Behavioral/Cognitive

Orbitofrontal Dopamine Depletion Upregulates Caudate Dopamine and Alters Behavior via Changes in Reinforcement Sensitivity

H. F. Clarke,[1,3]* R. N. Cardinal,[3,4,6]* R. Rygula,[2,3] Y. T. Hong,[5] T. D. Fryer,[5] S. J. Sawiak,[3,5] V. Ferrari,[5] G. Cockcroft,[1,3] F. I. Aigbirhio,[3,5] T. W. Robbins,[2,3] and A. C. Roberts[1,3]

[1]Departments of Physiology, Development and Neuroscience, [2]Psychology, and [3]Behavioral and Clinical Neuroscience Institute, University of Cambridge, Cambridge CB2 3EB, United Kingdom, [4]Department of Psychiatry and [5]Wolfson Brain Imaging Centre, Addenbrooke's Hospital, Cambridge CB2 0QQ, United Kingdom, and [6]Liaison Psychiatry Service, Cambridge and Peterborough NHS Foundation Trust, Cambridge CB2 0QQ, United Kingdom

Schizophrenia is associated with upregulation of dopamine (DA) release in the caudate nucleus. The caudate has dense connections with the orbitofrontal cortex (OFC) via the frontostriatal loops, and both areas exhibit pathophysiological change in schizophrenia. Despite evidence that abnormalities in dopaminergic neurotransmission and prefrontal cortex function co-occur in schizophrenia, the influence of OFC DA on caudate DA and reinforcement processing is poorly understood. To test the hypothesis that OFC dopaminergic dysfunction disrupts caudate dopamine function, we selectively depleted dopamine from the OFC of marmoset monkeys and measured striatal extracellular dopamine levels (using microdialysis) and dopamine D2/D3 receptor binding (using positron emission tomography), while modeling reinforcement-related behavior in a discrimination learning paradigm. OFC dopamine depletion caused an increase in tonic dopamine levels in the caudate nucleus and a corresponding reduction in D2/D3 receptor binding. Computational modeling of behavior showed that the lesion increased response exploration, reducing the tendency to persist with a recently chosen response side. This effect is akin to increased response switching previously seen in schizophrenia and was correlated with striatal but not OFC D2/D3 receptor binding. These results demonstrate that OFC dopamine depletion is sufficient to induce striatal hyperdopaminergia and changes in reinforcement learning relevant to schizophrenia.

Key words: **behavior; caudate nucleus; dopamine; orbitofrontal cortex; PET; schizophrenia**

Introduction

Modern versions of the dopamine (DA) hypothesis of schizophrenia suggest that important changes in DA function occur at two sites, the striatum and prefrontal cortex (PFC; Weinberger, 1987). In the striatum increased presynaptic DA synthesis and increased striatal D2 receptors correlate with the magnitude of positive symptoms in schizophrenia (Miyake et al., 2011) and blockade of striatal D2 receptors (Davis et al., 1991; Kapur and Remington, 2001) alleviates such symptoms. Moreover, the onset of psychosis is heralded by changes in DA function specifically within the caudate nucleus (Howes et al., 2009; Fusar-Poli et al., 2010), a key site of the increased D2 receptor availability seen in schizophrenia (Miyake et al., 2011). Decreased D1 receptor neurotransmission in the PFC is proposed to cause the "negative" (cognitive deficit) symptoms (Weinberger, 1987), although DA D3/D4 receptor mRNA is also downregulated in the orbitofrontal cortex (OFC; Meador-Woodruff et al., 1997) and in the cognitive-deficit syndrome of schizophrenia (Kanahara et al., 2013).

This raises the question as to whether these striatal and orbitofrontal changes observed in schizophrenia are causally related. Previous studies have provided evidence for interactions between other prefrontal cortical regions and striatal dopamine activity (Pycock et al., 1980; Roberts et al., 1994; Kolachana et al., 1995; Scornaiencki et al., 2009). Furthermore, the OFC not only innervates the caudate nucleus, but also projects directly and indirectly to the midbrain ascending DA systems (Leichnetz and Astruc, 1975; Haber et al., 1995) where it inhibits ventral tegmental area (VTA) neurons (Lodge, 2011), whereas glucose metabolism in the OFC correlates with D2 receptor availability in the human striatum (Volkow et al., 2001). Finally, prolonged psychological stress, a known risk factor and trigger for schizophrenia (van Winkel et al., 2008), reduces PFC DA transmission (Mizoguchi et al., 2000) and increases striatal DA uptake (Copeland et al., 2005). However, the specific relationship between DA in the OFC and the striatum has not yet been studied.

Received Feb. 20, 2014; revised April 1, 2014; accepted April 21, 2014.

Author contributions: H.F.C., R.R., T.W.R., and A.C.R. designed research; H.F.C., R.R., Y.T.H., T.D.F., S.J.S., and G.C. performed research; V.F. and F.I.A. contributed unpublished reagents/analytic tools; H.F.C., R.N.C., Y.T.H., T.D.F., and S.J.S. analyzed data; H.F.C., R.N.C., T.W.R., and A.C.R. wrote the paper.

This work was supported by a Wellcome Trust Grant (089589/Z/09/Z) to T.W.R., B. J. Everitt, A.C.R., J. W. Dalley, and B. J. Sahakian, and was conducted at the Behavioural and Clinical Neuroscience Institute, which is supported by a joint award from the Medical Research Council and Wellcome Trust (G000001354). R.N.C. was supported by a Wellcome Trust postdoctoral fellowship. We thank Nathaniel Daw and Paul Cumming for helpful advice and discussion.

The authors declare no competing financial interests.

*H.F.C. and R.N.C. contributed equally to this work.

This article is freely available online through the *J Neurosci* Author Open Choice option.

Correspondence should be addressed to Dr Hannah Clarke, Department of Physiology, Development, and Neuroscience, University of Cambridge, Downing Street, Cambridge CB2 3DY, UK. E-mail: hfc23@cam.ac.uk.

DOI:10.1523/JNEUROSCI.0718-14.2014

Thus, the present study determined whether depletions of dopamine, specifically within the OFC, can cause changes in D2 receptor transmission in the caudate nucleus. In a New World primate, the common marmoset, OFC dopamine was reduced using the neurotoxin 6-hydroxydopamine (6-OHDA) and its effects on striatal DA were assessed using ^{18}F-fallypride positron emission tomography (PET) to quantify D2/3 receptor binding, and *in vivo* microdialysis to assess levels of extracellular DA. In addition, the effects of OFC DA reductions were determined on performance of a probabilistic discrimination task in which marmosets had to learn which of two visual stimuli was more associated with reward. Patients with schizophrenia can show two distinct behavioral changes compared with controls on such tasks: they can adopt different strategies, such as switching response location at different rates (Frith and Done, 1983), and they can show altered sensitivity to positive or negative feedback that impacts upon learning (Waltz et al., 2007). How such behavioral changes relate to altered prefrontostriatal DA function is unclear. Therefore, we applied computational reinforcement learning models to subjects' performance to test for changes in either strategy or reinforcement learning.

Materials and Methods

Overview and behavioral methods

Subjects ($n = 7$) completed a probabilistic discrimination learning task consisting of multiple discriminations, each comprising two abstract multicolored stimuli presented on a touch-sensitive computer screen as described previously (Clarke et al., 2007). All monkeys were trained to enter a clear plastic transport box for marshmallow reward, familiarized with the testing apparatus, and trained to respond to the touchscreen. They learned through trial and error which stimulus was usually (70 or 80%) associated with a 5 s banana milkshake reward and sometimes punished (30 or 20%) with a 0.3 s 100 dB loud noise, and vice versa (Fig. 1). Subjects completed one session of 40 trials per day and an individual discrimination was considered learned when they reached a criterion of 90% or more correct choices in one session. A new discrimination was then started the next day. The rate of learning was assessed by calculating how many incorrect choices were made during each discrimination. While learning the preoperative discriminations, they were scanned with ^{18}F-fallypride to assess their D2/D3 receptor nondisplaceable binding potential [BP_{ND}] (D2RB). Once the task was learned, they underwent a 6-OHDA-induced selective depletion of DA within the OFC or a control procedure. When recovered, they continued with postoperative discriminations, and ~16 weeks after surgery were rescanned with ^{18}F-fallypride to assess their postoperative DR2B and microdialyzed to assess the levels of extracellular DA in the caudate nucleus.

Subjects and housing

Seven common marmosets (*Callithrix jacchus*; 3 females, 4 males) bred on site at the University of Cambridge Marmoset Breeding Colony were housed in pairs. All monkeys were fed 20 g of MP.E1 primate diet (Special Diet Services) and two pieces of carrot 5 d per week after the daily behavioral testing session, with simultaneous access to water for 2 h. On weekends, their diet was supplemented with fruit, rusk, malt loaf, eggs, bread, and treats and they had *ad libitum* access to water. Their cages contained a variety of environmental enrichment aids that were varied regularly and all procedures were performed in accordance with the UK Animals (Scientific Procedures) Act, 1986. One sham-operated control subject contributed to imaging data and dialysis ($n = 7$) but was not part of the behavioral study ($n = 6$) and its postmortem data were lost due to a freezer malfunction.

Structural magnetic resonance imaging

Subjects were premedicated with ketamine hydrochloride (Pharmacia and Upjohn, 0.05 ml of a 100 mg/ml solution, i.m.) and given a long-lasting prophylactic analgesic (Carprieve; 0.03 ml of 50 mg/ml carprofen, s.c.; Pfizer). The tail vein was cannulated (Intraflon 2 i.v. catheter attached to a Lock Stopper with injectable membrane; Vygon), the cannula was flushed with 0.5 ml saline and 0.25 ml heparinized saline and the monkey subsequently intubated and maintained on isoflurane gas anesthetic (flow rate: 2.0–2.5% isoflurane in 0.3 l/min O_2; Novartis). The monkey was positioned in the MRI scanner and monitored throughout (pulse oximetry, temperature).

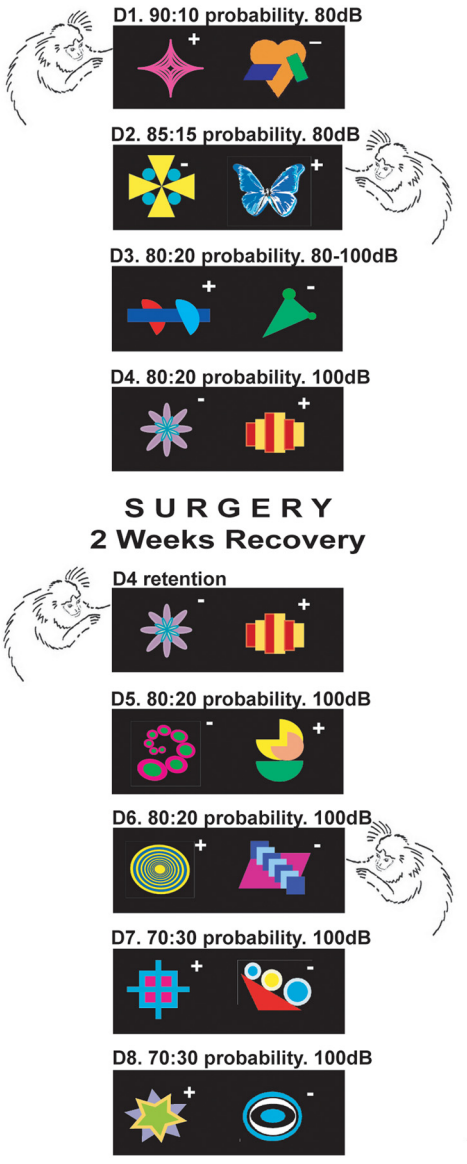

Figure 1. Task sequence (D, discrimination). Representative stimuli are shown, labeled + for correct and − for incorrect. Reinforcement probabilities are shown: for example, "90:10 probability" indicates that P(reward | correct stimulus selected) = P(punishment | incorrect stimulus selected) = 0.9 and P(punishment | correct stimulus selected) = P(reward | incorrect stimulus selected) = 0.1. The intensity of auditory punishment is shown in dB SPL.

Animals were scanned supine in a Bruker PharmaScan 47/16 system, using a locally built birdcage coil for signal transmission and reception. Structural images were obtained using a RARE sequence optimized for contrast between gray and white matter (TR/TE_{eff} 7455/36 ms, echo train length 8, field-of-view 7.68 × 7.68 cm, matrix 256 × 192, reconstructed to final resolution 300 × 300 μm, 50 slices of thickness 1 mm with gap 0.2 mm). Regions-of-interest (ROIs) were delineated for each subject independently on a slice-by-slice basis by a single expert reviewer (A.C.R.) using Analyze 8.1 (Mayo Clinic). ROIs were drawn for the orbitofrontal cortex, ventrolateral prefrontal cortex, ventromedial caudate, caudate body, dorsolateral caudate, putamen, nucleus accumbens, amygdala, ventral hippocampus, and cerebellum. Upon completion of the MRI scans, the monkeys were transferred to the PET scanner while still unconscious, and the PET scan commenced.

^{18}F-Fallypride positron emission tomography

To determine the effects of an OFC hypodopaminergic state on D2 receptor availability in the striatum we used PET imaging with the highly selective dopamine D2/D3 receptor radioligand ^{18}F-fallypride. The high affinity of ^{18}F-fallypride allows the investigation of areas of both high and low D2/D3 receptor density (e.g., the striatum and PFC respectively; Lataster et al., 2011). The marmoset OFC preferentially innervates the ventromedial caudate in the striatum (Roberts et al., 2007), the caudate being implicated in the increase in D2 receptor availability seen in schizophrenia (Miyake et al., 2011). Thus, the caudate nucleus was an a priori ROI. As it has been demonstrated that baseline striatal DA synthesis (Cools et al., 2009) and D2 receptor binding (Groman et al., 2011) varies between individuals, the monkeys were scanned both before and after surgery so that each monkey could act as its own control when assessing the effects of OFC DA depletion on caudate D2/D3 receptor binding.

Animals were imaged using a microPET P4 scanner (Concorde Microsystems). The brain was located centrally in the field of view of the scanner (78 mm axial × 200 mm diameter) to maximize sensitivity and spatial resolution. The amount of ^{18}F-fallypride injected was governed by the desire to minimize any mass-related perturbation of receptor availability, while also providing adequate counting statistics. Consequently, 0.49 ± 0.04 nmol/kg was injected, which corresponded to an activity range of 5.1–23.1 MBq across the animals. ^{18}F-Fallypride was injected intravenously as a bolus over 10 s, followed by a 10 s heparinized saline flush. List-mode data acquired over 180 min after injection were subsequently histogrammed into the following time frames: 10 × 10 s, 3 × 20 s, 6 × 30 s, 10 × 60 s, 10 × 120 s, and 29 × 300 s. The energy and timing windows used were 350–650 keV and 6 ns, respectively. Before injection, windowed coincidence mode transmission data were collected for 11 min with a rotating ^{68}Ge point source (~100 MBq) to allowed measured attenuation correction.

The images were reconstructed using the PROMIS 3D filtered backprojection algorithm (Kinahan and Rogers, 1989), adapted locally for the specific scanner. Corrections for randoms, dead time, background, normalization, attenuation, and sensitivity were applied to the data during reconstruction. Images were reconstructed into 0.5 × 0.5 × 0.5 mm^3 voxels in a 180 × 180 × 151 array, and a Hann window cutoff at the Nyquist frequency was incorporated into the reconstruction filters to give an image resolution of ~2.3 mm (full-width, half-maximum). For each scan an added image (120–180 min) was coregistered to its own MRI using rigid coregistration. ROIs delineated on the MRI were applied to the coregistered dynamic PET images to extract ROI time-activity curves (TACs).

ROI nondisplaceable binding potential was estimated from the ROI TACs with the simplified reference tissue model (reference tissue: cerebellum) using basis functions (sRTM; Gunn et al., 1997). One-hundred-fifty basis functions spaced logarithmically in the range of $0.009 \leq \theta_3 \leq 0.60$ 1/min were used.

Depletion of dopamine from the orbitofrontal cortex

Subjects were premedicated, given an analgesic, and anesthetized as above, before being placed in a stereotaxic frame modified for the marmoset (David Kopf). Anesthesia was monitored clinically and by pulse oximetry with capnography.

Lesions of the dopaminergic innervation of the OFC were made using 6-OHDA (Sigma-Aldrich; 6 μg/μl) in saline/0.1% L-ascorbic acid. To protect the serotoninergic innervation of the OFC from the 6-OHDA the selective serotonin reuptake inhibitor citalopram (Lundbeck; 5 mg/kg) was administered concomitantly in the infusate. Injections (0.04 μl/20 s) were made into five sites on each side within the OFC, using a 30 gauge cannula attached to a 2 μl Hamilton syringe. All injections were made 0.7 mm above the base of the brain. The coordinates and volumes used were as follows: AP +16.75: LM ± 2.5 (0.4 μl) and LM ± 3.5 (0.4 μl); AP +17.75: LM ± 2.0 (0.4 μl) and LM ± 3.0 (0.4 μl); and AP +18.5: LM ± 2.0 (0.6 μl), having been adjusted where necessary *in situ* according to cortical depth (Roberts et al., 2007). Sham surgery was identical except for the omission of the toxin from the infusion. Postoperatively, all monkeys received the analgesic meloxicam (0.1 ml of a 1.5 mg/ml oral suspension; Boehringer Ingelheim) before being returned to their home cage for 10 d of "weekend diet" and water *ad libitum* to allow complete recovery before returning to testing.

In vivo *striatal microdialysis*

Following isoflurane anesthesia, commercially available BASi brain microdialysis probes with a 2 mm membrane (BASI MD-2200, BR-2, Bioanalytical Systems) were implanted acutely into the ventromedial (AP +12.5 mm; L 2.3 mm; DV +9.8 mm) and lateral (AP +12 mm; L 3.5 mm; DV +11.0) caudate nucleus and used for collection of the dialysate. Harvard microsyringe pumps with 2.5 ml gas-tight syringes were used to perfuse artificial CSF (aCSF) through the dialysis probe at a flow rate of 1.0 μl/min. The aCSF had the following composition (in mM): NaCl 147, KCl 3.0, $CaCl_2$ 1.3, $MgCl_2$ 1.0, NaH_2PO_4 0.2, and Na_2HPO_4 1.3. After allowing 3 h for the implanted probes to equilibrate, dialysate fractions were collected every 20 min into 2 μl 0.01 M perchloric acid. After three baseline samples, monkeys received a 75 mM K^+ challenge for 10 min, which was followed by a further four baseline samples. Samples were stored at −80°C before being analyzed using reversed phase high-performance liquid chromatography (HPLC) and electrochemical detection as described previously (Clarke et al., 2007). The signal was integrated using Chromeleon software (v6.2, Dionex). Due to HPLC malfunction there was loss of data from the ventromedial caudate of one monkey and the lateral caudate from another. As the values were similar across the two regions values from all animals were averaged across the two areas, where available.

Postmortem histochemical assessment

The specificity and extent of OFC DA depletion following 6-OHDA infusions into the OFC was assessed by postmortem analysis of monoamine levels in cortical and subcortical regions 448.75 ± 5.70 (mean ± SEM) days after administration of the neurotoxin, as described previously (Clarke et al., 2007). Tissue samples were homogenized in 200 μl 0.2 M perchloric acid for 1.5 min and centrifuged at 6000 rpm for 20 min at 4°C. The supernatant (75 μl) was subsequently analyzed using HPLC as described above.

Statistics

Behavioral, D2/3 binding and DA-depletion data were analyzed using R (http://www.R-project.org/) and SPSS (IBM). For ANOVA, homogeneity of variance was verified using Levene's test; type III sums of squares and full factorial models were used unless stated. For designs with within-subjects factors, where applicable, the Huynh–Feldt correction was used to correct for any violations of the sphericity assumption as assessed by the Greenhouse–Geisser test. Computational model parameters were estimated using a hierarchical Bayesian analysis. Rather than confidence intervals, this produces credible intervals, specifically highest posterior density intervals (HDI). An *x*% HDI is the narrowest interval containing *x*% of the posterior probability mass. For example, if the 50% HDI for a parameter excludes zero, then it is more likely than not that the parameter is non-zero; if the 95% HDI excludes zero, then the probability that the parameter is non-zero exceeds 0.95. A 95% HDI excluding zero is, therefore, in general better evidence for a parameter being non-zero than a 95% confidence interval, which merely describes the likelihood of the data given the null hypothesis.

Computational modeling of behavior

We analyzed behavior in several ways, including the fitting of several computational models of reinforcement learning to the behavioral data. We aimed to address several behavioral possibilities:

(1) We analyzed behavior according to the reinforcement occurring on the immediately preceding trial, in a win-stay/lose-shift analysis, as is common (den Ouden et al., 2013). (2) The analyses in (1) indicated that OFC-depleted and control groups differed in their response to reinforcement veracity (whether reinforcement on the preceding trial was "true," meaning in the majority, or "false," meaning in the minority and misleading as to the best stimulus). This suggests an effect of prior history, so we examined the dependence of choice on preceding reward/punishment, and also on subjects' prior stimulus choices (to account for stimulus bound perseveration) in terms of several preceding trials, using an n-back analysis with a family of conditional logit regression models (Lau

and Glimcher, 2005; Seymour et al., 2012). However this family of models did not explain the group differences in the win-stay/lose-shift analysis, and even the best of them provided a poor description of behavior (as judged by the Bayesian Information Criterion; BIC) compared with state-based reinforcement learning models, considered below. We do not present full n-back analyses for reasons of space. (3) We considered the possibility that subjects used "model-based" (declarative) learning (Wunderlich et al., 2012), such as tracking reinforcement probabilities and their certainties about those probabilities in a Bayesian or similar fashion, and altering their estimates of probability less when their certainty is already high. (4) We considered a family of conventional value-/state-based ("model free") reinforcement learning rules, in which subjects update simple representations of their environment after each trial.

Model fitting and comparison

Likelihood calculation and maximum a posteriori fitting. Several regression and reinforcement learning (RL) models were compared. Each applies its own algorithm, with a certain number of parameters, to the sequence of stimuli and rewards experienced by the subjects. Sessions were treated as contiguous. In all cases, the model M, having parameters θ, calculated the probability of choosing each possible action (i.e., of selecting each of two given stimuli). The vector of actions actually chosen by subject s was denoted a_s, or $a_{s,t}$ at each trial t. The model's performance was evaluated by calculating the likelihood function P(choice actually made $\mid M, \theta$) for each trial. The log-likelihood (LL) was calculated as follows:

$$\ln(\text{likelihood}) = \text{LL} = \sum\nolimits_t \ln(P(a_{s,t}|M)).$$

We conducted maximum a posteriori (MAP) fitting using priors as follows (Daw, 2011): learning rates and other parameters that are in principle constrained in the range (0,1) were given priors of Beta(1.1, 1.1), whereas softmax inverse temperatures (β) and "stickiness" maxima (see below) were given priors of Gamma(shape = 1.2, scale = 5).

We selected model parameters $\hat{\theta}_M$ for a given model M to maximize the probability of obtained data D given the model and its parameters:

$$P(D|M,\hat{\theta}_M)\, P(\hat{\theta}_M|M),$$

by maximizing

$$\ln(P(D|M,\hat{\theta}_M)) + \ln(P(\hat{\theta}_M|M)),$$

for each subject using the optim() function of *R* (R Core Team, 2012). Logarithms are to base *e* throughout.

Model selection. Models were selected using the Bayesian information criterion $\text{BIC} = -2\,LL + k\ln(n)$, where k is the number of parameters in the model and n is the number of observations (trials; Schwartz, 1978; Burnham and Anderson, 2004). Lower BIC values indicated a better fit after penalization for the number of parameters. The BIC was computed across all subjects, such that $k = zs$, where s is the number of subjects and z is the number of parameters per subject entering the reinforcement learning model. This method gives more weight to subjects contributing more trials, but correctly so in terms of optimizing the overall fit, because such subjects contribute more information about the common model identity. There were no major differences if the corrected Akaike information criterion was used instead: $\text{AIC}_c = [2\,k - 2\,LL] + [2\,k\,(k+1)/(n-k-1)]$.

Reference model BIC. We included a model choosing at random ($p = 0.5$ for each trial, for $n = 4814$ trials) for comparison of BIC values.

Exceedance probability. Following MAP estimation, we also calculated the model that was most likely (across all subjects) based on the random-effects analysis of Stephan et al. (2009), which treats the model identity as a random variable.

Parameter comparison. For some models, we compared model parameters across groups using summary statistics (Daw, 2011). Because the number of trials varied by subject, in some cases we also compared model fits by comparing the mean LL per trial, calculated for each subject, across groups. For the best model, we estimated parameters and group differences using a full Bayesian hierarchical method, described below.

Optimal Bayesian choice algorithm

A hypothetical ideal subject would estimate the probability of reinforcement for each stimulus, represent its uncertainty about those estimates, and choose so as to maximize the reward obtained. We modeled this behavior using an optimal Bayesian method.

The probability of reward for each stimulus was represented by a probability density function (PDF) for each stimulus. The prior PDF was uniform (that is, before a discrimination begins, a subject is assumed to believe equally strongly that a given stimulus will always deliver reward, never deliver reward, deliver reward with a 40% probability, and so on). For a uniform prior, the posterior probability density function after T trials with r rewards and $s = T - r$ punishments is given by the following:

$$f(p) = \frac{1}{B(r+1,\, s+1)}\, p^r(1-p)^s,$$

where B() is the β function. The probability of choosing a stimulus was determined by randomized probability matching (RPM; Scott, 2010). In RPM, an agent selects a series of actions a_t at time t, and observes a sequence of rewards $y_t = (y_1, \ldots, y_t)$. For our purposes, reward occurs or does not occur on each trial, so $y_t \in \{0,1\}$. Each action generates reward independently from the reward distribution $f_{a_t}(y \mid \theta)$, where θ is an unknown parameter vector; for our purposes, $f_a(y \mid \theta)$ is the Bernoulli distribution with success probability θ_a. The quantity $\mu_a(\theta) = E(y_t \mid \theta, a_t = a)$ is the expected reward from $f_a(y \mid \theta)$. If θ were known, the optimal strategy would be to choose the option with the largest $\mu_a(\theta)$. RPM calculates the quantity:

$$w_{at} = P(\mu_a = \max\{\mu_1,\ldots,\mu_k\}|\mathrm{y_t}),$$

and allocates choice $t+1$ to option a with probability w_{at}. When rewards are drawn from independent Bernoulli random variables (a "binomial bandit"), as in the current situation, the optimality probability (Scott, 2010) is given by:

$$w_{at} = \int_0^1 \text{Be}(\theta_a|Y_{at}+1,\, T_{at}-Y_{at}+1)\prod\nolimits_{j\neq a} P(\theta_j < \theta_a|Y_{jt}+1,\, T_{jt}-Y_{jt}+1)\mathrm{d}\theta_a,$$

calculated across the actions on offer, where $\text{Be}(\theta \mid \alpha, \beta)$ is the density of the β distribution for a random variable θ with parameters α and β, and Y_{at} and T_{at} are the cumulative number of successes and trials respectively observed for action a up to time t.

RPM has no parameters and therefore requires no fitting. We used this model in isolation, but also added a softmax stage:

$$P(a,t) = \text{softmax}_\beta^a\,(w_{1,t},\ldots,w_{k,t}).$$

Value-/state-based RL models

Delta-rule updates of stimulus value. Simple value-based RL algorithms assign a value to each stimulus or action, and choose accordingly; the values are updated according to rules and parameters determining the impact of reward or punishment, but (unlike models such as RPM) they do not represent the statistical structure of the environment in a more complex way. Subjects' behavior was modeled using a delta-rule update function that allowed different speeds of response to reward (τ_r) and to punishment (τ_p):

$$\text{Prop}(x,a,b) = (1-x)a + xb$$

$$V_{t+1}^i \leftarrow \begin{cases} \text{Prop}(\tau_r, V_t^i, 1) \text{ when action } i \text{ chosen and rewarded on trial } t \\ \text{Prop}(\tau_p, V_t^i, 0) \text{ when } i \text{ chosen and punished on trial } t \\ V_t^i \text{ when not chosen/not presented} \end{cases}.$$

For each subject, the eight stimuli presented in discriminations D5–D8 were each assigned a value, which was updated according to this rule. The initial values of all stimuli were set to 0.5, midway between the target value for reward (1) and punishment (0).

In other model variants, the constraint $\tau_r = \tau_p = \tau_{rp}$ was applied.

Stickiness. In a subset of models, the tendency to repeat choices (Lau and Glimcher, 2005; Seymour et al., 2012) was modeled using additional parameters τ_c and c:

$$C^i_{t+1} \leftarrow \begin{cases} \text{Prop}(\tau_c, C^i_t, c) \text{ when action } i \text{ chosen on trial } t \\ \text{Prop}(\tau_c, C^i_t, 0) \text{ when not chosen/not presented} \end{cases}.$$

The initial C value for each stimulus was set to 0.

In variants, the constraint $\tau_c = \tau_r$ or the constraint $c = 1$ was applied.

Side bias. In some models, one of two possible sources of side bias was included. In one, based on the subject's own behavior only, the left side was favored as a result of previous choices to the left side:

$${}^{C}L^i_{t+1} \leftarrow \begin{cases} \text{Prop}(\tau_{LC}, {}^{C}L^i_t, d_{LC}) \text{ when left chosen on trial } t \\ \text{Prop}(\tau_{LC}, {}^{C}L^i_t, -d_{LC}) \text{ when right chosen} \end{cases}.$$

In the other, based on reinforcement, the left side was favored as a result of previous reinforcement following choice of the left side, or punishment following choice of the right side:

$${}^{R}L^i_{t+1} \leftarrow \begin{cases} \text{Prop}(\tau_{LR}, {}^{R}L^i_t, d_{LR}) \text{ if (left chosen and rewarded, or} \\ \quad \text{right chosen and punished) on trial } t \\ \text{Prop}(\tau_{LR}, {}^{R}L^i_t, -d_{LR}) \text{ otherwise} \end{cases}.$$

Bias values were initially set to 0.

Softmax. The probability of responding was calculated according to a softmax rule, applied only across the two stimuli presented on each trial ($n = 2$):

$$P(i,t) = \frac{e^{\beta(V^i_t + C^i_t + L^i_t)}}{\sum_{k=1}^{n} e^{\beta(V^k_t + C^k_t + L^k_t)}}.$$

The softmax (soft maximization) function takes a number of inputs and provides the same number of outputs. The outputs sum to 1. The largest input produces the largest output (maximization) and the proportion of the output captured by the largest input is determined by the softmax parameter (soft maximization rather than hard, or absolute, maximization). It has a variable inverse temperature β (low β, or high temperature, leads to nearly equiprobable actions). The use of β rather than temperature $1/\beta$ is for computational reasons, to avoid division by zero following underflow.

In variants, the constraint $\beta = 1$ was used. (This constraint was always used when τ_r and τ_p were separate parameters, since to vary it would have been confounded with the difference between τ_r and τ_p and would lead to over-parameterization, for example, for a given τ_r, either an increase in β or an increase in τ_p would lead to exaggerated preferences between any pairs of stimuli; thus, β and τ_p would be negatively coupled.)

Simple approximation to stimulus reliability in predicting reinforcement

We also tested models that calculated the total number of trials T_{at} and the total number of rewards Y_{at} for each action a up to time t, as for RPM, but updated values according to the simpler rules:

$$r_{it} = Y_{it}/T_{it}$$

$$r_{it} \leftarrow \begin{cases} Y_{it}/T_{it} \text{ when } T_{it} > 0 \\ 0.5 \text{ when } T_{it} = 0 \end{cases}$$

$$V^i_{t+1} \leftarrow \begin{cases} \text{Prop}(\text{Prop}(\rho_r, \tau_r, r_{it}\tau_r), V^i_t, 1) \text{ when action } i \text{ chosen} \\ \quad \text{and rewarded on trial } t \\ \text{Prop}(\text{Prop}(\rho_r, \tau_p, (1 - r_{it})\tau_p), V^i_t, 0) \textit{ when } i \text{ chosen} \\ \quad \text{and punished on trial } t \\ V^i_t \text{ when not chosen/not presented} \end{cases}.$$

In this model, subjects weight the effect of reward or punishment by the "reliability" measure r, and have a fixed propensity (ρ_r, ρ_p) to do so. This reduces to the previous model of value updating when $\rho_r = \rho_p = 0$. As an example, in the case where $\rho_r = \rho_p = 1$, this model would weight the effect of reward by 0.8 for an action that had been rewarded on 80% of previous trials, and weight the effect of punishment by 0.2 for the same action.

Hybrid models incorporating Bayesian and value-based elements

Finally, we created models that blended simple value-based and optimal Bayesian responding. We calculated decision probabilities based on the simple delta-rule models, and separately according to RPM (with or without an RPM softmax stage), with a further parameter for each subject, $0 \leq w \leq 1$, such that its decision probability for each action at each time point was $wp_{\text{rpm}} + (1 - w)p_{\text{delta}}$.

The complete set of reinforcement learning models tested is shown in Table 1.

Hierarchical Bayesian modeling of the optimal RL model

The best model (Delta1C-LC) was subjected to a full hierarchical Bayesian analysis using Stan (Stan Development Team, 2014), with the following parameters: (1) a shared group SD for each parameter: these had a prior distribution of the half-Cauchy(0, 5) distribution and constraints of $[0,+\infty)$; (2) a per-group mean for each parameter: group mean values of τ_{rp}, τ_c, τ_{LC} had prior distributions of Beta(1.1, 1.1) and were constrained to the range [0,1]; group mean values of c, d_{LC} had prior distributions of Gamma(shape = 1.2, scale = 5) and constraints of $[0,+\infty)$; (3) per-subject parameters, with similar constraints, drawn from normal distributions defined by the group-level parameters; (4) per-trial probabilities of choosing the best stimulus, calculated deterministically from the per-subject parameters according to the RL algorithm; and (5) actual choices assumed to be drawn from Bernoulli distributions defined by the per-trial probabilities.

The Stan software uses Hamiltonian Monte Carlo sampling, a form of Markov chain Monte Carlo sampling, to sample from the posterior distributions of the parameters. The values of primary interest were the posterior probability distributions of the differences in-group means.

To compare the model's predictions to the behavioral analysis of "obey" probabilities, the probabilities of obeying preceding feedback of different types were sampled from the best-fit computational model of behavior. Per-subject estimates were sampled of the mean probability (as determined by the model) of choosing an option that would correspond to "obeying," given the actual choice made and actual reinforcement obtained on the previous trial.

To compare the model's predictions to the behavioral analysis of errors to criterion, six "virtual" subjects chose according to the computational model and their mean posterior per-subject parameter values. For each subject, the model's probabilities of choosing the correct stimulus were converted into actual choices and fed into a virtual environment embodying a simple model of the task (in which the probability of valid reinforcement was 0.8, the probability of the correct stimulus being on the left or right was 0.5, using sessions of 30 trials each, with a stopping criterion of 90% correct in a session just as for the monkeys). Reinforcement from the virtual environment was fed back into the model, to update its state for the next trial. The mean number of errors to criterion was measured, across 1000 iterations of the task, for each virtual subject.

To establish the necessity and sufficiency of model parameter changes to cause behavioral effects on errors to criterion and obey probabilities, multiple ($n = 1000$) virtual subjects per group were similarly simulated, with either all parameters varying (all subjects taking their group mean value for each parameter), one parameter set varying [either the reinforcement rate parameter (τ_{rp}), the stimulus stickiness parameters (τ_c and c), or the side stickiness parameters (τ_{LC} and d_{LC}) varied between groups, with all other parameters taking the mean overall values, or two parameter sets varying and the remaining parameter taking the overall mean values.

Results

Orbitofrontal DA depletion increased caudate extracellular DA and reduced caudate dopamine D2 receptor availability

OFC DA-depleted monkeys exhibited a significant reduction in ^{18}F-fallypride D2RB compared with controls, which was localized to the caudate nucleus ($t_{(5)} = 3.616$, $p = 0.015$) (Fig. 2*A*,*B*)

Table 1. Summary of computational models tested

Model name	Description	Parameters	BIC (control)	Rank (control)	BIC (OFC)	Rank (OFC)	BIC (all)	Rank (all)
Random choice (reference model)								
Random	Random choice, $p = 0.5$	None	4400	32	2274	32	6674	32
Bayesian								
PureRPM	Randomized probability matching	None	36469	33	6118	33	42587	33
RPMSoftmax	Softmax of the RPM optimal probabilities	β_{rpm}	3798	17	1716	15	5518	17
Value-based								
Delta1	Delta rule with identical learning rate for reward and punishment	τ_{rp}	4106	28	2010	27	6121	28
Delta1S	Delta1 with inverse softmax temperature	τ_{rp}, β	3991	26	1850	26	5850	25
Delta2	Delta rule with learning rates for reward and punishment	τ_r, τ_p	4161	31	2028	28	6199	31
Delta1C	Delta rule with a single reinforcement rate parameter, plus dependence of choice on previous choice (stimulus stickiness)	τ_{rp}, τ_c, c	3504	5	1638	1	5156	4
Delta1CS	As for Delta1C, with a variable softmax temperature	τ_{rp}, τ_c, c, β	3540	10	1671	8	5229	9
Delta2C	Delta rule with learning rates for reward and punishment, plus dependence of choice on previous choice (stimulus stickiness)	τ_r, τ_p, τ_c, c	3535	9	1659	4	5212	8
Delta1CM	As for Delta1C, with τ_c constrained	τ_{rpc}, c	4143	30	2045	31	6197	30
Delta1CR	As for Delta1C, with c constrained	τ_{rp}, τ_c	4130	29	2032	29	6172	29
Delta1C-LR	Delta1C plus side bias based on reinforcement	τ_{rp}, τ_c, c, τ_{LR}, d_{LR}	3559	12	1663	6	5245	11
Delta1C-LC	**Delta1C plus side bias based on choice (winning model overall)**	$\boldsymbol{\tau_{rp}}$, $\boldsymbol{\tau_c}$, $\boldsymbol{c}$, $\boldsymbol{\tau_{LC}}$, $\boldsymbol{d_{LC}}$	**3371**	**1**	**1639**	**2**	**5033**	**1**
Delta1C-LC-SSMAX	As for Delta1C-LC, with τ_c constrained	τ_{rp}, τ_c, c, d_{LC}	3541	11	1673	9	5232	10
Delta1C-LC-SSRATE	As for Delta1C-LC, with d_{LC} constrained	τ_{rp}, τ_c, c, τ_{LC}	3529	7	1660	5	5207	6
Delta1-LC	Delta1 plus side bias based on choice	τ_{rp}, τ_{LC}, d_{LC}	3980	25	2040	30	6033	27
Delta2C-LC	Delta2C plus side bias based on choice	τ_r, τ_p, τ_c, c, τ_{LC}, d_{LC}	3420	3	1665	7	5112	2
Delta1C-LC-S	Delta1C-LC plus variable softmax temperature	$_{\tau rp}$, τ_c, c, τ_{LC}, d_{LC}, β	3412	2	1678	10	5117	3
Reliability								
ReliabilityS	Reward and punishment rates, dependence of reward/punishment update on the reliability of stimuli, and a softmax parameter	τ_r, τ_p, ρ_r, ρ_p, β	4054	27	1843	25	5920	26
ReliabilityC-LC-S	As for ReliabilityS, adding stimulus stickiness and side stickiness	τ_r, τ_p, ρ_r, ρ_p, τ_c, c, τ_{LC}, d_{LC} β	3505	6	1738	19	5284	13
Hybrid between RPM and delta rules								
RPMDelta1	Weighted combination of Delta1 and RPM	τ_{rp}, w	3801	18	1713	14	5523	18
RPMDelta1S	Weighted combination of Delta1S and RPM	τ_{rp}, β, w	3844	21	1739	20	5597	21
RPMDelta2	Weighted combination of Delta2 and RPM	τ_r, τ_p, w	3834	20	1732	18	5579	19
RPMDelta1C	Weighted combination of Delta1C and RPM.	τ_{rp}, τ_c, c, w	3534	8	1658	3	5210	7
RPMDelta1CS	Weighted combination of Delta1CS and RPM	τ_{rp}, τ_c, c, β, w	3563	13	1691	11	5276	12
DeltaRPMSoftmax	Weighted combination of Delta1 and RPM, with a single softmax function applied to the combination	τ_{rp}, β, w	3828	19	1748	21	5589	20
SRPMDelta1	Weighted combination of Delta1 and RPMSoftmax.	τ_{rp}, w, β_{rpm}	3847	22	1759	22	5620	22
SRPMDelta1S	Weighted combination of Delta1S and RPMSoftmax	τ_{rp}, β, w, β_{rpm}	3883	24	1790	24	5691	24
SRPMDelta2	Weighted combination of Delta2 and RPMSoftmax	τ_r, τ_p, w, β_{rpm}	3871	23	1782	23	5671	23
SRPMDelta1C	Weighted combination of Delta1C and RPMSoftmax	τ_{rp}, τ_c, c, w, β_{rpm}	3572	14	1696	12	5290	14
SRPMDelta1CS	Weighted combination of Delta1CS and RPMSoftmax	τ_{rp}, τ_c, c, β, w, β_{rpm}	3600	15	1728	17	5355	16
SRPMDelta2C	Weighted combination of Delta2C and RPMSoftmax	τ_r, τ_p, τ_c, c, w, β_{rpm}	3602	16	1721	16	5350	15
SRPMDelta1C-LC	Weighted combination of Delta1C-LC and RPMSoftmax	τ_{rp}, τ_c, c, τ_{LC}, d_{LC}, w, β_{rpm}	3428	4	1706	13	5165	5

Their performance was assessed by the BIC. Low BIC values indicate a better fit, having penalized the models for their complexity. BIC values and corresponding model ranks (1 being best) are shown for sham-operated control subjects, for OFC DA-depleted subjects, and for all subjects together. Ranking by BIC across all subjects gives more weight to subjects contributing more trials, which is correct in terms of optimizing the overall fit, because such subjects contribute more information about the common model identity. The best model overall (Delta1C_LC) was also the best model for control subjects. It was the second-best model for lesioned subjects with a BIC difference of only 1 from the best model for lesioned subjects (Delta1C), and it differed structurally from that model only in parameters whose values were demonstrably different in the OFC DA-depleted group (Fig. 5*Q*). β, Softmax inverse temperature; τ, rate constant; τ_r, learning rate for reward; τ_p, learning rate for punishment; τ_{rp}, learning rate for reinforcement whether reward or punishment; τ_c, learning rate for stimulus stickiness; τ_{LC}, learning rate for side stickiness; τ_{LR}, learning rate for side bias based on reinforcement; c, maximum for stimulus stickiness; d_{LC}, maximum for side stickiness; d_{LR}, maximum for side bias based on reinforcement; ρ, propensity to weight reward or punishment by its reliability; w, fraction of decision making based on Bayesian (vs delta rule) processes.

and did not extend to the putamen ($t_{(5)} = 2.012$, $p = 0.1$), nucleus accumbens ($t_{(5)} = -0.38$, $p = 0.971$), OFC ($t_{(5)} = 0.589$, $p = 0.581$), ventrolateral PFC ($t_{(2.075)} = 0.737$, $p = 0.535$), amygdala ($t_{(5)} = 0.815$, $p = 0.452$), or anterior hippocampus ($t_{(5)} = 1.452$, $p = 0.206$). Subsequent analysis of caudate subregions revealed significant reductions in D2RB in the ventromedial caudate ($t_{(5)} = 2.772$, $p = 0.039$) and the body of the caudate nucleus ($t_{(5)} = 2.497$, $p = 0.03$), with a nonsignificant reduction in the dorsolateral caudate ($t_{(5)} = 2.541$, $p = 0.052$). These effects are likely mediated primarily by D2 receptors, as the caudate nucleus is an area of low D3 receptor expression in the rat (Bouthenet et al., 1991), although the situation in primates is less clear. The lack of a reduction in D2/3 binding in the OFC after the OFC-DA depletion suggests that the residual dopamine is sufficient to occupy OFC D2 receptors at presurgical levels. Indeed, there are relatively few D2/3 receptors in the OFC compared with D1 receptors which further suggests that the upregulation of DA in the caudate is mediated via a D1 pathway in the OFC (Lidow et al., 1991).

A potential cause of this decreased caudate D2RB was competition from increased extracellular DA that reduced radioligand binding. To assess this hypothesis, all subjects underwent caudate microdialysis. Extracellular DA levels were measured at baseline and following 75 mM K^+. These measures are taken to reflect tonic and phasic DA release, respectively as the influx of K^+ mimics the arrival

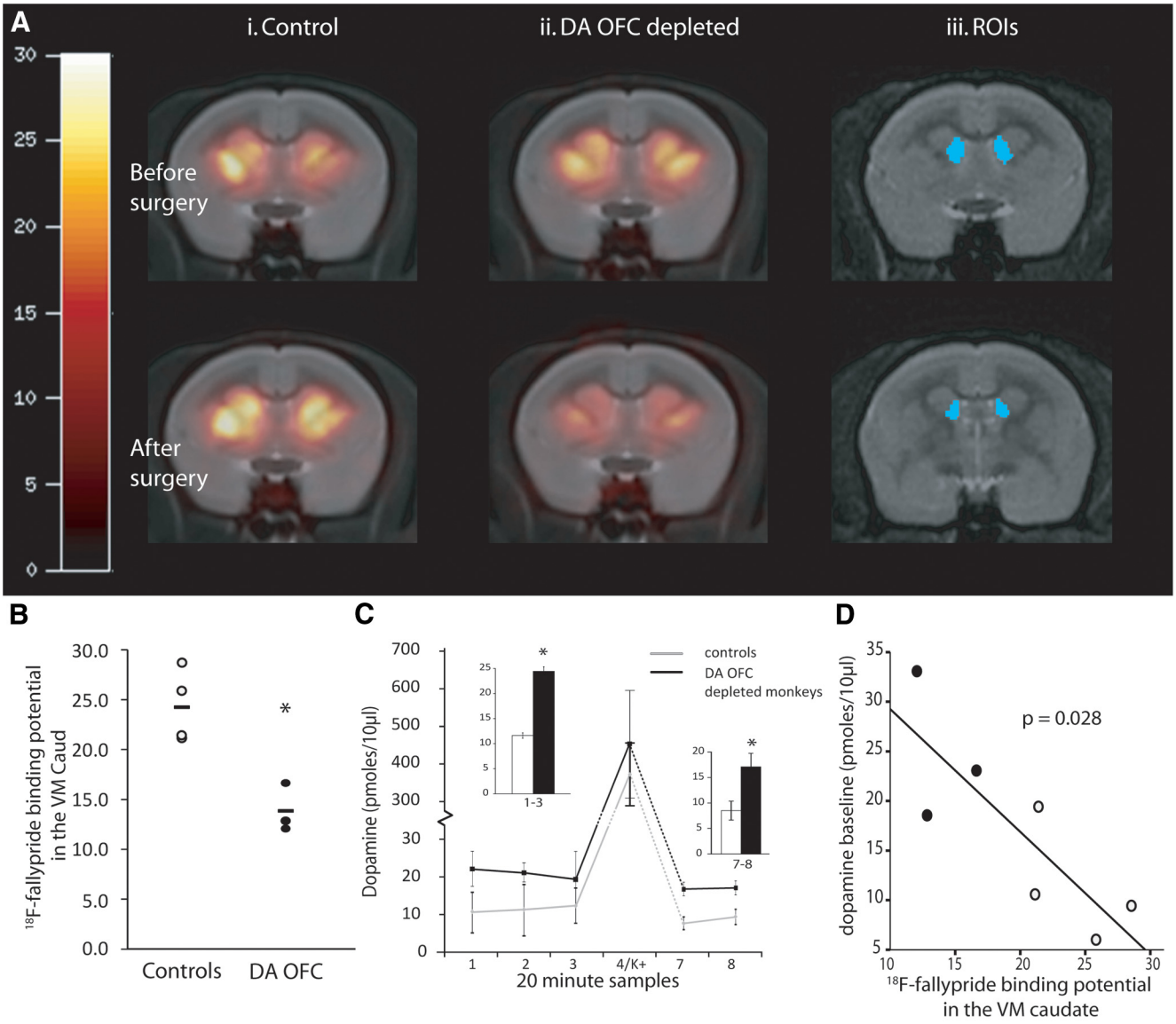

Figure 2. Increased ventromedial caudate dopamine levels following 6-OHDA lesions of the OFC, shown by reduced binding potential of the selective D2/3 receptor antagonist ^{18}F-fallypride and increased baseline extracellular DA levels. ***A***, Coronal coregistered MRI and PET images of ^{18}F-fallypride nondisplaceable binding potential (BP_{ND}) in the striatum of a control animal (***Ai***), and an OFC DA-depleted animal (***Aii***), before and after surgery, depicted at coronal section AP +10. ***Aiii***, Ventromedial caudate ROIs (blue) are depicted at coronal section AP +10 (top) and AP +9.75 (bottom); BP_{ND} threshold = 30. ***B***, Significantly reduced ^{18}F-fallypride BP_{ND} in the ventromedial caudate of OFC DA-depleted monkeys, compared with controls (*$p < 0.05$). ***C***, Significantly increased baseline extracellular DA in the ventromedial (VM) caudate both before and after the K^+ challenge of OFC DA-depleted monkeys, measured by microdialysis (insets: samples 1–3, $p = 0.041$; samples 7–8, $p = 0.04$). ***D***, Extracellular DA levels before K^+ challenge correlated negatively with ^{18}F-fallypride BP_{ND} in the ventromedial caudate ($p = 0.028$; filled circles, OFC DA-depleted group; open circles, controls).

of an action potential and thus induces the release of DA in a phasic manner. Consistent with this hypothesis, OFC DA-depleted monkeys showed a significant increase in baseline DA compared with controls (samples 1–3, $t_{(5)} = -2.745$, $p = 0.041$; Fig. 2*C*) that was maintained after the K^+ challenge (samples 7–8, $p = 0.04$). However, differences in DA release in response to K^+ challenge were not seen ($t_{(5)} = -0.410$, $p = 0.699$), suggesting that OFC DA only modulates tonic rather than evoked striatal DA release.

Furthermore, the extent of caudate extracellular DA release correlated negatively with the reduced levels of ^{18}F-fallypride binding seen in the ventromedial caudate of the OFC DA-depleted monkeys ($r = -0.807$, $p = 0.028$; Fig. 2*D*). These findings are consistent with competition between endogenous extracellular DA and ^{18}F-fallypride binding and demonstrate that OFC DA dysfunction modulates caudate DA levels.

Postmortem analysis at ~448 d after surgery, confirmed that injection of 6-OHDA caused a significant DA depletion (45 ± 7.9%) in the OFC compared with controls ($t_{(3)} = 4.27$, $p = 0.024$; Table 2). Because our previous work has shown that these OFC DA depletions show considerable recovery over time, we also analyzed the time course of DA OFC depletion obtained from our analysis of identical lesions from previous studies at earlier timepoints (81% depletion at 16 d, 75% depletion at 84 ± 3 d, and 51% depletion at 370 ± 23 d). Thus, the time period during which the behavioral analysis, imaging, and microdialysis were performed corresponds to periods of very high (in excess of 70%) OFC DA depletion (Fig.3). 5-HT levels were unaffected, and although the medial prefrontal and OFC/lateral PFC also showed significant depletions of DA and NA respectively (medial PFC DA, $t_{(2)} = 11.063$, $p = 0.008$; OFC NA, $t_{(3)} = 11.145$, $p = 0.002$;

Table 2. Tissue monoamine levels assessed postmortem by HPLC

	Control monoamine levels (pmol/mg)			Decrease in OFC DA-depleted monkeys (%)		
	DA	5-HT	NA	DA	5-HT	NA
OFC	0.33 ± 0.06	0.83 ± 0.32	1.05 ± 0.12	45.5 ± 7.9*	10.53 ± 20.9	48.4 ± 5.3*
Medial PFC	0.35 ± 0.01	0.89 ± 0.21	1.25 ± 0.06	59.1 ± 2.38*	8.29 ± 23.1	28.4 ± 12.5
Lateral PFC	0.29 ± 0.03	0.69 ± 0.15	0.99 ± 0.06	18.18 ± 11.4	15.9 ± 26.04	36.6 ± 11.5*
Dorsal PFC	0.51 ± 0.06	0.51 ± 0.13	1.24 ± 0.10	10.3 ± 14.3	11.6 ± 18.13	28.13 ± 10.1
Motor cortex	0.44 ± 0.05	0.55 ± 0.11	1.44 ± 0.03	14.1 ± 24.4	8.3 ± 10.21	9.901 ± 7.4
Amygdala	6.33 ± 1.68	2.69 ± 0.66	1.76 ± 0.36	35.9 ± 8.14	18.3 ± 23.01	35.4 ± 13.8
Nucleus Accumbens	27.7 ± 7.29	2.55 ± 1.16	4.69 ± 1.11	18.6 ± 14.03	−18.6 ± 35.01	33.3 ± 12.0
Caudate: medial head	50.06 ± 8.6	0.91 ± 0.12	0.37 ± 0.11	−29.7 ± 21.9	−31.3 ± 42.7	38.4 ± 10.6
Caudate: lateral head	51.2 ± 6.23	0.99 ± 0.28	0.38 ± 0.07	−5.88 ± 20.3	−2.39 ± 11.8	56.5 ± 5.55
Caudate: body	70.3 ± 8.9	1.15 ± 0.32	0.37 ± 0.11	−7.38 ± 4.3	1.505 ± 5.2	28.7 ± 19.4
Putamen	69.3 ± 10.8	1.43 ± 0.36	0.52 ± 0.07	0.43 ± 4.22	−5.52 ± 11.7	−1.36 ± 12.6

Absolute levels of DA, 5-HT, and NA in the striatum, amygdala, and anterior cortices of sham-operated control marmosets (pmol/mg, mean ± SEM), and corresponding decreases in marmosets with 6-OHDA infusions into the OFC (percentage decrease from the sham-operated control group, mean ± SEM; negative changes indicate percentage increases); *$p < 0.05$ for the group difference (from direct comparison of raw values from sham-operated and 6-OHDA-lesioned groups).

lateral PFC NA, $t_{(3)} = 3.892$, $p = 0.03$), these depletions are not apparent at any earlier time points, suggesting that they may be due to compensatory processes that occur later than our period of interest.

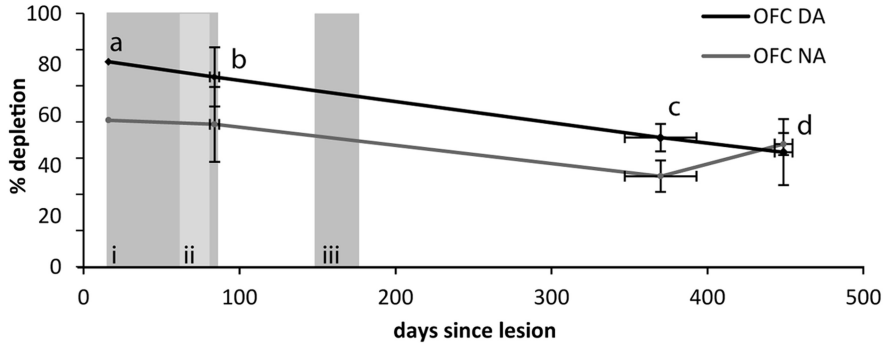

Figure 3. Postmortem depletions of DA and NA in the OFC as a function of time since surgery in OFC-depleted monkeys. The gray regions indicate the time periods in which the behavior (***i***), the second MRI/PET scan (***ii***), and *in vivo* microdialysis (***iii***) were completed by the DA OFC depleted monkeys. Each gray region extends from the earliest starting point to the latest endpoint (and for behavior, their edges represent the monkeys that were the "fastest" and "slowest" to complete the discriminations). All three components of this study therefore occurred during high levels of OFC DA depletion. ***a***, One DA OFC-depleted animal, 16 d after surgery; ***b***, four DA OFC-depleted animals averaging 84 d after surgery (Clarke et al., 2007); ***c***, Four DA OFC-depleted animals averaging 370 d after surgery (Walker et al., 2009); and ***d***, current behavioral study.

OFC-DA depletion improved overall behavioral performance and resulted in a decreased sensitivity to false punishment

To assess how behavior was altered by the OFC-DA depletion we focused not only on overall errors to criterion but also how the feedback on the immediately preceding trial impacted upon stimulus selection on the current trial, using a win-stay/lose-shift analysis approach. The latter approach has frequently been used (Waltz and Gold, 2007; den Ouden et al., 2013) to reveal how sensitivity to positive or negative feedback governs subsequent behavioral choice (the premise being that positive feedback should lead to repeated choice of a given stimulus, whereas negative feedback should lead to a shift in choice to an alternate stimulus) and has been successfully used to reveal differences in reinforcement learning in schizophrenia (Waltz et al., 2007). We therefore defined "shifting" as choosing a different stimulus to that chosen on the previous trial and "staying" as its converse, and calculated the probability of obeying reinforcement (staying after reward and shifting after punishment). We analyzed responding on trial X according to (1) valence: whether the response on trial X − 1 was rewarded or punished, and (2) veracity: whether that reinforcement was true (majority; e.g., reward following selection of the "correct" stimulus) or false/misleading (minority; e.g., punishment following selection of the correct stimulus).

Presurgical performance

There were no between-group differences in either errors to criterion or in any win-stay/lose-shift parameters (D1–D4, all $p >$ 0.05, NS; Fig. 4).

Postsurgical performance

OFC DA-depleted monkeys were faster to learn, i.e., made fewer errors to criterion, than sham-operated controls (mean D5–D8, $p = 0.05$) compared with their presurgical performance (Fig. 5*A*).

Preliminary win-stay/lose-shift analysis established that no term involving the reinforcement probabilities (80:20 vs 70:30) was significant ($F \leq 1.34$, $p \geq 0.311$), so this term was dropped from subsequent analyses.

OFC DA-depleted subjects were less likely to obey false punishment (Fig. 5*B*). There was a three-way interaction between lesion, valence, and veracity ($F_{(1,4)} = 9.10$, $p = 0.0392$). This interaction was analyzed further by considering reward and punishment separately. For reward, there was no effect of lesion and no lesion × veracity interaction ($F < 1$, NS). For punishment, the groups differed (lesion × veracity, $F_{(1,4)} = 8.93$, $p = 0.040$). Although the groups did not differ in their response to true punishment (lesion: $F_{(1,4)} = 1.75$, $p = 0.256$), OFC DA-depleted subjects were significantly less likely to obey false punishment (lesion: $F_{(1,4)} = 15.0$, $p = 0.0180$).

Behavior was best described by a simple computational model of reinforcement learning

To examine the behavioral strategy used by the subjects and to characterize this change better, several computational models of behavior were compared (Table 1). The best model for controls was one in which subjects' choices were governed by reinforcement, their own recent choice of stimulus, and their own recent choice of response side (Model Delta1C-LC; BIC 3371 across controls; Table 1). The best model for OFC DA-depleted subjects was either the same model (BIC 1639) or a similar model lacking

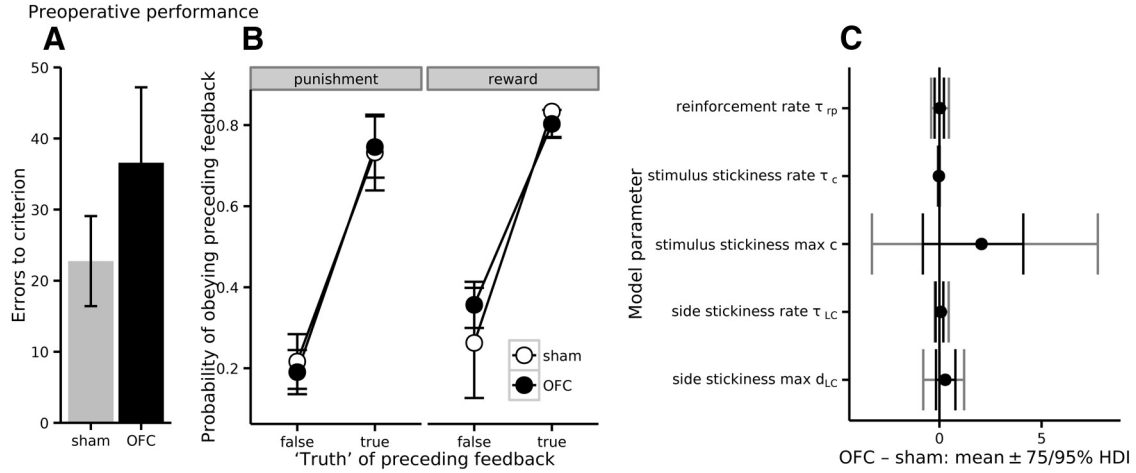

Figure 4. Preoperative behavioral performance. The monkeys did not show any group differences ($p > 0.05$) in either (***A***) their ability to learn the preoperative probabilistic discriminations (D1–D4; compare Fig. 5*A*), or (***B***) their win-stay/lose-shift behavior (last preoperative discrimination, D4; compare Fig. 5*B*). ***C***, Similarly, there were no group differences in parameters for the best computational model when fitted to preoperative discrimination D4 except a fractionally higher τ_c (75% HDI [0.0025, 0.051], 95% HDI [−0.019, 0.081]; HDIs for all other parameters included zero; compare Fig. 5*C*).

the dependence on their own recent choice of response side (Model Delta1C; BIC 1638).

Model Delta1C-LC also had the highest exceedance probability, at 0.705 (the probability that this model is more likely than any other model tested), and the lowest BIC across all subjects. Thus, this model was selected as the winner. It incorporated parameters for (1) sensitivity to reinforcement (τ_{rp}, rate), without the need for different response rates to reward and punishment; (2) stimulus stickiness, the tendency to repeat choices to stimuli that have been recently chosen (τ_c, rate; c, maximum effect relative to reinforcement); and (3) side stickiness, the tendency to repeat choices to the side (left vs right) that had been recently chosen (τ_{LC}, rate; d_{LC}, maximum effect relative to reinforcement). The stickiness parameters govern a process analogous to exploration versus exploitation strategies (Lau and Glimcher, 2005; Seymour et al., 2012).

OFC DA depletion increased response exploration and reinforcement sensitivity, as assessed by computational modeling of behavior

This model was a good descriptor of both groups and there were no preoperative group differences in its parameters. Postoperatively however, OFC DA-depleted animals exhibited alterations in both strategy and reinforcement-related behavior.

OFC DA-depletion made subjects less reliant on a strategy of reselecting a recently chosen side (Fig. 5*C*: strong evidence for reduced d_{LC} (the maximum for side stickiness; probability of non-zero difference = 0.051), with some evidence for increased τ_{LC} (learning rate for side stickiness; probability of non-zero difference = 0.938), indicating that their side stickiness strategy had a significantly lesser effect on behavior overall and altered rapidly). OFC DA-depleted animals also showed an enhancement of reinforcement sensitivity (Fig. 5*C*). The reinforcement rate parameter (τ_{rp}) had a posterior probability of 0.778 of being non-zero (being much stronger evidence for a difference than a frequentist p value of 0.222). There were no group differences in the parameters governing stimulus stickiness.

To assess the contribution of the alterations in strategy (reduced side stickiness) and the increased reinforcement sensitivity to these predictions independently, the analyses were replicated in simulations that allowed only subsets of the parameters to vary between groups (see Materials and Methods). This revealed that the reduced side stickiness did not impact upon win-stay/lose-shift behavior, but that changes in the overall sensitivity to reinforcement (regardless of whether it was reward or punishment) were necessary and sufficient to reproduce the reductions in both errors to criterion and sensitivity to false punishment shown by the OFC DA-depleted group behaviorally (Fig. 5*E*).

In summary, the OFC DA-depleted animals showed an increase in reinforcement sensitivity, and a decrease in side stickiness. Of these two changes, the increase in reinforcement sensitivity was responsible for the changes in errors to criterion and sensitivity to false punishment shown behaviorally.

There was no effect of OFC DA-depletion on response latencies. Latencies from postoperative discriminations (D5–D8) were analyzed using lesion (OFC/controls) and correct (correct/incorrect) as factors; there was no effect of either predictor and no interaction (lesion: $F_{(1,4)} = 2.09$, $p = 0.222$; other terms: $F < 1$, NS).

Behavioral changes correlated with D2 receptor binding in the caudate nucleus but not with OFC D2 receptor binding

Given the specificity of these changes in D2RB to the caudate nucleus, and previous evidence suggesting that D2RB in the caudate is implicated in the ability to shift responding to changing feedback (Groman et al., 2011), we investigated whether (1) the reduction in side stickiness, (2) increased reinforcement sensitivity, or (3) the resulting reduction in insensitivity to misleading negative feedback induced by OFC DA-depletion were related to changes in caudate DA and D2RB.

The reinforcement sensitivity parameter (τ_{rp}) correlated negatively with D2RB in the caudate nucleus (examined as an ROI: ventromedial caudate $r = -0.857$, uncorrected $p = 0.037$; dorsolateral caudate $r = -0.839$, $p = 0.029$; Fig. 6*A*), but not with D2RB in the OFC ($r = -0.207$, $p = 0.694$; Fig. 6*B*).

Side stickiness maximum (d_{LC}) was positively correlated with D2RB in the caudate nucleus (ventromedial caudate, $r = 0.887$, uncorrected $p = 0.018$; dorsolateral caudate, $r = 0.870$, $p = 0.024$ and caudate body, $r = 0.884$, $p = 0.019$; Fig. 6*C*) but not with D2RB in the OFC ($r = -0.020$, uncorrected $p = 0.971$; Fig. 6*D*). We did not examine the relationship with τ_{LC} (learning rate for side stickiness) as well, because τ_{LC} and d_{LC} were themselves

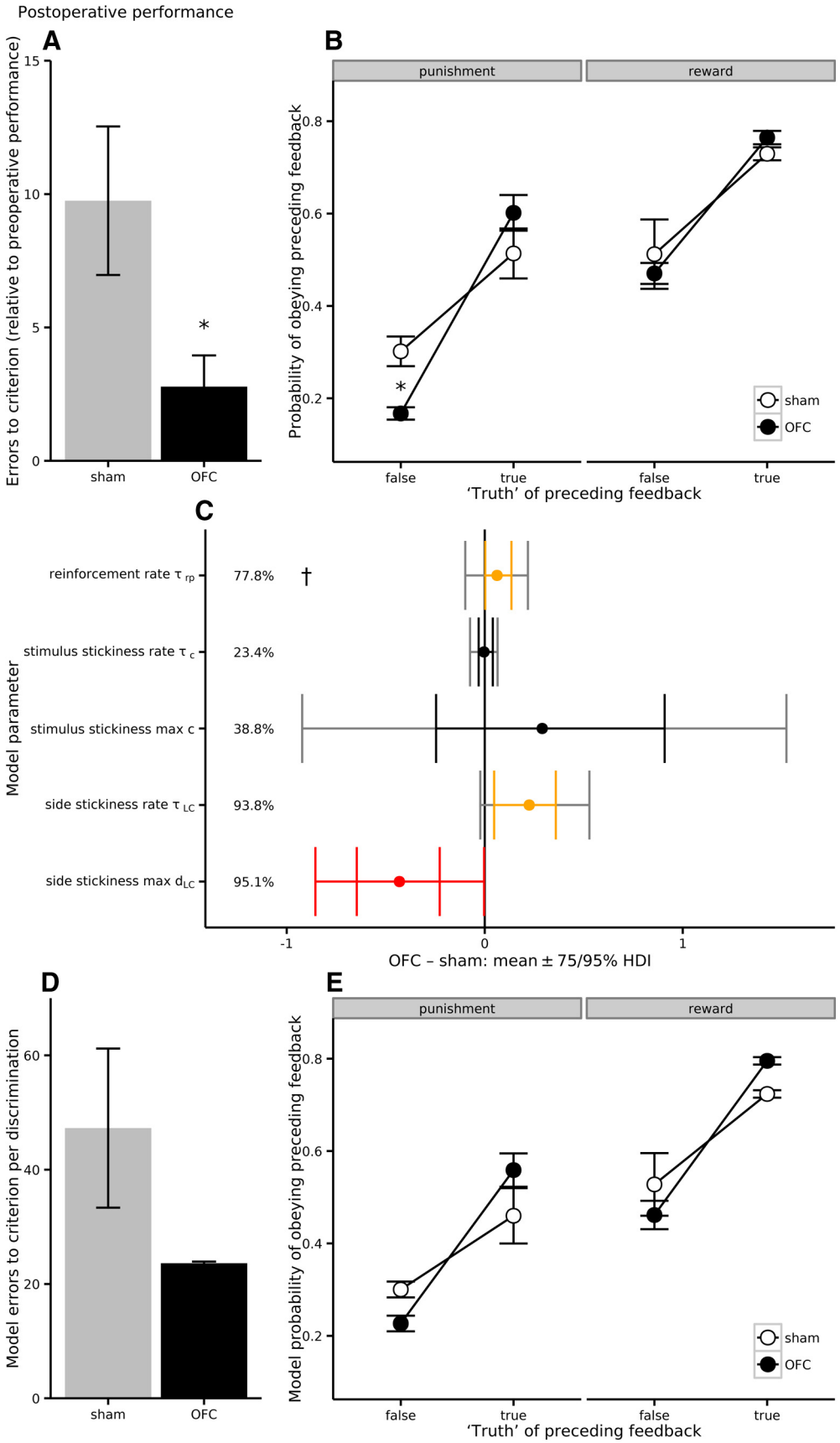

Figure 5. Behavioral performance. ***A***, Faster learning in OFC DA-depleted monkeys in a probabilistic visual discrimination learning task, with fewer errors to criterion compared with their preoperative performance ($p \leq 0.05$). ***B***, OFC DA-depleted monkeys showed intact reward-related behavior but a decreased probability of shifting their responding to the other stimulus after misleading (false) negative feedback (†$p = 0.018$). ***C***, The optimal computational model of behavior had parameters representing sensitivity to reinforcement (*Figure legend continues.*)

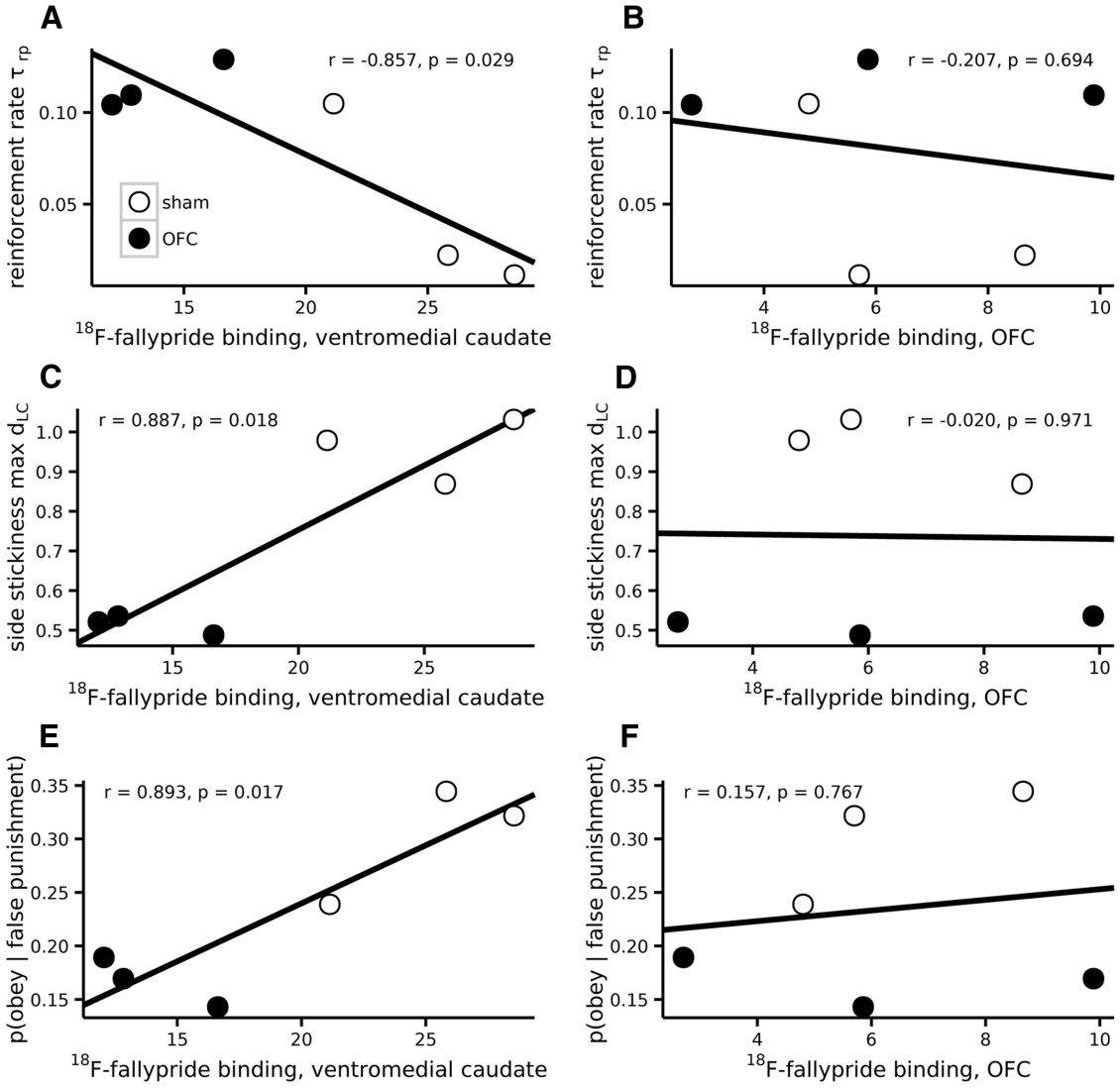

Figure 6. Relationship between behavior and striatal dopamine. ***A***, The d_{LC} parameter correlated with ^{18}F-fallypride BP_{ND} in the caudate (ventromedial caudate shown) but (***B***) not the OFC. ***C***, Similarly, the τ_{rp} parameter correlated with ^{18}F-fallypride BP_{ND} in the caudate but (***D***) not the OFC. The probability of shifting after false-negative feedback correlated with (***E***) the reduced levels of ventromedial caudate ^{18}F-fallypride BP_{ND} seen in the OFC DA-depleted monkeys but not (***F***) the levels of ^{18}F-fallypride BP_{ND} seen in the OFC.

strongly anticorrelated ($r = -0.959$, $p = 0.002$). Similarly, the probability of shifting following false punishment correlated with D2RB in both the ventromedial caudate ($r = 0.893$, $p = 0.017$; Fig. 6*E*) and the caudate body ($r = 0.841$, $p = 0.036$) but not the OFC ($r = 0.157$, $p = 0.767$; Fig. 6*F*). Neither τ_c (learning rate for stimulus stickiness) nor c (maximum for stimulus stickiness) correlated with D2RB in the caudate or OFC.

Because the computational model revealed reinforcement sensitivity as the key driver of changes in overall task performance, this suggests that performance improves as caudate DA increases and reinforcement sensitivity increases. Similar results were found with a voxel-based approach, and the differences did not exist preoperatively (data not shown).

Discussion

Depleting OFC DA led to an upregulation of tonic extracellular striatal DA levels, measured by microdialysis, with a corresponding decrease in DA D2/D3 receptor binding potential, measured by PET. This depletion improved subjects' ability to learn visual discriminations in a task offering partially ambiguous feedback. OFC DA-depleted subjects were less driven by a tendency to persist in choosing a recently chosen side, as established by computational modeling, although this change did not explain their behavioral alterations. They also showed an increase in reinforcement sensitivity, which did predict the observed behavioral

←

(*Figure legend continued.*) (τ_{rp}), a tendency to repeat choices to recently chosen stimuli (τ_c, c), and a tendency to repeat choices to recently chosen sides (τ_{LC}, d_{LC}). Lesioned subjects showed increased sensitivity to reinforcement (higher τ_{rp}). They also showed less side stickiness (shown both by a lower d_{LC}, indicating a reduction in the overall influence of side stickiness compared with that of reinforcement, and a higher τ_{LC}, indicating that the influence of side stickiness was less long-lasting). The dagger (†) indicates that between-group differences in τ_{rp} were necessary and sufficient for the other behavioral effects shown in ***D***, ***E*** (see Materials and Methods, and Results). Error bars show the posterior distributions of group differences in group mean parameter values, as highest-density intervals (HDIs; orange, 75% HDI excludes zero; red, 95% HDI excludes zero). Percentages are the posterior probabilities that the parameter differs from zero (width of the largest HDI excluding zero), as described in the Materials and Methods; they are not frequentist p values. ***D***, This computational model predicted fewer errors to criterion in the OFC DA-depleted group (compare with ***A***). ***E***, Moreover, the computational model predicted the differences in responding to false punishment in the behavioral data (compare with ***B***).

changes, namely a reduction in shifting away from the better stimulus in the face of punishment and a reduction in the number of errors made before criterion performance was attained. Parameters representing reinforcement sensitivity and the tendency to choose a recently chosen side were anticorrelated and correlated (respectively) with striatal D2RB, an inverse measure of striatal DA itself, but were not related to OFC D2RB. These results suggest that OFC DA depletion increases behavioral switching and reinforcement sensitivity via increases in striatal DA release.

The novel finding that DA depletion specifically within the OFC induces selective caudate DA excess is relevant to models of schizophrenia. Most previous work on the relationship between PFC and striatal DA relates to the dorsolateral PFC, the whole PFC, or the rodent ventromedial PFC, rather than the OFC. Catecholamine depletion of the ventromedial PFC in rats increases DA throughout the dorsal and ventral striatum (Pycock et al., 1980), whereas *N*-acetyl-aspartate levels (a putative marker of neuronal integrity) in the dorsolateral PFC predict striatal D2 receptor availability in schizophrenia (Bertolino et al., 1999). PFC DA receptor binding is abnormal in schizophrenia (Okubo et al., 1997) and the magnitude of prefrontal dysfunction predicts increased striatal DA uptake during the Wisconsin card-sorting task in schizophrenia (Meyer-Lindenberg et al., 2002) and the prodromal state (Fusar-Poli et al., 2010), supporting the hypothesis that abnormal frontostriatal interactions contribute to the development of this disorder. It is known that the OFC inhibits firing in the VTA (Lodge, 2011) and that OFC damage disrupts striatal dopaminergic signaling and learning from unexpected outcomes in rats (Takahashi et al., 2009, 2011) and humans (Tsuchida et al., 2010). Here, we demonstrate for the first time that a reduction in primate OFC DA elevates DA levels in the caudate (perhaps also via VTA disinhibition), the site where changes in dopaminergic function are associated with the onset of psychosis (Fusar-Poli et al., 2010).

The instrumental behavior required by the probabilistic discrimination task can be generated by several interacting neuropsychological systems (Cardinal et al., 2002). It can be habitual, using "model-free" reinforcement learning driven by reward prediction errors without representing the causal structure of the world, or goal-directed (model-based), based on an internal model of the consequences of actions derived from experience of their outcomes. The OFC is implicated in aspects of model-based learning (McDannald et al., 2011), and the balance between model-based and model-free learning can be altered by DA manipulations (Wunderlich et al., 2012). Our behavioral results were not well described by a shift between model-free and model-based learning systems but our task was not explicitly designed to compare the two, and may be underpowered to detect such effects. Indeed, computational models of model-based strategies described behavior poorly in both the lesion group and controls. Our results are also not explicable simply by changes in motor function: response latencies were unaffected.

The most parsimonious account of our behavioral results was offered by a model free computational model in which learning was driven by reinforcement (according to a simple delta rule operating at the same rate for reward and punishment), by stimulus stickiness (the tendency to choose the stimulus chosen on the previous trial) and side stickiness (the tendency to respond to the side of the testing chamber chosen on the previous trial). Like schizophrenia patients who show alterations in both strategy and reinforcement learning, OFC DA-depleted monkeys were less strongly influenced by their recently chosen side, and more influenced by reinforcement. This effect was predicted by caudate but not OFC D2RB. This model also retrodicted the behavioral outcomes that OFC DA-depleted monkeys learned the task faster and that their choice selection was less affected by unpredicted negative outcomes.

The effect on side stickiness can be viewed as favoring exploration of stimulus locations over exploitation, or as an increase in the rate of response-based or side-lateralized switching. Our results are compatible with Humphries et al. (2012) theory that tonic striatal DA influences the trade-off between exploration and exploitation. Their network simulations suggest that in a two-choice task, high tonic dopamine promotes exploration under certain circumstances and the exploration-exploitation trade-off can alter win-stay/lose-shift probabilities and overall measures of task success, as seen in our data. They provide a potential neurobiological substrate for the increase in response switching between two locations induced by the indirect DA agonist amphetamine (Evenden and Robbins, 1983; Ridley et al., 1988) in a manner similar to that seen in schizophrenia (Frith and Done, 1983). They are also consistent with the well established roles of striatal DA in controlling responding within egocentric space (Cook and Kesner, 1988). It may be that the effects on side stickiness and reinforcement sensitivity are neurally separable, because changes in reinforcement sensitivity (but not side stickiness) were capable of driving changes in win-stay/lose-shift behavior. If so, it might be that DA in the caudate body mediates changes in side stickiness specifically, given that D2RB changes in the caudate body correlated with side stickiness and not reinforcement sensitivity (which was limited to the head of the caudate).

Much interest has centered on the role of the striatum in reinforcement learning and it is of note that had we not selected among competing computational models, our win-stay/lose-shift outcomes would have found a ready explanation in reward prediction error signaling theories of the striatum (Sutton and Barto, 1998). Midbrain DA neurons fire in response to unexpected rewards (for convenience we will term these "blips"), and reduce their firing in response to unexpected omission of reward ("dips"; Schultz, 2002). In our study, OFC DA depletion increased tonic striatal DA without affecting K^+-induced phasic DA release. This increase in tonic DA might mask the dips when reward is unexpectedly not delivered (and a mildly aversive outcome delivered instead), without affecting the blips in response to unexpected reward. Accordingly, one would expect a selective decrease in the normal behavioral response to unexpected punishment/reward omission, as observed. However, although convenient, this interpretation would imply that the changes in behavior were related to the difference between unexpected and expected reward, or between reward and punishment, or both. Instead, our results were explicable in terms of a simpler change in reinforcement sensitivity. This could be viewed as an enhancement of a model-free reinforcement learning system due to increased caudate DA. It is no surprise that an increase in reinforcement sensitivity was associated with fewer errors to criterion. Because the majority of reinforcement is valid, the invalid, minority feedback impacts upon behavior less, and thus animals with increased sensitivity will be more likely to ignore misleading feedback. Our results also emphasize the importance of considering the reinforcement-independent functions of the striatum, because a change in response strategy can influence simple behavioral measures often assumed to depend on reinforcement learning (Humphries et al., 2012).

Striatal dopamine increases have been suggested to contribute to changes in salience processing and psychosis, particularly early

in the course of schizophrenia (Kapur, 2003; Fusar-Poli et al., 2010). In particular, because established schizophrenia is associated with impairments in reinforcement learning (Waltz et al., 2007), an important question arising from our results is whether prodromal or early psychosis, via increases in striatal dopamine, can be associated with improvements in reinforcement learning under some circumstances; certainly, global performance improvements have at times been reported in schizophrenia (Kasanova et al., 2011). Alternatively, the overall improvement apparent in our animal model may be a consequence of our control subjects not relying on a model-based system that is so prominent in humans. Schizophrenia involves many neural changes and animal models such as this one do not attempt to reproduce the entire disorder. Nevertheless, reproducing individual aspects of the disorder's complex neurobiology is helpful in isolating the cause of the individual neurobehavioral sequelae that do present in this complex disorder.

Orbitofrontal cortex DA function is abnormal in schizophrenia (Meador-Woodruff et al., 1997). The cause or causes of these abnormalities remain unknown, and it is uncertain whether they contribute to symptoms of the disorder, though it has long been hypothesized that prefrontal dopaminergic dysfunction is responsible for the striatal dopaminergic hyperfunction (Weinberger, 1987). One potential mechanism is via genetic changes affecting the OFC. Knock-out of the DISC1 schizophrenia susceptibility gene reduces OFC tyrosine hydroxylase expression (Sekiguchi et al., 2011). Another is via stress, as prolonged psychological stress reduces PFC DA transmission (Mizoguchi et al., 2000). A third is via distant cortical damage. For example, early ventromedial temporal lobe lesions damage PFC and impair dorsolateral prefrontal cortical regulation of striatal DA (Saunders et al., 1998; Bertolino et al., 2002), with dorsolateral PFC DA abnormalities also seen in schizophrenia (Davis et al., 1991). Here, using a combined behavioral, neuroimaging and computational approach we have demonstrated (to our knowledge for the first time) a specific additional mechanism of prefrontal-striatal regulation, in which DA depletion of the primate OFC causes an increase in tonic DA in the caudate nucleus. Behaviorally, this depletion caused an increase in the tendency to switch response location, a feature of choice behavior observed in patients with schizophrenia, and an increase in reinforcement sensitivity, both of which correlated with striatal but not OFC D2/D3 receptor binding. These results provide causal evidence that altered OFC DA transmission contributes to the striatal hyperdopaminergia known to contribute to behavioral dysfunction in schizophrenia.

References

Bertolino A, Knable MB, Saunders RC, Callicott JH, Kolachana B, Mattay VS, Bachevalier J, Frank JA, Egan M, Weinberger DR (1999) The relationship between dorsolateral prefrontal N-acetylaspartate measures and striatal dopamine activity in schizophrenia. Biol Psychiatry 45:660–667. CrossRef Medline

Bertolino A, Roffman JL, Lipska BK, van Gelderen P, Olson A, Weinberger DR (2002) Reduced *N*-acetylaspartate in prefrontal cortex of adult rats with neonatal hippocampal damage. Cereb Cortex 12:983–990. CrossRef Medline

Bouthenet ML, Souil E, Martres MP, Sokoloff P, Giros B, Schwartz JC (1991) Localization of dopamine D3 receptor mRNA in the rat brain using in situ hybridization histochemistry: comparison with dopamine D2 receptor mRNA. Brain Res 564:203–219. CrossRef Medline

Burnham KP, Anderson DR (2004) Multimodel inference: understanding AIC and BIC in model selection. Sociol Methods Res 33:261–304. CrossRef

Cardinal RN, Parkinson JA, Hall J, Everitt BJ (2002) Emotion and motivation: the role of the amygdala, ventral striatum, and prefrontal cortex. Neurosci Biobehav Rev 26:321–352. CrossRef Medline

Clarke HF, Walker SC, Dalley JW, Robbins TW, Roberts AC (2007) Cognitive inflexibility after prefrontal serotonin depletion is behaviorally and neurochemically specific. Cereb Cortex 17:18–27. CrossRef Medline

Cook D, Kesner RP (1988) Caudate nucleus and memory for egocentric localization. Behav Neural Biol 49:332–343. CrossRef Medline

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. CrossRef Medline

Copeland BJ, Neff NH, Hadjiconstantinou M (2005) Enhanced dopamine uptake in the striatum following repeated restraint stress. Synapse 57:167–174. CrossRef Medline

Davis KL, Kahn RS, Ko G, Davidson M (1991) Dopamine in schizophrenia: a review and reconceptualization. Am J Psychiatry 148:1474–1486. Medline

Daw N (2011) Trial-by-trial data analysis using computational models. In: Decision making, affect, and learning: attention and performance XXIII (M. R. Delgado, E. Phelps, T. Robbins, eds). Oxford: Oxford UP. CrossRef

den Ouden HE, Daw ND, Fernandez G, Elshout JA, Rijpkema M, Hoogman M, Franke B, Cools R (2013) Dissociable effects of dopamine and serotonin on reversal learning. Neuron 80:1090–1100. CrossRef Medline

Evenden JL, Robbins TW (1983) Increased response switching, perseveration and perseverative switching following d-amphetamine in the rat. Psychopharmacology 80:67–73. CrossRef Medline

Frith CD, Done DJ (1983) Stereotyped responding by schizophrenic patients on a two-choice guessing task. Psychol Med 13:779–786. CrossRef Medline

Fusar-Poli P, Howes OD, Allen P, Broome M, Valli I, Asselin MC, Grasby PM, McGuire PK (2010) Abnormal frontostriatal interactions in people with prodromal signs of psychosis: a multimodal imaging study. Arch Gen Psychiatry 67:683–691. CrossRef Medline

Groman SM, Lee B, London ED, Mandelkern MA, James AS, Feiler K, Rivera R, Dahlbom M, Sossi V, Vandervoort E, Jentsch JD (2011) Dorsal striatal D2-like receptor availability covaries with sensitivity to positive reinforcement during discrimination learning. J Neurosci 31:7291–7299. CrossRef Medline

Gunn RN, Lammertsma AA, Hume SP, Cunningham VJ (1997) Parametric imaging of ligand-receptor binding in PET using a simplified reference region model. Neuroimage 6:279–287. CrossRef Medline

Haber SN, Kunishio K, Mizobuchi M, Lynd-Balta E (1995) The orbital and medial prefrontal circuit through the primate basal ganglia. J Neurosci 15:4851–4867. Medline

Howes OD, Montgomery AJ, Asselin MC, Murray RM, Valli I, Tabraham P, Bramon-Bosch E, Valmaggia L, Johns L, Broome M, McGuire PK, Grasby PM (2009) Elevated striatal dopamine function linked to prodromal signs of schizophrenia. Arch Gen Psychiatry 66:13–20. CrossRef Medline

Humphries MD, Khamassi M, Gurney K (2012) Dopaminergic control of the exploration-exploitation trade-off via the basal ganglia. Front Neurosci 6:9. CrossRef Medline

Kanahara N, Sekine Y, Haraguchi T, Uchida Y, Hashimoto K, Shimizu E, Iyo M (2013) Orbitofrontal cortex abnormality and deficit schizophrenia. Schizophr Res 143:246–252. CrossRef Medline

Kapur S (2003) Psychosis as a state of aberrant salience: a framework linking biology, phenomenology, and pharmacology in schizophrenia. Am J Psychiatry 160:13–23. CrossRef Medline

Kapur S, Remington G (2001) Dopamine D(2) receptors and their role in atypical antipsychotic action: still necessary and may even be sufficient. Biol Psychiatry 50:873–883. CrossRef Medline

Kasanova Z, Waltz JA, Strauss GP, Frank MJ, Gold JM (2011) Optimizing vs. matching: response strategy in a probabilistic learning task is associated with negative symptoms of schizophrenia. Schizophr Res 127:215–222. CrossRef Medline

Kinahan PE, Rogers JG (1989) Analytic 3D image reconstruction using all detected events. IEEE Trans Nucl Sci 36:964–968. CrossRef

Kolachana BS, Saunders RC, Weinberger DR (1995) Augmentation of prefrontal cortical monoaminergic activity inhibits dopamine release in the caudate nucleus: an in vivo neurochemical assessment in the rhesus monkey. Neuroscience 69:859–868. CrossRef Medline

Lataster J, Collip D, Ceccarini J, Haas D, Booij L, van Os J, Pruessner J, Van Laere K, Myin-Germeys I (2011) Psychosocial stress is associated with in vivo dopamine release in human ventromedial prefrontal cortex: a posi-

tron emission tomography study using [(1) (8)F]fallypride. Neuroimage 58:1081–1089. CrossRef Medline

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

Leichnetz GR, Astruc J (1975) Efferent connections of the orbitofrontal cortex in the marmoset (*Saguinus oedipus*). Brain Res 84:169–180. CrossRef Medline

Lidow MS, Goldman-Rakic PS, Gallager DW, Rakic P (1991) Distribution of dopaminergic receptors in the primate cerebral cortex: quantitative autoradiographic analysis using [3H]raclopride, [3H]spiperone and [3H]SCH23390. Neuroscience 40:657–671. CrossRef Medline

Lodge DJ (2011) The medial prefrontal and orbitofrontal cortices differentially regulate dopamine system function. Neuropsychopharmacology 36: 1227–1236. CrossRef Medline

McDannald MA, Lucantonio F, Burke KA, Niv Y, Schoenbaum G (2011) Ventral striatum and orbitofrontal cortex are both required for model-based, but not model-free, reinforcement learning. J Neurosci 31:2700–2705. CrossRef Medline

Meador-Woodruff JH, Haroutunian V, Powchik P, Davidson M, Davis KL, Watson SJ (1997) Dopamine receptor transcript expression in striatum and prefrontal and occipital cortex: focal abnormalities in orbitofrontal cortex in schizophrenia. Arch Gen Psychiatry 54:1089–1095. CrossRef Medline

Meyer-Lindenberg A, Miletich RS, Kohn PD, Esposito G, Carson RE, Quarantelli M, Weinberger DR, Berman KF (2002) Reduced prefrontal activity predicts exaggerated striatal dopaminergic function in schizophrenia. Nat Neurosci 5:267–271. CrossRef Medline

Miyake N, Thompson J, Skinbjerg M, Abi-Dargham A (2011) Presynaptic dopamine in schizophrenia. CNS Neurosci Ther 17:104–109. CrossRef Medline

Mizoguchi K, Yuzurihara M, Ishige A, Sasaki H, Chui DH, Tabira T (2000) Chronic stress induces impairment of spatial working memory because of prefrontal dopaminergic dysfunction. J Neurosci 20:1568–1574. Medline

Okubo Y, Suhara T, Suzuki K, Kobayashi K, Inoue O, Terasaki O, Someya Y, Sassa T, Sudo Y, Matsushima E, Iyo M, Tateno Y, Toru M (1997) Decreased prefrontal dopamine D1 receptors in schizophrenia revealed by PET. Nature 385:634–636. CrossRef Medline

Pycock CJ, Kerwin RW, Carter CJ (1980) Effect of lesion of cortical dopamine terminals on subcortical dopamine receptors in rats. Nature 286: 74–76. CrossRef Medline

R Core Team (2012) R: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing.

Ridley RM, Baker HF, Frith CD, Dowdy J, Crow TJ (1988) Stereotyped responding on a two-choice guessing task by marmosets and humans treated with amphetamine. Psychopharmacology (Berl) 95:560–564. Medline

Roberts AC, De Salvia MA, Wilkinson LS, Collins P, Muir JL, Everitt BJ, Robbins TW (1994) 6-Hydroxydopamine lesions of the prefrontal cortex in monkeys enhance performance on an analog of the Wisconsin card sort test: possible interactions with subcortical dopamine. J Neurosci 14: 2531–2544. Medline

Roberts AC, Tomic DL, Parkinson CH, Roeling TA, Cutter DJ, Robbins TW, Everitt BJ (2007) Forebrain connectivity of the prefrontal cortex in the marmoset monkey (*Callithrix jacchus*): an anterograde and retrograde tract-tracing study. J Comp Neurol 502:86–112. CrossRef Medline

Saunders RC, Kolachana BS, Bachevalier J, Weinberger DR (1998) Neonatal lesions of the medial temporal lobe disrupt prefrontal cortical regulation of striatal dopamine. Nature 393:169–171. CrossRef Medline

Schultz W (2002) Getting formal with dopamine and reward. Neuron 36: 241–263. CrossRef Medline

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

Scornaiencki R, Cantrup R, Rushlow WJ, Rajakumar N (2009) Prefrontal cortical D1 dopamine receptors modulate subcortical D2 dopamine receptor-mediated stress responsiveness. Int J Neuropsychopharmacol 12:1195–1208. CrossRef Medline

Scott L (2010) A modern Bayesian look at the multi-armed bandit. Appl Stochastic Models Bus Ind 26:639–658. CrossRef

Sekiguchi H, Iritani S, Habuchi C, Torii Y, Kuroda K, Kaibuchi K, Ozaki N (2011) Impairment of the tyrosine hydroxylase neuronal network in the orbitofrontal cortex of a genetically modified mouse model of schizophrenia. Brain Res 1392:47–53. CrossRef Medline

Seymour B, Daw ND, Roiser JP, Dayan P, Dolan R (2012) Serotonin selectively modulates reward value in human decision-making. J Neurosci 32:5833–5842. CrossRef Medline

Stan Development Team (2014) Stan: a C++ library for probability and sampling, version 2.2. http://mc-stan.org.

Stephan KE, Penny WD, Daunizeau J, Moran RJ, Friston KJ (2009) Bayesian model selection for group studies. Neuroimage 46:1004–1017. CrossRef Medline

Sutton R, Barto A (1998) Reinforcement learning. Cambridge, MA: MIT.

Takahashi YK, Roesch MR, Stalnaker TA, Haney RZ, Calu DJ, Taylor AR, Burke KA, Schoenbaum G (2009) The orbitofrontal cortex and ventral tegmental area are necessary for learning from unexpected outcomes. Neuron 62:269–280. CrossRef Medline

Takahashi YK, Roesch MR, Wilson RC, Toreson K, O'Donnell P, Niv Y, Schoenbaum G (2011) Expectancy-related changes in firing of dopamine neurons depend on orbitofrontal cortex. Nat Neurosci 14:1590–1597. CrossRef Medline

Tsuchida A, Doll BB, Fellows LK (2010) Beyond reversal: a critical role for human orbitofrontal cortex in flexible learning from probabilistic feedback. J Neurosci 30:16868–16875. CrossRef Medline

van Winkel R, Stefanis NC, Myin-Germeys I (2008) Psychosocial stress and psychosis: a review of the neurobiological mechanisms and the evidence for gene-stress interaction. Schizophr Bull 34:1095–1105. CrossRef Medline

Volkow ND, Chang L, Wang GJ, Fowler JS, Ding YS, Sedler M, Logan J, Franceschi D, Gatley J, Hitzemann R, Gifford A, Wong C, Pappas N (2001) Low level of brain dopamine D2 receptors in methamphetamine abusers: association with metabolism in the orbitofrontal cortex. Am J Psychiatry 158:2015–2021. CrossRef Medline

Walker SC, Robbins TW, Roberts AC (2009) Differential contributions of dopamine and serotonin to orbitofrontal cortex function in the marmoset. Cereb Cortex 19:889–898. CrossRef Medline

Waltz JA, Gold JM (2007) Probabilistic reversal learning impairments in schizophrenia: further evidence of orbitofrontal dysfunction. Schizophr Res 93:296–303. CrossRef Medline

Waltz JA, Frank MJ, Robinson BM, Gold JM (2007) Selective reinforcement learning deficits in schizophrenia support predictions from computational models of striatal-cortical dysfunction. Biol Psychiatry.

Weinberger DR (1987) Implications of normal brain development for the pathogenesis of schizophrenia. Arch Gen Psychiatry 44:660–669. CrossRef Medline

Wunderlich K, Smittenaar P, Dolan RJ (2012) Dopamine enhances model-based over model-free choice behavior. Neuron 75:418–424. CrossRef Medline