Dorsal Striatal–Midbrain Connectivity in Humans Predicts How Reinforcements Are Used to Guide Decisions

Thorsten Kahnt[1,2*], Soyoung Q Park[1*], Michael X Cohen[3], Anne Beck[1], Andreas Heinz[1], and Jana Wrase[1]

Abstract

■ It has been suggested that the target areas of dopaminergic midbrain neurons, the dorsal (DS) and ventral striatum (VS), are differently involved in reinforcement learning especially as actor and critic. Whereas the critic learns to predict rewards, the actor maintains action values to guide future decisions. The different midbrain connections to the DS and the VS seem to play a critical role in this functional distinction. Here, subjects performed a dynamic, reward-based decision-making task during fMRI acquisition. A computational model of reinforcement learning was used to estimate the different effects of positive and negative reinforcements on future decisions for each subject individually. We found that activity in both the DS and the VS correlated with reward prediction errors. Using functional connectivity, we show that the DS and the VS are differentially connected to different midbrain regions (possibly corresponding to the substantia nigra [SN] and the ventral tegmental area [VTA], respectively). However, only functional connectivity between the DS and the putative SN predicted the impact of different reinforcement types on future behavior. These results suggest that connections between the putative SN and the DS are critical for modulating action values in the DS according to both positive and negative reinforcements to guide future decision making. ■

INTRODUCTION

Learning which action to take in an uncertain environment to maximize reward and minimize punishment is critical for survival. Both positive and negative outcomes of current decisions can contribute differentially to the way individuals decide in the future. The dopaminergic midbrain system and its prominent target areas, especially the striatum, play key roles in this process. Midbrain dopamine neurons are thought to contribute to reinforcement learning by sending a teaching signal to the striatum that biases action selection according to the previous action–outcome history (Samejima & Doya, 2007; Samejima, Ueda, Doya, & Kimura, 2005; Schultz & Dickinson, 2000; Hollerman & Schultz, 1998). Specifically, bursts in dopamine after positive outcomes are thought to facilitate the current response, whereas dips in dopamine after negative outcomes may support the inhibition of the current response (Schultz, 2002; Hollerman & Schultz, 1998). The ventral and dorsal compartments of the striatum have been implicated in maintaining and updating reward predictions and action values, respectively. Whereas the ventral part (ventral striatum: VS) learns to predict future rewards, the dorsal compartment (dorsal striatum: DS) maintains information about the outcomes of decisions to enable that the better option is chosen more often (Atallah, Lopez-Paniagua, Rudy, & O'Reilly, 2007; Schonberg, Daw, Joel, & O'Doherty, 2007; Williams & Eskandar, 2006; Samejima et al., 2005; O'Doherty et al., 2004; Joel, Niv, & Ruppin, 2002; Joel & Weiner, 2000). In computational reinforcement learning theory, the critic uses a prediction error signal to update predictions about future rewards, whereas the actor uses the same signal to update action values that biases future decisions toward advantageous options (O'Doherty et al., 2004; Sutton & Barto, 1998).

As neurophysiological investigations in primates and rats have shown, the DS and the VS receive dopaminergic input from different but somewhat overlapping midbrain regions building an ascending midbrain–striatal loop (Ikemoto, 2007; Haber, Fudge, & McFarland, 2000). Specifically, the ventral tegmental area (VTA) is reciprocally interconnected with the VS, whereas the substantia nigra (SN) receives input from the VS and is reciprocally connected to the DS (Haber, 2003; Haber et al., 2000). The DS and the VS are also differentially connected to the frontal cortex. The VS is connected with

[1]Charité—Universitätsmedizin Berlin (Charité Campus Mitte), Germany, [2]Bernstein Center for Computational Neuroscience Berlin, Germany, [3]University of Arizona, Tucson
*These authors contributed equally to the work.

medial prefrontal and orbito-frontal cortices, whereas the DS is connected with dorsal prefrontal and motor cortices (Lehericy et al., 2004; Alexander, Crutcher, & DeLong, 1990). Due to its connections, associations between sensory cues and motor behavior that lead to reward might be strengthened in the striatum by dopaminergic projections from the midbrain (Williams & Eskandar, 2006; Reynolds, Hyland, & Wickens, 2001).

In situations with two different options, one can learn to choose *X* over *Y* by either learning that *X* leads to positive feedback or that *Y* leads to negative feedback, or both. In such situations, it is challenging to disentangle the degree to which positive and negative reinforcements contribute to learning. However, it is possible to differentiate both processes and subjects differ in how they use positive and negative outcomes to guide their decisions (Frank, Woroch, & Curran, 2005; Frank, Seeberger, & O'Reilly, 2004). These individual differences in learning from either positive or negative reinforcements have been linked to dopamine genetics, dopamine treatment and neurological conditions in which dopamine activity is changed (Cohen, Krohn-Grimberghe, Elger, & Weber, 2007; Frank, Moustafa, Haughey, Curran, & Hutchison, 2007; Frank, Scheres, & Sherman, 2007; Klein et al., 2007; Frank, 2005). However, in humans, although there is an increasing amount of research about the striatum and its involvement in reward-learning (Cohen, 2007; Delgado, 2007; Pessiglione, Seymour, Flandin, Dolan, & Frith, 2006; Cohen & Ranganath, 2005; O'Doherty, Dayan, Friston, Critchley, & Dolan, 2003), so far, little is known about the specific contributions of midbrain–striatal connections to the way reinforcements are used to guide decision making. A recent study showed prediction error-related responses in the VS and the VTA, and provided evidence for a functional coupling between both regions (D'Ardenne, McClure, Nystrom, & Cohen, 2008). Here, we went beyond this study by showing how these striatal–midbrain circuits might be involved in mediating dynamic adjustments in decision making.

Using fMRI and a computational model of reinforcement learning, we aimed to disentangle the use of positive and negative reinforcements to guide decision making in a two-arm bandit task with dynamically changing reward allocations. Furthermore, by means of functional connectivity analyses, we investigated specific contributions of different striatal–midbrain connections to how different reinforcements impact future decision making.

Because both the DS and the VS have been suggested to play specific roles in reinforcement learning by learning the action and reward value, respectively (Williams & Eskandar, 2006; O'Doherty et al., 2004; Joel et al., 2002), we hypothesized that information about reinforcements is represented in form of reward prediction errors in the VS and the DS. Additionally, according to animal studies (Haber, 2003), we hypothesized that the VS and the DS are differentially connected to different midbrain regions. Finally, we hypothesized that the integrity of midbrain–striatal connectivity plays a critical role in the use of reinforcements to guide future decision making (Belin & Everitt, 2008; Faure, Haberland, Conde, & El Massioui, 2005). Specifically, if reinforcement-related activity in the midbrain is used to update action values in the DS, the way different subjects use this information should depend on the functional connectivity between the midbrain and the DS, but not necessarily the VS.

METHODS

Participants

Nineteen right-handed subjects (10 women, aged 24–31 years, mean = 25.7 ± 0.45 years) participated in the study. Subjects had normal or corrected-to-normal vision, reported having no psychiatric conditions, and gave written consent to participate. The study was approved by the local ethics review committee of the Charité—Universitätsmedizin Berlin.

Experimental Design

During fMRI acquisition, subjects performed a reward-based decision-making task with dynamically changing action–reward contingencies (Figure 1A). In each of 200 trials (100 per session), subjects first saw two abstract targets on the screen and were asked to choose one of them as quickly as possible by pressing the left or right button with the left or right thumb on a response box (maximum decision time: 2 sec). A blue box surrounding their chosen target and feedback (green smiling face for positive feedback or a red frowning face for negative feedback) was simultaneously shown for 1 sec. The trials were separated with a jittered interval of 1 to 6.5 sec. In each session two abstract targets were randomly selected for each subject and displayed one on the left and the other one on the right side of the screen and remained there over the whole session. In each trial, one side leads to a positive feedback and the other to a negative feedback (no neutral feedback; only positive or negative). The task provided three different allocations of reward probability for left versus right responses (i.e., rule) that changed unpredictably for the subject during the experiment: 20:80, 50:50, and 80:20 (left/right) probability of reward. The rule reversed after a minimum of 10 trials and after the subject had reached at least 70% accuracy. If the rule was not learned after 16 trials, the task went over to the next condition automatically. Before entering the scanner, subjects performed a practice version of the task (without reversal component) to be introduced to the probabilistic character of the task. Subjects were instructed to win as often as possible.

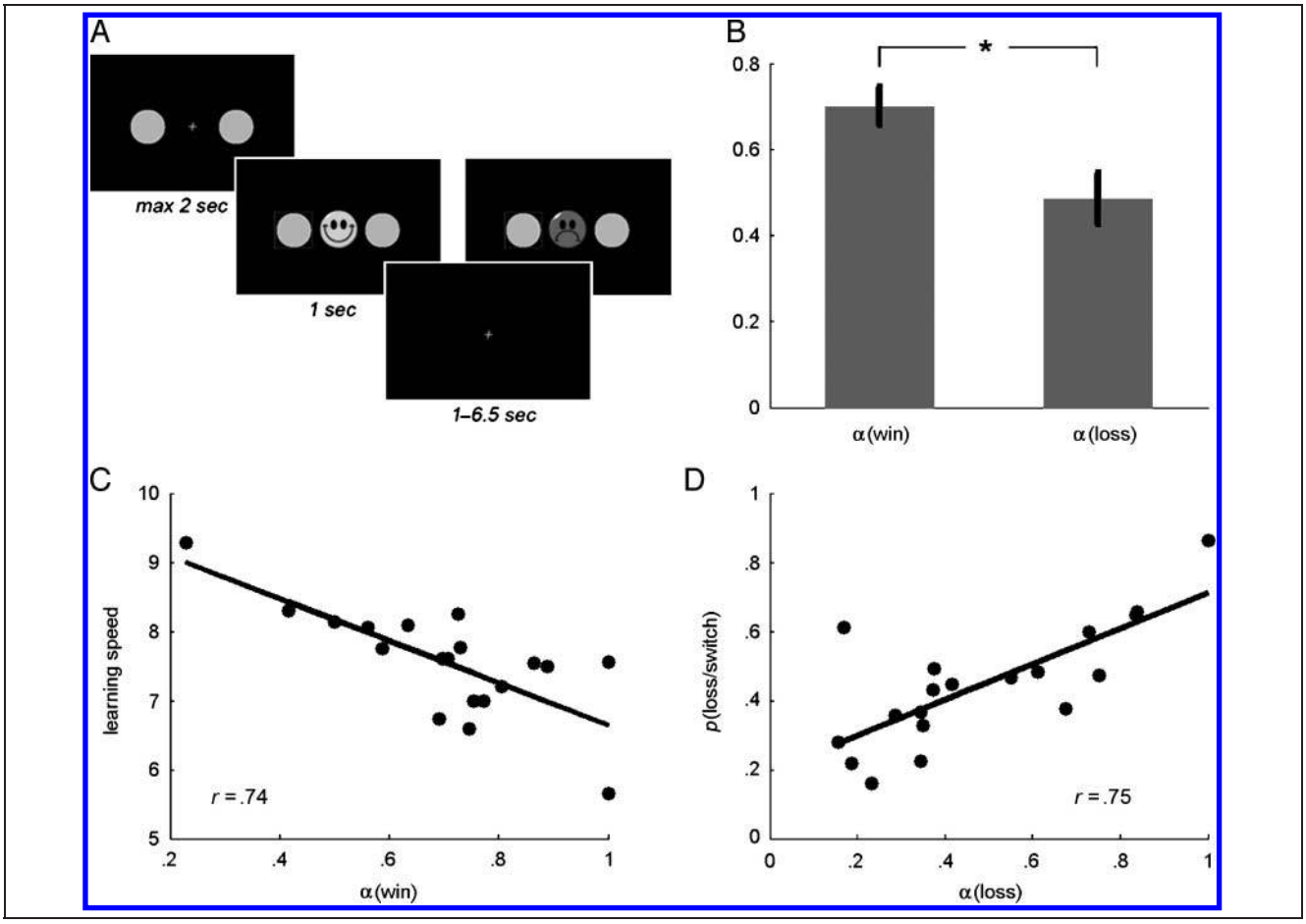

Figure 1. Experimental design and behavioral results. (A) Structure of the dynamic decision-making task. Subjects first saw two targets for up to 2 sec (or reaction time). After selecting one with a button press, a blue frame surrounded the chosen target and either positive (reward) or negative (loss) feedback (reward or loss, left or right middle panel, respectively) was shown for 1 to 6.5 sec. (B) Average of individual learning rates α(win) and α(loss). Asterisk indicates significant difference at the $p < .001$ level and error bars indicate standard error of mean. (C) Scatterplot depicts relationship between α(win) (*x*-axis) and learning speed (*y*-axis, average number of trials until the rule was learned). (D) Scatterplot depicts relationship between α(loss) (*x*-axis) and *p*(loss/switch) (*y*-axis, proportion of loss trials after which subjects switched to the opposite target to all loss trials). Solid black lines represent best fitting regression lines.

Reinforcement Learning Model

Blood oxygenation level-dependent (BOLD) data and behavioral responses were analyzed using a standard reinforcement learning model (Sutton & Barto, 1998). Similar models have been used previously to analyze behavioral and neural data (Cohen, 2007; Cohen & Ranganath, 2007; Haruno & Kawato, 2006; Pessiglione et al., 2006; Cohen & Ranganath, 2005; Samejima et al., 2005).

The model used a reward prediction error (δ) to update action values or decision weights (w) associated with each response (left and right) (Schultz, 2004; Holroyd & Coles, 2002; Egelman, Person, & Montague, 1998; Schultz, Dayan, & Montague, 1997). Thus, after receiving a positive feedback, the model generates a positive prediction error which is used to increase the size of the action value of the chosen option (e.g., the right-hand target). In contrast, after receiving negative feedback, the model generates a negative prediction error, which is used to decrease the size of the action value of the chosen option, making the model less likely to choose that decision option on the following trial. Specifically, the model uses the soft-max mechanism to generate the probability (p) of choosing the right-hand target on trial t as the logit transform of the difference in the action values in each trial (w_t) associated with each target, passed through a biasing sigmoid function (Montague, Hyman, & Cohen, 2004; Egelman et al., 1998).

$$p(\text{right})_t = \frac{e^{w(\text{right})_t}}{e^{w(\text{right})_t} + e^{w(\text{left})_t}}$$

After each trial, a prediction error (δ) is calculated as the difference between the outcome received (r = 0

or 1 for losses and wins) and the action value for the chosen target:

$$\delta_t = r_t - w(\text{chosen})_t$$

for example, $\delta = 1 - w(\text{right})_t$ in case of a positive outcome after choosing the right-hand target. The action values are then updated according to:

$$w_{t+1} = w_t + \pi \times \alpha(\text{outcome}) \times \delta_t$$

where π is 1 for the chosen and 0 for the unchosen target, α(outcome) is a set of learning rates for positive (α(win)) and negative outcomes (α(loss)), which scale the effect of the prediction error on future action values, with a high learning rate indicating a high impact of that type of reinforcement on future behavior. Given that information about rewards and punishments are differently used by the brain to guide future behavior, these parameters should predict different aspects of subjects' behavior and brain processes. Learning rates were individually estimated by fitting the model predictions (p(right)) to subjects' actual behavior. We used the multivariate constrained minimization function (fmincon) implemented in MATLAB 6.5 for this fitting procedure. Initial values for learning rates were α(win) = α(loss) = 0.5 and for action values, w(left) = w(right) = 0.5.

Behavioral Analyses

Two dependent variables of behavioral performance were used: (1) learning speed and (2) p(loss/switch). Learning speed was defined as the average number of trials until the rule in 20:80 and 80:20 conditions was learned (80% correct responses over a sliding window of 5 trials). p(loss/switch) was defined as the proportion of loss trials after which subjects switched to the opposite target, to the total number of loss trials. Thus, learning speed is an indicator of maintaining the rewarded action even in the face of probabilistic losses, whereas p(loss/switch) is an indicator of avoiding the unrewarded option.

To test whether the individually estimated learning rates α(win) and α(loss) predict different aspects of subjects' behavior, both learning rates were simultaneously regressed against p(loss/switch) and learning speed, respectively, using multiple regression.

In order to examine the correspondence between model predictions and subjects' behavior, model predictions were compared with the actual behavior on a trial-by-trial basis. To do this, we gave the model (provided with individual learning rates) the unique history of choices and reinforcements of each subject to receive a vector of models' probability of choosing the right-hand target (p(right)) for each trial. Subjects' trial-wise behavior was coded as zeros and ones for left- and right-hand responses, respectively. Model predictions were then regressed against the vector of subjects' choices and individual b-coefficients were taken to a second-level random effect analysis using a one-sample t test.

MRI Acquisition and Preprocessing

Functional imaging was conducted using a 3.0-Tesla GE Signa scanner with an eight-channel head coil to acquire gradient-echo, T2*-weighted echo-planar images. For each of the two sessions, 310 volumes (~12 min) containing 29 slices (4 mm thick) were acquired. The imaging parameter were as follows: repetition time (TR) = 2.3 sec, echo time (TE) = 27 msec, $\alpha = 90^\circ$, matrix size = 96 × 96, and a field of view (FOV) of 260 mm, thus yielding an in-plane voxel resolution of 2.71 mm^2. We were unable to acquire data from the ventromedial part of the orbito-frontal cortex (OFC) due to susceptibility artifacts at air–tissue interfaces. A T1-weighted structural dataset was collected for the purpose of anatomical localization. The parameters were as follows: TE = 3.2 msec, matrix size = 196 × 196, FOV = 240 mm, 1 mm slice thickness, $\alpha = 20^\circ$. Due to technical problems, we were unable to acquire structural scans from two subjects.

Functional data were analyzed using SPM5 (Wellcome Department of Imaging Neuroscience, Institute of Neurology, London, UK). The first three volumes of each session were discarded to allow for magnetic saturation effects. For preprocessing, images were slice time corrected, realigned and unwrapped, spatially normalized to a standard T2* template of the Montreal Neurological Institute (MNI), resampled to 2.5 mm isotropic voxels, and spatially smoothed with an 8-mm FWHM kernel.

fMRI Data Analyses

To investigate the neural responses to feedback valence, independent of learning conditions, we set up a general linear model (GLM) with the onsets of each feedback type as regressors. Three feedback types were included: (1) reward outcomes, (2) loss outcomes followed by a switch (loss/switch), and (3) loss outcomes not followed by a switch (loss/stay) in behavior. Reward trials could not be broken down into similar subgroups because of too few reward/switch trials. The stick functions were convolved with a hemodynamic response function (HRF) provided by SPM5 to account for the sluggishness of the BOLD signal. The regressors were simultaneously regressed against the BOLD signal in each voxel using the least squares criteria, and contrast images were computed from the resulting parameter estimates.

To examine neural responses that correlate with ongoing reward prediction errors during reinforcement guided decision-making, we set up a second GLM with a parametric design (Buchel, Holmes, Rees, & Friston, 1998). In this model, the stimulus functions for reward and loss feedback were parametrically modulated by the trial-wise prediction errors derived from the reinforcement learning model using individually estimated learning rates. The modulated stick functions were then convolved with an HRF to provide the regressors used in the GLM. These regressors were then orthogonalized with respect to the onset regressors of reward and loss trials and regressed against the BOLD signal. Individual contrast images were computed for prediction error-related responses and taken to a second-level random effect analysis using one-sample t test. Thresholds were set to $p < .001$, uncorrected with an extend threshold of 15 continuous voxels. Because reward prediction errors are thought to act as a teaching signal, this analysis should reveal regions involved in updating action values.

To investigate the interplay between striatal subregions and midbrain during reinforcement-guided decision making, functional connectivity of the DS and the VS was assessed using the ''psychophysiological interaction'' term (Cohen, Elger, & Weber, 2008; Cohen, Heller, & Ranganath, 2005; Pessoa, Gutierrez, Bandettini, & Ungerleider, 2002; Friston et al., 1997). Two psychophysiological interaction models were set up to assess functional connectivity of the (1) DS and (2) VS independently. Clusters in the DS and VS, in which activity correlated significantly ($p < .001$, $k = 15$) with reward prediction errors, were used as seed regions of interest (ROIs). The method used here relies on correlations in the observed BOLD time-series data and makes no assumptions about the nature of the neural event that contributed to the BOLD signal (Cohen et al., 2008). For each model, the entire time series over the experiment was extracted from each subject in the clusters of the left and right (dorsal or ventral) striatum, respectively. Regressors were then created by multiplying the normalized time series of the left or right striatum with condition vectors that contain ones for six TRs after each Right versus Left hand × Reward versus Loss feedback, respectively, and zeros otherwise. Thus, the four condition vectors for Right versus Left hand × Reward versus Loss feedbacks were each multiplied with the time course of the contralateral striatum. These regressors were then used as covariates in a separate regression. The time series between the left and right hemispheres within each striatal subregion were highly correlated (averages across runs and subjects were $r = .82$ and $r = .70$ in the DS and the VS, respectively). Therefore, after estimation, parameter estimates of left- and right-hand regressors were collapsed, and thus, represent the extent to which feedback-related activity in each voxel correlates with feedback-related activity in the DS and the VS, respectively. In other words, connectivity estimates represent the extent to which activity in the VS and the DS, respectively, contribute to the responsiveness of distinct other regions to reward or loss. Individual contrast images for reward > loss feedback were then computed for both models and entered into second-level one-sample t tests. To identify significant functional connectivity, we used a more stringent threshold of $p < .05$, family-wise error (FWE) corrected for whole brain, with a cluster threshold of $k = 10$ voxels.

In order to confirm the statistical significance of the finding from the whole-brain analyses of different patterns of functional connectivity between the DS and the VS on the one hand, and different midbrain regions on the other, a three-way ANOVA on connectivity estimates in midbrain ROIs was performed. For this, ROIs within the anatomical boundaries of the midbrain were defined as follows: left and right dorsal/posterior (d/p) midbrain ROIs were defined from significant clusters in the reward > loss contrast of the DS seed model. On the other hand, left and right ventral/anterior (v/a) midbrain ROIs were defined from significant clusters in the reward > loss contrast of the VS seed model. Functional connectivity parameter estimates from both models (DS and VS seed) were then extracted from the reward > loss contrast in these four midbrain ROIs and entered into a 2 × 2 × 2 (VS vs. DS seed × v/a vs. d/p midbrain ROI × Left vs. right side) repeated measures ANOVA. We hypothesized that the DS and the VS were differentially connected to d/p and v/a midbrain regions. Thus, the critical effect in the ANOVA is the Striatal-seed by Midbrain-ROI interaction.

To test whether the impact of positive and negative reinforcements on subsequent decision making depends on the integrity of functional DS–d/p midbrain connectivity, individual functional connectivity parameter estimates during reward and loss feedback were correlated with individual estimates of α(win) and α(loss), respectively. Because α(win) and α(loss) represent the individual degree of learning from either reinforcement, we predicted that α(win) correlates positively with DS–d/p midbrain connectivity during reward trials and that α(loss) correlates positively with DS–d/p midbrain connectivity during loss trials but not vice versa [i.e., α(win) with DS–d/p during loss and α(loss) with DS–d/p during win). Due to these directed hypotheses, one-tailed tests of significance were used.

Anatomical localizations were carried out by overlaying statistical maps on a normalized structural T1-weighted image averaged across subjects and with reference to an anatomical atlas (Duvernoy, 1999). Additionally, MNI coordinates were transformed in the Talairach space and corresponding areas were identified with reference to the atlas provided by Talairach and Tournoux (1988). Precise anatomical localization of midbrain structures is difficult in fMRI, even at high

spatial resolution. As mentioned in the Discussion section, our results are consistent with anatomical sources in the VTA and the SN; however, we refer to these activations as ''dorsal/posterior (d/p)'' (possibly corresponding to the SN) and ''ventral/anterior (v/a)'' (possibly corresponding to the VTA). These activations are also consistent with others who have reported fMRI activations in midbrain structures (D'Ardenne et al., 2008; Wittmann, Schiltz, Boehler, & Duzel, 2008; Adcock, Thangavel, Whitfield-Gabrieli, Knutson, & Gabrieli, 2006; Bunzeck & Duzel, 2006; Menon & Levitin, 2005; Wittmann et al., 2005; Aron et al., 2004).

RESULTS

Behavioral Results

On average, reaction times were 561 msec (±23) and subjects won in 71% (±0.5) of trials. Subjects needed 7.56 trials (±0.18) on average to learn the rule and switched after 44.6% (±4.0) of all loss trials to the opposite target.

Average estimated learning rates were 0.70 (±0.04) and 0.49 (±0.06) for α(win) and α(loss), respectively, and differed significantly [$t(18) = 4.38$, $p < .001$], indicating that positive and negative reinforcements had different effects on subsequent decision making (Figure 1B). Besides that, α(win) and α(loss) were correlated to some degree ($r = .58$, $p < .01$), indicating that although rewards and punishments contributed differently to reinforcement guided decision making, there seems to be a tendency to learn from experience per se that varied between subjects.

To assess whether different learning rates captured different aspects of behavior, α(win) and α(loss) were simultaneously regressed against the two dependent variables of this study [learning speed and p(loss/switch)]. A multiple regression of both learning rates on learning speed fitted significantly [$R^2 = .56$, $F(2, 16) = 10.36$, $p < .001$] but only α(win) contributed significantly to the regression [$b_{\alpha(win)} = -2.69$, $t(18) = -3.25$, $p < .005$; Figure 1C], whereas α(loss) did not [$b_{\alpha(loss)} = -0.46$, $t(18) = -0.65$, $p = .46$]. In contrast, in the regression against p(loss/switch) [$R^2 = .57$, $F(2, 16) = 10.45$, $p < .001$], α(loss) [$b_{\alpha(loss)} = 0.49$, $t(18) = 3.54$, $p < .005$; Figure 1D], but not α(win) [$b_{\alpha(win)} = 0.06$, $t(18) = 0.33$, $p = .75$] contributed significantly. This double dissociation indicates that both learning rates captured different behavioral aspects of reinforcement-guided decision making, and thus, their validity.

The reinforcement learning model with individual learning rates predicted subjects' behavior quite well; the average b-coefficient was significantly above zero [$b = 3.82 \pm 1.57$, $t(18) = 10.62$, $p < .001$]. Indeed, this regression coefficient was statistically significant in each subject (all ts > 10.58). However, there was still a large variability in the model fits (variance of $b = 2.47$), that is, across subjects the model predicted behavior with different accuracy.

fMRI Results

A GLM with the onsets of each feedback type (reward, loss/switch, loss/stay) as regressors revealed significant activity in the left amygdala/hippocampus, the medial prefrontal cortex (mPFC, BA 10; Figure 2A), the precuneus/posterior cingulate cortex (BA 30), and the inferior parietal cortex (BA 40) in the reward > loss contrast (Table S1). The opposite contrast, loss > reward, revealed enhanced activity in the anterior cingulate cortex (ACC, BA 32; Figure 2B), extending to premotor areas (BA 8/6), and the lateral OFC (BA 47), extending to the insula cortex (BA 13, Table S1). Enhanced activity in the loss/switch > loss/stay contrast was revealed in the left and right lateral OFC (BA 47; Figure 2C) as well as in the dorsal ACC (BA 32), bilateral parietal areas, and pre- and postcentral gyrus (Table S1). No region showed significant activation in loss/stay > loss/switch trials at the $p < .001$ level.

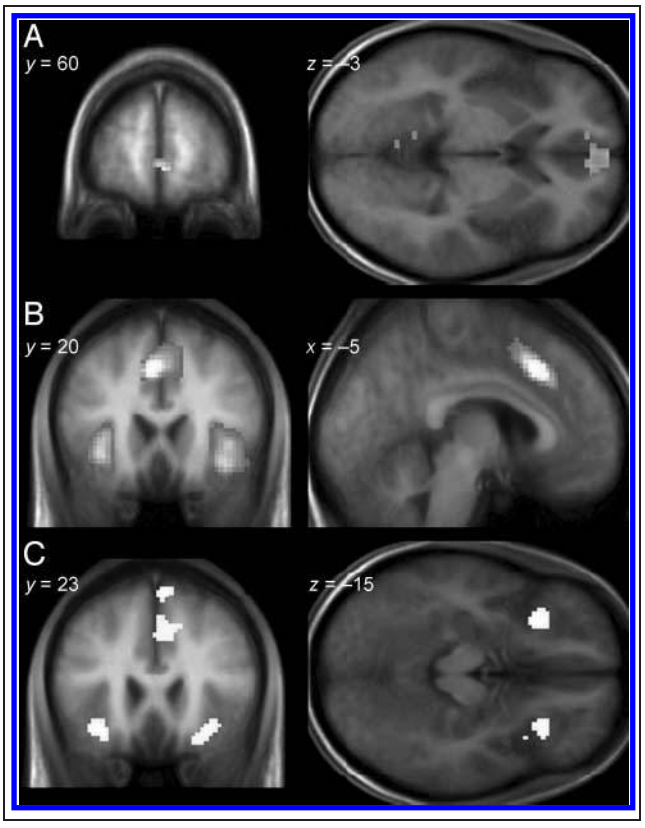

Figure 2. BOLD responses to feedback. (A) The mPFC showing enhanced activity to reward > loss outcomes. (B) The ACC and the insula showing enhanced activity to loss > reward trials. (C) The left and right lateral OFC and the dorsal ACC showing enhanced activity during loss/switch compared to loss/stay trials. Statistical maps are thresholded at $p < .001$, uncorrected ($k = 15$) and overlaid on a normalized structural image averaged across subjects.

Model-based fMRI Results

Trial-wise reward prediction errors correlated with BOLD signal in bilateral DS and bilateral VS in four separable clusters (Figure 3A) as well as in the ACC (BA 32) extending to the premotor cortex (BA 8), the dorsolateral prefrontal cortex, the OFC, and the parietal cortex (Table S2). Activity in the VS was localized at the ventral intersection between the putamen and the caudate nucleus [MNI x, y, z coordinates, left: -10 8 -5, $t(18) = 5.51$; right: 15 5 -5, $t(18) = 4.28$], whereas in the DS, activity peaked in the dorsal anterior caudate nucleus [left: -10 5 10, $t(18) = 4.33$; right: 15 8 15, $t(18) = 4.62$]. Figure 3B depicts mean parameter estimates of the correlation between BOLD responses and model-generated prediction errors separately for positive (win trials) and negative prediction errors (loss trials) in all four activity clusters. In the VS and the DS, BOLD responses were positively correlated with both positive and negative prediction errors; the higher the positive prediction error, the higher the BOLD response, whereas the lower (more negative) the negative prediction error, the lower (more negative) the BOLD response. Parameter estimates of positive prediction errors were significantly higher compared to that of negative prediction errors in the VS [$t(18) = 2.22$, $p < .05$ and $t(18) = 2.36$, $p < .05$, for left and right VS, respectively], but not in the DS ($p = .31$ and $p = .19$, for left and right DS, respectively).

To determine whether the fit of the model to the neural data was related to the model fit of the behavioral data (Cohen, 2007), mean parameter estimates from left and right, DS and VS clusters were correlated with individual b-coefficients of behavioral fit. As shown in Figure 3C, the better the model predicted the behavior, the better the model fit to the neural data (left VS: $r = .74$, right VS: $r = .61$, left DS: $r = .64$, right DS: $r = .77$; all ps $< .005$). This further demonstrates that these striatal prediction error responses are related to reinforcement learning.

Functional Connectivity Results

We found significantly ($p < .05$, FWE corrected, $k = 10$) enhanced connectivity during reward compared to loss trials between the VS seed and bilateral v/a midbrain regions (Figure 4A), the right hippocampus, the pons, and the cerebellum (Table 1). In contrast, the bilateral d/p midbrain regions (Figure 4B), the right hippocampus, the thalamus, and the cerebellum showed significantly enhanced connectivity with the DS in reward > loss (Table 1). Detailed maps of both midbrain regions are shown in Figure S1. The distances between the peak voxels in v/a and d/p midbrain are 15.1 mm and 9.9 mm in the left and right hemispheres, respectively.

A Striatal-seed (DS vs. VS) × Midbrain-ROI (d/p vs. v/a) × Side (left vs. right) ANOVA on connectivity parameter estimates during reward > loss revealed a significant main effect for midbrain ROI [$F(1, 18) = 13.10$, $p < .01$], indicating enhanced connectivity of DS and VS with d/p compared to v/a midbrain but neither a main effect of striatal-seed region [$F(1, 18) = 0.78$, $p = .39$] nor side [$F(1, 18) = 1.05$, $p = .32$]. Additionally, we found a significant Striatal-seed × Midbrain-ROI × Side interaction [$F(1, 18) = 5.24$, $p < .05$] but neither a Striatal-seed × Side [$F(1, 18) = 1.92$, $p = .18$] nor a Midbrain-ROI × Side interaction [$F(1, 18) = 2.11$, $p = .16$]. The critical Striatal-seed × Midbrain-ROI interaction was significant [$F(1, 18) = 17.25$, $p < .001$],

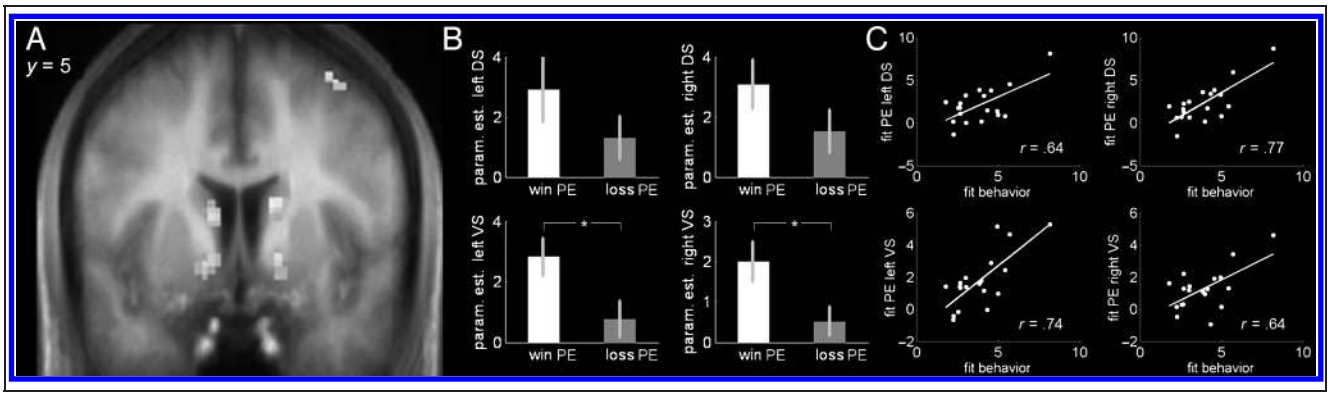

Figure 3. Prediction error responses. (A) Regions in the ventral striatum (VS) and the dorsal striatum (DS) in which BOLD signal was significantly correlated with reward prediction errors. Statistical map is thresholded at $p < .001$, uncorrected, $k = 15$, and overlaid on a normalized structural image averaged across subjects. (B) Parameter estimates of the correlation between model-generated prediction errors and BOLD response in the left DS, right DS, left VS, and right VS (from left to right, top to bottom) for win (white bar) and loss (gray bar) prediction errors separately. Asterisks indicate significant differences at the $p < .05$ level and error bars indicate standard error of mean. (C) Scatterplots depict the relationship between the fit of the model to the behavioral data (x-axis, b-coefficients) and the fit of the model to the BOLD signal in the left DS and right DS, in the left VS and right VS (y-axis, parameter estimates), respectively. Solid lines represent best fitting regression lines.

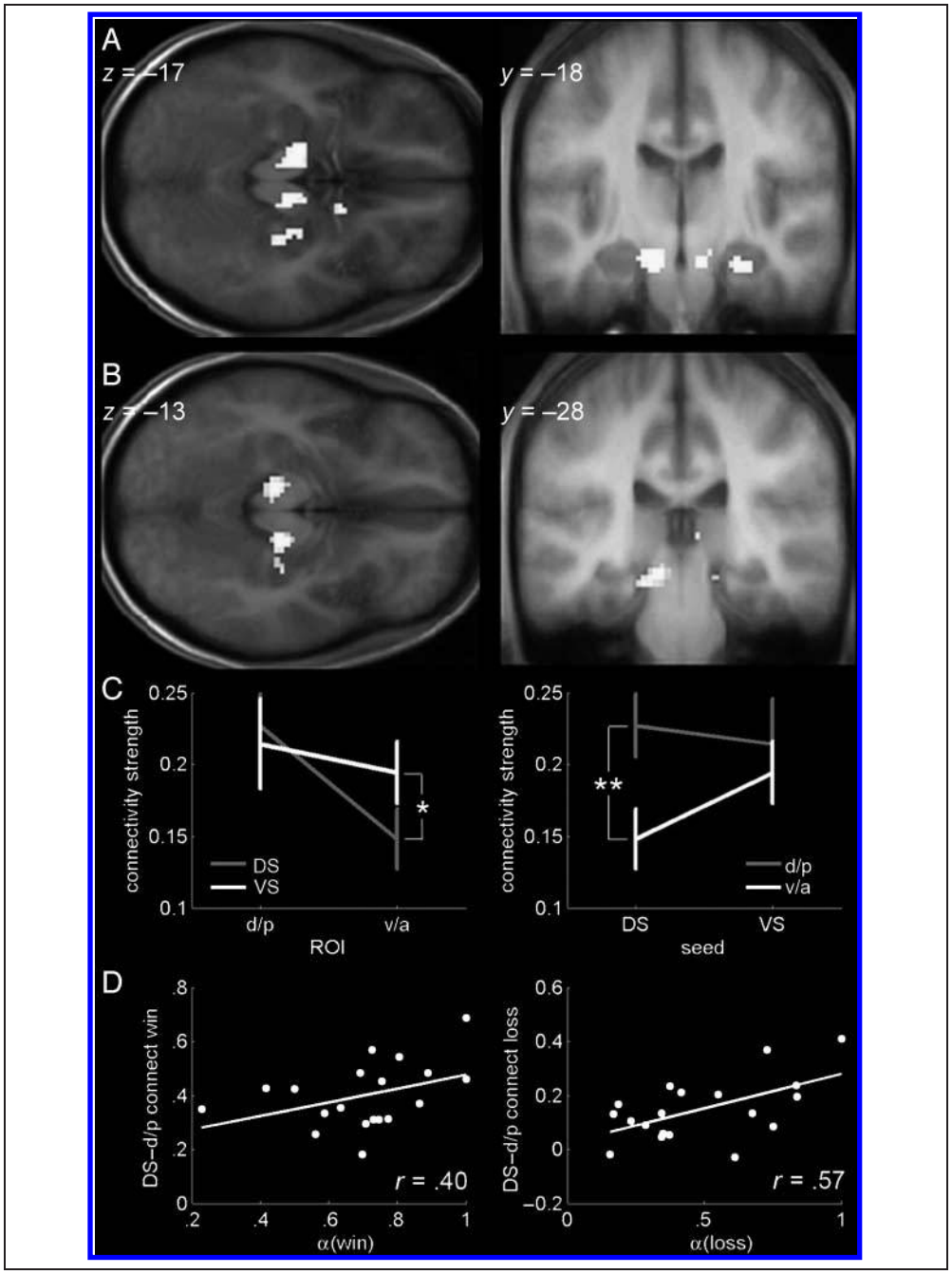

Figure 4. Functional connectivity with the ventral striatum (VS) and the dorsal striatum (DS). Midbrain regions that show enhanced functional connectivity with (A) the VS and (B) the DS during reward compared to losses. Statistical maps are thresholded at $p < .05$, FWE corrected for multiple comparisons (whole brain), $k = 10$, and overlaid on a normalized structural image averaged across subjects. (C) Interaction plots of functional connectivity estimates (reward > loss) between the DS and VS seeds and the dorsal/posterior (d/p) and ventral/anterior (v/a) midbrain ROIs. Left: Connectivity estimates of DS (gray line) and VS (white line) seed plotted as a function of midbrain ROI (d/p vs. v/a midbrain). Right: Connectivity estimates in d/p (gray line) and v/a midbrain (white line) ROI plotted as a function of striatal seed (DS vs. VS). Error bars indicate standard error of mean. Asterisks indicate significant differences at the $*p < .05$ and $**p < .001$ level. (D) Scatterplots depict the significant relationship between α(win) (left) and α(loss) (right) on the one hand (*x*-axes) and functional DS–left d/p midbrain connectivity in win (left panel) and loss trials (right panel) on the other hand (*y*-axes). Solid lines represent best fitting regression lines.

indicating that the DS and the VS are differentially functionally connected to different midbrain regions. Specifically, as shown in Figure 4C, whereas the VS was connected to both d/p and v/a midbrain regions, the DS was significantly more strongly functionally connected to the d/p compared to v/a midbrain [$t(18) = 5.77$, $p < .001$] and the VS was significantly more strongly functionally connected to the v/a midbrain than the DS [$t(18) = 2.49$, $p < .05$]. This Striatal-seed × ROI interaction was significant in both the left- and right-hemisphere midbrain as indicated by two separate Striatal-seed × Midbrain-ROI repeated measures ANOVAs [$F(1, 18) = 18.91$, $p < .001$ and $F(1, 18) = 7.91$, $p < .05$, for left and right midbrain ROIs, respectively].

According to findings regarding the specific role of DS in reinforcement learning (Williams & Eskandar, 2006) and its connection to the SN (Haber, 2003), we hypothesized that the effect of different reinforcements on future behavior depends on the integrity of DS–d/p midbrain connections. Specifically, we hypothesized that α(win) should positively correlate with DS–d/p midbrain connectivity during reward and α(loss) should correlate positively with DS–d/p midbrain connectivity during

Table 1. Functional Connectivity Results

Region Name	Brodmann's Area (BA)	MNI x	MNI y	MNI z	t
DS Connectivity Reward > Loss, FWE Corrected, $p < .05$, $k = 10$					
L dorsal/posterior midbrain		−10	−28	−13	8.63
R dorsal/posterior midbrain		15	−23	−15	11.79
R hippocampus		25	−20	−18	7.54
L parahippocampal gyrus	34	−18	5	−20	8.48
L superior temporal gyrus	22	−53	−13	3	7.92
L insula	13	−43	−23	0	8.09
L thalamus		−15	−5	15	10.30
L caudate		−10	5	10	9.29
R caudate		13	5	10	8.69
L cerebellum		−28	−38	−35	8.69
VS Connectivity Reward > Loss, FWE Corrected, $p < .05$, $k = 10$					
L ventral/anterior midbrain		−13	−15	−20	9.04
R ventral/anterior midbrain		8	−18	−20	7.57
R dorsal/posterior midbrain		13	−23	−15	8.01
R hippocampus		28	−20	−20	8.14
L pons		−5	−30	−28	10.47
L ventral striatum		−18	0	−10	7.17
R ventral striatum		15	3	−8	7.94
L cerebellum		−10	−45	−28	8.61
R cerebellum		10	−45	−28	8.89

loss outcomes but not vice versa. The results confirmed this hypothesis: In the left hemisphere, α(win) was significantly positively correlated with DS–d/p connectivity during reward outcomes ($r = .40$, $p < .05$, one-tailed; Figure 4D, left) and α(loss) was significantly positively correlated with DS–d/p connectivity during loss outcomes ($r = .57$, $p < .05$, one-tailed; Figure 4D, right). Critically, the opposite correlations [i.e., α(win) with DS–d/p during loss and α(loss) with DS–d/p during win] as well as the correlations between learning rates and VS–v/a midbrain connectivity during reward and loss outcomes were not significant (all ps > .15). In the right hemisphere, α(loss) was significantly positively correlated with DS–d/p during loss outcomes ($r = .39$, $p < .05$, one-tailed), but α(win) was not significantly positively correlated with DS–d/p connectivity during reward outcomes ($p = .44$, one-tailed). Again, the opposite correlations (all ps > .17) as well as the correlations between learning rates and VS–v/a midbrain connectivity during reward and loss outcomes were not significant (all ps > .08).

Furthermore, we tested for the same pattern of correlations using the behavioral variables p(loss/switch) and learning speed instead of α(loss) and α(win), respectively. The strength of functional connectivity between DS and d/p midbrain during loss outcomes was significantly positively correlated with p(loss/switch) in both hemispheres ($r = .67$, $p < .05$ and $r = .63$, $p < .05$, one-tailed for left and right hemispheres, respectively). Accordingly, there was a trend toward a negative correlation between the strength of functional DS–d/p midbrain connectivity during win and learning speed (number of trials to reach learning criteria) in both hemispheres ($r = -.38$, $p = .06$ and $r = -.32$, $p = .09$, one-tailed for left and right hemispheres, respectively).

Because the task used in this study was a spatial learning task and it has been suggested that striatal–hippocampus connections are essential for space-related reward learning (Izquierdo et al., 2006; Rossato et al., 2006), we also searched for similar patterns of correlations in functional striatal–hippocampus connectivity. However, we did not find any significant correlation between striatal–hippocampus connectivity and learning rates in either direction (all ps > .10). Thus, the observed pattern of correlations seems to be specific to DS–d/p midbrain connectivity.

DISCUSSION

In this study, we examined functional connectivity between striatal and midbrain subregions and the neural basis of how positive and negative reinforcements are used to guide decision making. We found that different striatal regions, specifically its dorsal and ventral compartments, in which activity correlated with reward prediction errors, are differentially connected to midbrain subregions, particularly with regard to its d/p and v/a compartment. Notably, whereas the VS was functionally connected to both the v/a and d/p midbrain, the DS was significantly more strongly functionally connected to the d/p than the v/a midbrain. We used a computational reinforcement learning model to generate individual estimates of how subjects use rewards and losses separately to guide their behavior in a two-arm bandit task. Individual differences in functional connectivity between the DS and d/p midbrain predicted the impact of positive and negative outcomes on future decision making. Our results suggest a specific role of striatal–

midbrain connections in updating action values in the striatum, and thus, provide novel insights into the neural mechanisms underlying reinforcement-guided decision making.

We found activity in the mPFC, in the amygdala/hippocampus, and in the posterior cingulate cortex related to positive outcomes, which have been shown to code the appetitive value of both primary and secondary reinforcements (Gottfried & Dolan, 2004; Knutson, Fong, Bennett, Adams, & Hommer, 2003). Activity in the ACC, in the insula, and in the lateral OFC was related to negative reinforcement. It has been shown that these regions represent the value of aversive and punishing stimuli (Seymour et al., 2004; O'Doherty, Kringelbach, Rolls, Hornak, & Andrews, 2001). Activity related to behavioral switch after negative outcomes was observed in the lateral OFC and in the ACC. These results are in line with previous studies which have investigated BOLD responses to behavioral adjustment (Cohen et al., 2008; Wrase et al., 2007; Hampton, Bossaerts, & O'Doherty, 2006; O'Doherty, Critchley, Deichmann, & Dolan, 2003; Cools, Clark, Owen, & Robbins, 2002). In general, these findings replicate the results of previous studies on reinforcement processing, and thus, set the stage for the further investigation of the mechanisms of how reinforcements are used to guide decision making.

Intuitively, in situations with only two different options, it may seem difficult to disentangle whether an individual learns to choose the advantageous option by either learning that one option leads to positive feedback or that the other leads to negative feedback. Of course, both processes contribute to learning but there are individual differences in which process dominates (Cohen & Ranganath, 2005; Frank et al., 2004, 2005). Here, using a computational model of reinforcement learning, we were able to disentangle both processes by estimating learning rates for positive and negative outcomes that reflect the degree of learning from either reinforcement separately (i.e., their effect on subsequent behavior). We found a double dissociation such that either learning rate predicted a different aspect of subjects' behavior. Specifically, α(win) predicted the ability to maintain the rewarded action even in the face of probabilistic losses (learning speed), whereas α(loss) predicted the individual tendency to avoid the unrewarded option [p(loss/switch)] but not vice versa. This confirms the validity of these learning rates as shown in previous studies (Cohen & Ranganath, 2005).

Trial-by-trial prediction errors generated by the reinforcement learning model using individual learning rates correlated with activity in separable clusters in the dorsal and ventral striatum. Prediction error responses in the VS have been shown previously in classical as well as operant conditioning tasks (Cohen, 2007; Kim, Shimojo, & O'Doherty, 2006; Pessiglione et al., 2006; O'Doherty et al., 2004; McClure, Berns, & Montague, 2003; O'Doherty, Dayan, et al., 2003; Pagnoni, Zink, Montague, & Berns, 2002). These responses, measured with fMRI, are dopamine-modulated (Pessiglione et al., 2006), indicating that they indeed reflect the reward-related firing of dopaminergic midbrain neurons known from electrophysiologic investigations that are thought to guide learning and goal-directed behavior (Reynolds et al., 2001; Schultz & Dickinson, 2000; Hollerman & Schultz, 1998). Prediction error responses in the DS have been shown exclusively in learning tasks in which reward delivery was dependent on operant responses (Schonberg et al., 2007; Haruno & Kawato, 2006; Haruno et al., 2004; O'Doherty et al., 2004; Tricomi, Delgado, & Fiez, 2004). Schonberg et al. (2007) showed that prediction error responses in the DS but not in the VS distinguished learners from nonlearners and correlated with the individual degree of learning. Such results confirm findings from animal research, which have raised the notion of an actor–critic model of the basal ganglia, in which the VS is involved in reward prediction and the DS uses this information to bias action selection in favor of the advantageous option (Atallah et al., 2007; Williams & Eskandar, 2006; Samejima et al., 2005; O'Doherty et al., 2004; Ito, Dalley, Robbins, & Everitt, 2002; Joel et al., 2002).

Recently, Williams and Eskandar (2006) have shown that activity in the DS in rhesus monkeys correlates with instrumental learning and that microstimulation of the dorsal anterior caudate nucleus during the reinforcement period enhanced the learning of cue–action associations, and thus, strengthened action values. The more input the DS gets during reinforcement, the more reinforcement information can be used to bias future actions. Atallah et al. (2007) provided evidence that in rats, whereas the VS is critical for learning, the DS is important for selecting the appropriate option. According to them, there are two ways the VS might direct the DS. First, the VS could modulate activity in the OFC, which maintains action–reward contingencies that, in turn, exerts top–down control over the DS (Frank & Claus, 2006; Joel & Weiner, 1994). Here, the critical pathway that determines how reinforcements are used by the DS to guide future action would be the OFC–DS connection. Alternatively, the VS might provide reinforcement information to the DS by exerting modulatory control over dopaminergic projections from the SN to the DS. The VS projects to the SN, which in turn projects to the DS (Haber, 2003; Joel & Weiner, 2000). Thus, the critical connection would be the SN–DS pathway. Supporting the latter mechanism, Belin and Everitt (2008) provided evidence that the VS exerts control over dorsal striatal processes via dopaminergic midbrain connections.

Our results confirm these findings: To examine striatal–midbrain connections, we used the striatal regions identified in the prediction error analysis as seed regions in

a functional connectivity analysis. We found significant functional connectivity of the DS with d/p midbrain regions, whereas the VS was functionally connected to the v/a and d/p midbrain. These findings are consistent with animal studies showing that prediction error responses in the striatum are generated in the midbrain (Schultz & Dickinson, 2000; Hollerman & Schultz, 1998). Furthermore, we found a significant Striatal-seed × Midbrain-ROI interaction, indicating that the VS was connected to both midbrain regions, whereas the DS was connected to the d/p midbrain solely. Neurophysiologic investigations in primates have, indeed, shown that the DS and the VS are differently connected to the midbrain (Haber, 2003; Haber et al., 2000). Specifically, the VS is reciprocally connected to the VTA but also sends input to the SN, which in turn is reciprocally connected to the DS (Haber et al., 2000; Lynd-Balta & Haber, 1994b).

The anatomical subdivisions in the midbrain are relatively small compared to the size of an MRI voxel used in our study so that it is difficult to localize precisely the VTA and the SN. However, the pattern of findings in our study is consistent with BOLD generators in the VTA and the SN. First, the anatomical localization of our v/a and d/p midbrain activation is similar to the anatomical localization of the VTA and the SN, respectively in primates (Haber et al., 2000). Second, the pattern of functional connectivity to the DS and the VS itself is consistent with the pattern of anatomical connections found in animal studies (Haber et al., 2000; Lynd-Balta & Haber, 1994a, 1994b). Other studies were also able to investigate distinct midbrain regions, specifically the VTA, using fMRI (D'Ardenne et al., 2008; Wittmann et al., 2005, 2008; Wittmann, Bunzeck, Dolan, & Duzel, 2007; Adcock et al., 2006; Bunzeck & Duzel, 2006; Menon & Levitin, 2005). Additionally, one fMRI study identified functional connectivity between the midbrain and the striatum during learning (Aron et al., 2004). Furthermore, a recent study showed a correlation between reward-related BOLD responses in the VS and the VTA using high-resolution imaging (D'Ardenne et al., 2008). Therefore, it seems plausible to assume that the v/a and d/p midbrain identified in our study correspond to the VTA and the SN, respectively. Nevertheless, it might be interesting to replicate this pattern of functional connectivity using high-resolution fMRI.

The degree of functional connectivity between the DS and the d/p midbrain (presumably the SN) during reward and loss feedback correlated with individual learning rates, α(win) and α(loss), respectively. This was not the case for the functional connectivity between the VS and the v/a midbrain (presumably the VTA). The stronger the functional connectivity between the DS and the SN during a certain feedback type, the more that feedback had an effect on future actions; DS–SN connectivity predicted how subjects used reinforcements to guide future decisions. To our knowledge, this is the first human study that supports evidence in animal research indicating that inputs from the SN drive instrumental learning and action selection in the DS (Belin & Everitt, 2008; Williams & Eskandar, 2006; Faure et al., 2005; Joel & Weiner, 2000). Given this mechanism, our results imply that the more input the DS gets during reinforcement, the more this reinforcement biases action selection. This interpretation is consistent with results from primate studies (Williams & Eskandar, 2006). In our data, this pattern was selective to striatal–midbrain connectivity and was not present in the observed striatal–hippocampus connectivity. Connections between the striatum and the hippocampus, as observed in our study, have been implicated in space-related reward learning (Izquierdo et al., 2006; Rossato et al., 2006; Goto & Grace, 2005), but our data suggest that their integrity do not determine the degree of learning from different feedback types. Previous studies have shown that the degree of learning from either reinforcement depends on various dopaminergic conditions, such as dopamine genetics, dopaminergic drugs, and diseases that target dopaminergic transmission specifically in the SN (Cohen et al., 2007; Frank, Moustafa, et al., 2007; Frank, Scheres, et al., 2007; Klein et al., 2007; Frank, 2005). An interesting hypothesis regarding the impact of different reinforcements on learning is that this should also be revealed in a task that incorporates win versus nonwin and loss versus nonloss outcomes. In this case, a nonloss in the context of a possible loss might be interpreted as a reward (Kim et al., 2006), and thus, the reward learning rate should also correlate with the strength of DS–SN connectivity during nonloss feedback.

A limitation of our experimental design is that it does not allow the distinction between stimulus and action values. Therefore, it is possible that action and/or stimulus values are updated via SN–DS pathways. However, it is clear that the learning rates used here reflect mechanisms of updating behavior according to reinforcements. Future studies could use tasks that investigate the updating of stimulus and action representations separately. Another limitation is that the correlation between functional connectivity estimates and learning rates was modest and should be replicated in future studies. However, these correlations are orthogonal to the definition of the midbrain ROIs that were used. To our knowledge, this is the first investigation into the relation between functional connectivity and model-estimated behavioral performance. Furthermore, we acknowledge that it is not possible to interpret the direction of information flow or the underlying neurotransmitters using functional connectivity measurements. However, our results are in line with the notion that the SN sends information about reinforcements via dopaminergic projections to the DS that uses this information to bias action selection.

In conclusion, here we found reward prediction error responses in the VS and the DS during a dynamic reward-based decision-making task. The DS and the VS were differentially connected to different midbrain regions, the SN and the VTA, respectively. Furthermore, we have shown that the way different reinforcements are used to guide future decisions critically depends on the integrity of DS–SN connectivity. Therefore, our results support the hypothesis that the VTA sends a teaching signal in form of a reward prediction error to the VS that is used to predict rewards. The VS, on the other hand, projects this signal back to the SN, which in turn projects it to the DS in order to bias action selection in favor of the advantageous option (Belin & Everitt, 2008; Haber et al., 2000; Joel & Weiner, 2000). Our results further imply that the latter pathway is crucial for the way different reinforcements influence future decisions, but not the preceding pathways. Thus, VS–VTA connections might be important for reward predictions, whereas DS–SN connections contribute specifically to the use of reinforcements to guide actions. These findings might also help to shed light into the learning deficits observed in patients with Parkinson's disease (Shohamy et al., 2004; Myers et al., 2003; Knowlton, Mangels, & Squire, 1996), where dopaminergic neurons in the SN are diminished (Ito et al., 1999; Owen et al., 1992).

Acknowledgments

This study was supported by the German Research Foundation (Deutsche Forschungsgemeinschaft; HE 2597/4-3) and by the Bernstein Center for Computational Neuroscience Berlin (Bundesministerium für Bildung und Forschung grant 01GQ0411).

Reprint requests should be sent to Thorsten Kahnt, Department of Psychiatry and Psychotherapy, Charité—Universitätsmedizin Berlin (Charité Campus Mitte), Charitéplatz 1, 10117 Berlin, Germany, or via e-mail: thorsten.kahnt@gmail.com.

REFERENCES

Adcock, R. A., Thangavel, A., Whitfield-Gabrieli, S., Knutson, B., & Gabrieli, J. D. (2006). Reward-motivated learning: Mesolimbic activation precedes memory formation. *Neuron, 50,* 507–517.

Alexander, G. E., Crutcher, M. D., & DeLong, M. R. (1990). Basal ganglia–thalamocortical circuits: Parallel substrates for motor, oculomotor, ''prefrontal'' and ''limbic'' functions. *Progress in Brain Research, 85,* 119–146.

Aron, A. R., Shohamy, D., Clark, J., Myers, C., Gluck, M. A., & Poldrack, R. A. (2004). Human midbrain sensitivity to cognitive feedback and uncertainty during classification learning. *Journal of Neurophysiology, 92,* 1144–1152.

Atallah, H. E., Lopez-Paniagua, D., Rudy, J. W., & O'Reilly, R. C. (2007). Separate neural substrates for skill learning and performance in the ventral and dorsal striatum. *Nature Neuroscience, 10,* 126–131.

Belin, D., & Everitt, B. J. (2008). Cocaine seeking habits depend upon dopamine-dependent serial connectivity linking the ventral with the dorsal striatum. *Neuron, 57,* 432–441.

Buchel, C., Holmes, A. P., Rees, G., & Friston, K. J. (1998). Characterizing stimulus–response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage, 8,* 140–148.

Bunzeck, N., & Duzel, E. (2006). Absolute coding of stimulus novelty in the human substantia Nigra/VTA. *Neuron, 51,* 369–379.

Cohen, M. X (2007). Individual differences and the neural representations of reward expectation and reward prediction error. *Social Cognitive and Affective Neuroscience, 2,* 20–30.

Cohen, M. X, Elger, C. E., & Weber, B. (2008). Amygdala tractography predicts functional connectivity and learning during feedback-guided decision making. *Neuroimage, 39,* 1396–1407.

Cohen, M. X, Heller, A. S., & Ranganath, C. (2005). Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Brain Research, Cognitive Brain Research, 23,* 61–70.

Cohen, M. X, Krohn-Grimberghe, A., Elger, C. E., & Weber, B. (2007). Dopamine gene predicts the brain's response to dopaminergic drug. *European Journal of Neuroscience, 26,* 3652–3660.

Cohen, M. X, & Ranganath, C. (2005). Behavioral and neural predictors of upcoming decisions. *Cognitive, Affective & Behavioral Neuroscience, 5,* 117–126.

Cohen, M. X, & Ranganath, C. (2007). Reinforcement learning signals predict future decisions. *Journal of Neuroscience, 27,* 371–378.

Cools, R., Clark, L., Owen, A. M., & Robbins, T. W. (2002). Defining the neural mechanisms of probabilistic reversal learning using event-related functional magnetic resonance imaging. *Journal of Neuroscience, 22,* 4563–4567.

D'Ardenne, K., McClure, S. M., Nystrom, L. E., & Cohen, J. D. (2008). BOLD responses reflecting dopaminergic signals in the human ventral tegmental area. *Science, 319,* 1264–1267.

Delgado, M. R. (2007). Reward-related responses in the human striatum. *Annals of the New York Academy of Sciences, 1104,* 70–88.

Duvernoy, H. (1999). *The human brain. Surface, blood supply, and three-dimensional sectional anatomy* (2nd ed.). Vienna: Springer-Verlag.

Egelman, D. M., Person, C., & Montague, P. R. (1998). A computational role for dopamine delivery in human decision-making. *Journal of Cognitive Neuroscience, 10,* 623–630.

Faure, A., Haberland, U., Conde, F., & El Massioui, N. (2005). Lesion to the nigrostriatal dopamine system disrupts stimulus–response habit formation. *Journal of Neuroscience, 25,* 2771–2780.

Frank, M. J. (2005). Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *Journal of Cognitive Neuroscience, 17,* 51–72.

Frank, M. J., & Claus, E. D. (2006). Anatomy of a decision: Striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychological Review, 113,* 300–326.

Frank, M. J., Moustafa, A. A., Haughey, H. M., Curran, T., & Hutchison, K. E. (2007). Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proceedings of the National Academy of Sciences, U.S.A., 104,* 16311–16316.

Frank, M. J., Scheres, A., & Sherman, S. J. (2007). Understanding decision-making deficits in neurological conditions: Insights from models of natural action selection.

Philosophical Transactions of the Royal Society of London, Series B, Biological Sciences, 362, 1641–1654.
Frank, M. J., Seeberger, L. C., & O'Reilly, R. C. (2004). By carrot or by stick: Cognitive reinforcement learning in parkinsonism. *Science, 306,* 1940–1943.
Frank, M. J., Woroch, B. S., & Curran, T. (2005). Error-related negativity predicts reinforcement learning and conflict biases. *Neuron, 47,* 495–501.
Friston, K. J., Buechel, C., Fink, G. R., Morris, J., Rolls, E., & Dolan, R. J. (1997). Psychophysiological and modulatory interactions in neuroimaging. *Neuroimage, 6,* 218–229.
Goto, Y., & Grace, A. A. (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nature Neuroscience, 8,* 805–812.
Gottfried, J. A., & Dolan, R. J. (2004). Human orbitofrontal cortex mediates extinction learning while accessing conditioned representations of value. *Nature Neuroscience, 7,* 1145–1153.
Haber, S. N. (2003). The primate basal ganglia: Parallel and integrative networks. *Journal of Chemical Neuroanatomy, 26,* 317–330.
Haber, S. N., Fudge, J. L., & McFarland, N. R. (2000). Striatonigrostriatal pathways in primates form an ascending spiral from the shell to the dorsolateral striatum. *Journal of Neuroscience, 20,* 2369–2382.
Hampton, A. N., Bossaerts, P., & O'Doherty, J. P. (2006). The role of the ventromedial prefrontal cortex in abstract state-based inference during decision making in humans. *Journal of Neuroscience, 26,* 8360–8367.
Haruno, M., & Kawato, M. (2006). Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus–action–reward association learning. *Journal of Neurophysiology, 95,* 948–959.
Haruno, M., Kuroda, T., Doya, K., Toyama, K., Kimura, M., Samejima, K., et al. (2004). A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *Journal of Neuroscience, 24,* 1660–1665.
Hollerman, J. R., & Schultz, W. (1998). Dopamine neurons report an error in the temporal prediction of reward during learning. *Nature Neuroscience, 1,* 304–309.
Holroyd, C. B., & Coles, M. G. (2002). The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychological Review, 109,* 679–709.
Ikemoto, S. (2007). Dopamine reward circuitry: Two projection systems from the ventral midbrain to the nucleus accumbens–olfactory tubercle complex. *Brain Research Reviews, 56,* 27–78.
Ito, K., Morrish, P. K., Rakshi, J. S., Uema, T., Ashburner, J., Bailey, D. L., et al. (1999). Statistical parametric mapping with 18F-dopa PET shows bilaterally reduced striatal and nigral dopaminergic function in early Parkinson's disease. *Journal of Neurology, Neurosurgery and Psychiatry, 66,* 754–758.
Ito, R., Dalley, J., Robbins, T., & Everitt, B. (2002). Dopamine release in the dorsal striatum during cocaine seeking behavior under the control of a drug-associated cue. *Journal of Neuroscience, 22,* 6247–6253.
Izquierdo, I., Bevilaqua, L. R., Rossato, J. I., Bonini, J. S., Da Silva, W. C., Medina, J. H., et al. (2006). The connection between the hippocampal and the striatal memory systems of the brain: A review of recent findings. *Neurotoxicity Research, 10,* 113–121.
Joel, D., Niv, Y., & Ruppin, E. (2002). Actor–critic models of the basal ganglia: New anatomical and computational perspectives. *Neural Networks, 15,* 535–547.
Joel, D., & Weiner, I. (1994). The organization of the basal ganglia–thalamocortical circuits: Open interconnected rather than closed segregated. *Neuroscience, 63,* 363–379.
Joel, D., & Weiner, I. (2000). The connections of the dopaminergic system with the striatum in rats and primates: An analysis with respect to the functional and compartmental organization of the striatum. *Neuroscience, 96,* 451–474.
Kim, H., Shimojo, S., & O'Doherty, J. P. (2006). Is avoiding an aversive outcome rewarding? Neural substrates of avoidance learning in the human brain. *PLoS Biology, 4,* e233.
Klein, T. A., Neumann, J., Reuter, M., Hennig, J., von Cramon, D. Y., & Ullsperger, M. (2007). Genetically determined differences in learning from errors. *Science, 318,* 1642–1645.
Knowlton, B. J., Mangels, J. A., & Squire, L. R. (1996). A neostriatal habit learning system in humans. *Science, 273,* 1399–1402.
Knutson, B., Fong, G. W., Bennett, S. M., Adams, C. M., & Hommer, D. (2003). A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: Characterization with rapid event-related fMRI. *Neuroimage, 18,* 263–272.
Lehericy, S., Ducros, M., Van de Moortele, P. F., Francois, C., Thivard, L., Poupon, C., et al. (2004). Diffusion tensor fiber tracking shows distinct corticostriatal circuits in humans. *Annals of Neurology, 55,* 522–529.
Lynd-Balta, E., & Haber, S. N. (1994a). Primate striatonigral projections: A comparison of the sensorimotor-related striatum and the ventral striatum. *Journal of Comparative Neurology, 345,* 562–578.
Lynd-Balta, E., & Haber, S. N. (1994b). The organization of midbrain projections to the striatum in the primate: Sensorimotor-related striatum versus ventral striatum. *Neuroscience, 59,* 625–640.
McClure, S. M., Berns, G. S., & Montague, P. R. (2003). Temporal prediction errors in a passive learning task activate human striatum. *Neuron, 38,* 339–346.
Menon, V., & Levitin, D. J. (2005). The rewards of music listening: Response and physiological connectivity of the mesolimbic system. *Neuroimage, 28,* 175–184.
Montague, P. R., Hyman, S. E., & Cohen, J. D. (2004). Computational roles for dopamine in behavioural control. *Nature, 431,* 760–767.
Myers, C. E., Shohamy, D., Gluck, M. A., Grossman, S., Kluger, A., Ferris, S., et al. (2003). Dissociating hippocampal versus basal ganglia contributions to learning and transfer. *Journal of Cognitive Neuroscience, 15,* 185–193.
O'Doherty, J. P., Critchley, H., Deichmann, R., & Dolan, R. J. (2003). Dissociating valence of outcome from behavioral control in human orbital and ventral prefrontal cortices. *Journal of Neuroscience, 23,* 7931–7939.
O'Doherty, J. P., Dayan, P., Friston, K., Critchley, H., & Dolan, R. J. (2003). Temporal difference models and reward-related learning in the human brain. *Neuron, 38,* 329–337.
O'Doherty, J. P., Dayan, P., Schultz, J., Deichmann, R., Friston, K., & Dolan, R. J. (2004). Dissociable roles of ventral and dorsal striatum in instrumental conditioning. *Science, 304,* 452–454.
O'Doherty, J. P., Kringelbach, M. L., Rolls, E. T., Hornak, J., & Andrews, C. (2001). Abstract reward and punishment representations in the human orbitofrontal cortex. *Nature Neuroscience, 4,* 95–102.
Owen, A. M., James, M., Leigh, P. N., Summers, B. A., Marsden, C. D., Quinn, N. P., et al. (1992). Fronto-striatal cognitive deficits at different stages of Parkinson's disease. *Brain, 115,* 1727–1751.

Pagnoni, G., Zink, C. F., Montague, P. R., & Berns, G. S. (2002). Activity in human ventral striatum locked to errors of reward prediction. *Nature Neuroscience, 5,* 97–98.
Pessiglione, M., Seymour, B., Flandin, G., Dolan, R. J., & Frith, C. D. (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature, 442,* 1042–1045.
Pessoa, L., Gutierrez, E., Bandettini, P., & Ungerleider, L. (2002). Neural correlates of visual working memory: fMRI amplitude predicts task performance. *Neuron, 35,* 975–987.
Reynolds, J. N., Hyland, B. I., & Wickens, J. R. (2001). A cellular mechanism of reward-related learning. *Nature, 413,* 67–70.
Rossato, J. I., Zinn, C. G., Furini, C., Bevilaqua, L. R., Medina, J. H., Cammarota, M., et al. (2006). A link between the hippocampal and the striatal memory systems of the brain. *Annals of the Brazilian Academy of Sciences, 78,* 515–523.
Samejima, K., & Doya, K. (2007). Multiple representations of belief states and action values in corticobasal ganglia loops. *Annals of the New York Academy of Sciences, 1104,* 213–228.
Samejima, K., Ueda, Y., Doya, K., & Kimura, M. (2005). Representation of action-specific reward values in the striatum. *Science, 310,* 1337–1340.
Schonberg, T., Daw, N. D., Joel, D., & O'Doherty, J. P. (2007). Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *Journal of Neuroscience, 27,* 12860–12867.
Schultz, W. (2002). Getting formal with dopamine and reward. *Neuron, 36,* 241–263.
Schultz, W. (2004). Neural coding of basic reward terms of animal learning theory, game theory, microeconomics and behavioural ecology. *Current Opinion in Neurobiology, 14,* 139–147.
Schultz, W., Dayan, P., & Montague, P. R. (1997). A neural substrate of prediction and reward. *Science, 275,* 1593–1599.
Schultz, W., & Dickinson, A. (2000). Neuronal coding of prediction errors. *Annual Review of Neuroscience, 23,* 473–500.
Seymour, B., O'Doherty, J. P., Dayan, P., Koltzenburg, M., Jones, A. K., Dolan, R. J., et al. (2004). Temporal difference models describe higher-order learning in humans. *Nature, 429,* 664–667.
Shohamy, D., Myers, C. E., Grossman, S., Sage, J., Gluck, M. A., & Poldrack, R. A. (2004). Cortico-striatal contributions to feedback-based learning: Converging data from neuroimaging and neuropsychology. *Brain, 127,* 851–859.
Sutton, R., & Barto, A. (1998). *Reinforcement learning: An introduction*. Cambridge, MA: MIT Press.
Talairach, J., & Tournoux, P. (1988). *Co-planar stereotaxic atlas of the human brain*. New York: Thieme.
Tricomi, E. M., Delgado, M. R., & Fiez, J. A. (2004). Modulation of caudate activity by action contingency. *Neuron, 41,* 281–292.
Williams, Z. M., & Eskandar, E. N. (2006). Selective enhancement of associative learning by microstimulation of the anterior caudate. *Nature Neuroscience, 9,* 562–568.
Wittmann, B. C., Bunzeck, N., Dolan, R. J., & Duzel, E. (2007). Anticipation of novelty recruits reward system and hippocampus while promoting recollection. *Neuroimage, 38,* 194–202.
Wittmann, B. C., Schiltz, K., Boehler, C. N., & Duzel, E. (2008). Mesolimbic interaction of emotional valence and reward improves memory formation. *Neuropsychologia, 46,* 1000–1008.
Wittmann, B. C., Schott, B. H., Guderian, S., Frey, J. U., Heinze, H. J., & Duzel, E. (2005). Reward-related fMRI activation of dopaminergic midbrain is associated with enhanced hippocampus-dependent long-term memory formation. *Neuron, 45,* 459–467.
Wrase, J., Kahnt, T., Schlagenhauf, F., Beck, A., Cohen, M. X., Knutson, B., et al. (2007). Different neural systems adjust motor behavior in response to reward and punishment. *Neuroimage, 36,* 1253–1262.

This article has been cited by:

1. Cara Bohon, Eric Stice. 2012. Negative affect and neural response to palatable food intake in bulimia nervosa. *Appetite* **58**:3, 964-970. [CrossRef]
2. Anne G. E. Collins, Michael J. Frank. 2012. How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience* **35**:7, 1024-1035. [CrossRef]
3. Louis Anthony Tony Cox. 2012. Confronting Deep Uncertainties in Risk Analysis. *Risk Analysis* no-no. [CrossRef]
4. Florian Schlagenhauf, Michael A. Rapp, Quentin J. M. Huys, Anne Beck, Torsten Wüstenberg, Lorenz Deserno, Hans-Georg Buchholz, Jan Kalbitzer, Ralph Buchert, Michael Bauer, Thorsten Kienast, Paul Cumming, Michail Plotkin, Yoshitaka Kumakura, Anthony A. Grace, Raymond J. Dolan, Andreas Heinz. 2012. Ventral striatal prediction error signaling is associated with dopamine synthesis capacity and fluid intelligence. *Human Brain Mapping* n/a-n/a. [CrossRef]
5. W. van den Bos, M. X. Cohen, T. Kahnt, E. A. Crone. 2011. Striatum-Medial Prefrontal Cortex Connectivity Predicts Developmental Changes in Reinforcement Learning. *Cerebral Cortex* . [CrossRef]
6. Michael Michaelides, Panayotis K. Thanos, Ronald Kim, Jacob Cho, Mala Ananth, Gene-Jack Wang, Nora D. Volkow. 2011. PET imaging predicts future body weight and cocaine preference. *NeuroImage* . [CrossRef]
7. Thorsten Kahnt, Marcus Grueschow, Oliver Speck, John-Dylan Haynes. 2011. Perceptual Learning and Decision-Making in Human Medial Frontal Cortex. *Neuron* **70**:3, 549-559. [CrossRef]
8. Thorsten Kahnt, Jakob Heinzle, Soyoung Q. Park, John-Dylan Haynes. 2011. Decoding different roles for vmPFC and dlPFC in multi-attribute decision making. *NeuroImage* **56**:2, 709-715. [CrossRef]
9. Tobias Brosch, Geraldine Coppin, Klaus Scherer, Sophie Schwartz, David Sander. 2011. Generating value(s): Psychological value hierarchies reflect context-dependent sensitivity of the reward system. *Social Neuroscience* **6**:2, 198-208. [CrossRef]
10. Nicola Canessa, Matteo Motterlini, Federica Alemanno, Daniela Perani, Stefano F. Cappa. 2011. Learning from other people's experience: A neuroimaging study of decisional interactive-learning. *NeuroImage* **55**:1, 353-362. [CrossRef]
11. Daniel W. Hommer, James M. Bjork, Jodi M. Gilman. 2011. Imaging brain response to reward in addictive disorders. *Annals of the New York Academy of Sciences* **1216**:1, 50-61. [CrossRef]
12. Dongju Seo, Zhiru Jia, Cheryl M. Lacadie, Kristen A. Tsou, Keri Bergquist, Rajita Sinha. 2011. Sex differences in neural responses to stress and alcohol context cues. *Human Brain Mapping* n/a-n/a. [CrossRef]
13. A. Heinz, F. Schlagenhauf. 2010. Dopaminergic Dysfunction in Schizophrenia: Salience Attribution Revisited. *Schizophrenia Bulletin* **36**:3, 472-485. [CrossRef]
14. Bastian Sajonz, Thorsten Kahnt, Daniel S. Margulies, Soyoung Q. Park, André Wittmann, Meline Stoy, Andreas Ströhle, Andreas Heinz, Georg Northoff, Felix Bermpohl. 2010. Delineating self-referential processing from episodic memory retrieval: Common and dissociable networks. *NeuroImage* **50**:4, 1606-1617. [CrossRef]
15. M.J. Frank, K. Hutchison. 2009. Genetic contributions to avoidance-based decisions: striatal D2 receptor polymorphisms. *Neuroscience* **164**:1, 131-140. [CrossRef]