2004) Histological changes of the dopaminergic nigrostriatal system in aging. Cell Tissue Res 318:81–92.

Sugrue LP, Corrado GS, Newsome WT (2004) Matching behavior and the representation of value in the parietal cortex. Science 304:1782–1787.

Suri RE, Schultz W (1999) A neural network learns a spatial delayed response task with a dopamine-like reinforcement signal. Neurosci 91:871–890.

Sutton RS, Barto AG (1998) Reinforcement learning: an introduction. Cambridge, MA: MIT.

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.

Voon V, Hassan K, Zurowski M, Duff-Canning S, de Souza M, Fox S, Lang AE, Miyasaki J (2006) Prospective prevalence of pathologic gambling and medication association in Parkinson disease. Neurology 66:1750–1752.

Wightman RM, Amatore C, Engstrom RC, Hale PD, Kristensen EW, Kuhr WG, May LJ (1988) Real-time characterization of dopamine overflow and uptake in the rat striatum. Neuroscience 25:513–523.