NeuroImage 109 (2015) 95–101

Contents lists available at ScienceDirect

NeuroImage

journal homepage: www.elsevier.com/locate/ynimg

Striatal D1 and D2 signaling differentially predict learning from positive and negative outcomes

Sylvia M.L. Cox [a,b], Michael J. Frank [c], Kevin Larcher [a,b], Lesley K. Fellows [a], Crystal A. Clark [a], Marco Leyton [b], Alain Dagher [a,*]

[a] Montreal Neurological Institute, McGill University, 3801 University Street, Montreal, Quebec H3A 2B4, Canada
[b] Department of Psychiatry, McGill University, 1033 Pine Avenue West, Montreal, Quebec H3A 1A1, Canada
[c] Department of Cognitive, Linguistic & Psychological Sciences, Brown Institute for Brain Science, Brown University, 190 Thayer Street, Providence, RI 02912-1821, USA

ARTICLE INFO

Article history:
Accepted 27 December 2014
Available online 3 January 2015

Keywords:
Dopamine
Reinforcement
Learning
Positron emission tomography
Striatum

ABSTRACT

The extent to which we learn from positive and negative outcomes of decisions is modulated by the neurotransmitter dopamine. Dopamine neurons burst fire in response to unexpected rewards and pause following negative outcomes. This dual signaling mechanism is hypothesized to drive both approach and avoidance behavior. Here we test a prediction deriving from a computational reinforcement learning model, in which approach is mediated via activation of the direct cortico-striatal pathway due to striatal D1 receptor stimulation, while avoidance occurs via disinhibition of indirect pathway striatal neurons secondary to a reduction of D2 receptor stimulation. Using positron emission tomography with two separate radioligands, we demonstrate that individual differences in human approach and avoidance learning are predicted by variability in striatal D1 and D2 receptor binding, respectively. Moreover, transient dopamine precursor depletion improved learning from negative outcomes. These findings support a bidirectional modulatory role for striatal dopamine in reward and avoidance learning via segregated D1 and D2 cortico-striatal pathways.

Introduction

While striatal dopamine signaling is widely thought to play an important role in reward learning (Schultz, 2002), its contribution to learning from negative outcomes is more controversial. Dopamine neurons burst fire following the presentation of unexpected rewards, and pause when an expected reward has been withheld, allowing them to encode a reward prediction error (RPE) signal (Montague et al., 1996; Schultz et al., 1997). One class of reinforcement learning models suggests that phasic dopamine bursts support learning from positive outcomes, but, that, due to low dynamic range at the low end of dopamine signaling, other neuromodulators must be involved in learning from negative outcomes (Daw et al., 2002; Bayer and Glimcher, 2005). Another model, tested here, proposes that dopamine modulates both approach and avoidance learning via two segregated pathways in the cortico-striato-thalamocortical circuit (Frank, 2005; Frank and O'Reilly, 2006; Frank et al., 2007a). In this model (Fig. 1), striatal medium spiny neurons of the direct pathway, which express dopamine D1 receptors (D1R), facilitate the selection of rewarding actions encoded in the cortex. Those belonging to the indirect pathway, which express dopamine D2 receptors (D2R), help to suppress cortical patterns that encode maladaptive or non-rewarding actions (Surmeier et al., 2011). This opponent system model may better account for asymmetrical effects of dopamine manipulations on reward and punishment learning than single-value reinforcement learning models (Collins and Frank, 2014). One of its predictions is that positive reinforcement learning will be modulated by signaling within the D1 direct pathway while negative reinforcement learning will be modulated by signaling within the D2 indirect pathway.

In humans, there is indirect evidence that dopamine modulates approach and avoidance learning in opposite directions (Frank et al., 2004; Cools et al., 2006; Frank and O'Reilly, 2006; Pessiglione et al., 2006; Moustafa et al., 2008). Genetic studies suggest that these opposite modulations are related to D1R and D2R signaling (Frank et al., 2007b, 2009; Frank and Hutchison, 2009; Jocham et al., 2009; Doll et al., 2011), but direct evidence is lacking.

Here we assess the potentially distinct roles of D1R and D2R signaling in human learning and decision making by measuring receptor availabilities using positron emission tomography (PET) with separate tracers selective for D1R ([^{11}C]SCH23390) and D2R ([^{11}C]raclopride), and relating these measures to performance on the Probabilistic Selection Task (PST), which simultaneously measures reinforcement and avoidance learning. In a second study, we employed the PST and acute phenylalanine and tyrosine depletion (APTD) to transiently decrease dopamine levels, providing a causal test of the model prediction that this manipulation would improve learning from negative outcomes.

* Corresponding author at: Montreal Neurological Institute, 3801 University St., Montréal, QC H3A 2B4, Canada.
E-mail address: alain.dagher@mcgill.ca (A. Dagher).

http://dx.doi.org/10.1016/j.neuroimage.2014.12.070
1053-8119/

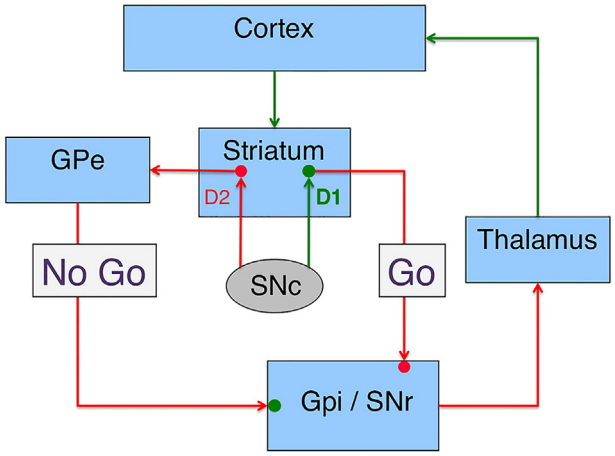

Fig. 1. Anatomy of the cortico-basal ganglia-thalamo-cortical loop. Dopamine modulates the execution of responses involved in learning via two segregated pathways in the cortico-striato-thalamocortical circuit: The direct Go pathway and the indirect NoGo pathway. Striatal medium spiny neurons of the direct pathway mainly express D1 receptors and project to the internal segment of the globus pallidus (Gpi) and the substantia nigra pars reticulata (SNr), which in turn disinhibits the thalamus, thereby facilitating thalamic projection to the cortex. Striatal neurons in the indirect pathway mostly express D2 receptors and project to the external segment of the globus pallidus (GPe), by which they reduce the tonic inhibition of the GPe on the GPi/SNr, which in turn leads to a suppression of the thalamic output to the cortex. Excitatory (inhibitory) projections in green (red).

Materials & methods

Subjects

Study 1

Twenty-eight young (age 20.6 ± 1.9) healthy volunteers (14 females) were recruited through an advertisement in the McGill University classified ads. Participants were excluded for any of the following reasons: regular current or past drug use including smoking (lifetime use of nicotine or marijuana > 20 occasions, lifetime use of any other illicit drug > 3 occasions), any current or past Axis-I disorder assessed by a brief version of the Structural Clinical Interview of DSM-IV diagnoses (First et al., 2002); any current or past neurological condition; any other current or past medical condition that might affect the interpretation of the study results; use of central nervous system active medications; positive pregnancy test; and positive urine toxicologic test results for illicit drugs (Triage-TM Panel for Drugs of Abuse; Biosite Diagnostics, San Diego, California). Subjects were preselected based on successful performance of the PST; they had to reach the performance criteria during training as described below. The study was carried out in accordance with the Declaration of Helsinki and approved by the Research Ethics Board of the Montreal Neurological Institute. All participants provided written, informed consent.

Study 2

A second, independent sample (N = 15; 6 females) of young (21.5 ± 2.6) healthy volunteers was recruited using the exclusion criteria described above. Ten subjects were kept for the final analysis as 5 subjects had to be excluded: two because the APTD drink did not produce a reduction in tyrosine and phenylalanine levels, and three because they did not successfully learn the PST (see below), choosing A > B in fewer than 50% of test phase trials. The study was carried out in accordance with the Declaration of Helsinki and approved by the Research Ethics Board of the Montreal Neurological Institute. All participants provided written, informed consent.

Procedure

Study 1

Subjects underwent four test sessions, consisting of one cognitive session, an anatomical MRI scan, and two PET scans: one with the D1R tracer [^{11}C]SCH23390, and one with the D2R tracer [^{11}C]raclopride.

Subjects were asked to abstain from caffeine for at least 4 h and from alcohol for at least 24 h prior to each test session and to eat a light meal 1 h prior to coming to the laboratory. All sessions took place between 10 am and 4 pm. At the start of each test session, mood and anxiety levels were measured using the Profile of Mood States questionnaire (Lorr et al., 1982) and the Spielberger State Anxiety Inventory (Spielberger et al., 1983) to ensure stable mood patterns across test days. Prior to each session, a urine screen was done to confirm the absence of any use of drugs of abuse or CNS active medications (including cocaine, amphetamines, barbiturates, benzodiazepines, Δ^9-tetrahydrocannabinol, opiates and phencyclidine). Women were tested during the follicular phase to ensure stable levels of estrogen across test days.

Study 2

Subjects came to the laboratory on two separate occasions during which they completed the PST after ingesting one of the following two amino acid (AA) mixtures: 1) An AA mixture that was deficient in dopamine precursors tyrosine and phenylalanine, known to temporarily cause a reduction in dopamine synthesis and 2) A nutritionally balanced AA mixture. Studies in animals (Palmour et al., 1998; McTavish et al., 1999; Brodnik et al., 2013) and humans (Montgomery et al., 2003; Leyton et al., 2004) have established that this method reduces CSF amine metabolite levels and brain dopamine levels by approximately 30% (Montgomery et al., 2003) for a period of several hours. The AA mixtures were administered double blind, in counterbalanced order. The APTD mixture's composition, preparation, and administration are based on a 100 g nutritionally balanced mixture with phenylalanine and tyrosine withheld, as described elsewhere (Leyton et al., 2000).

The day prior to each test session subjects followed a low protein diet provided by the investigators and fasted from midnight. On the actual test days, subjects came to the laboratory at 8.30 am, had a blood sample taken and ingested one of the two AA mixtures. Subjects were then asked to relax (but remain awake) for 3.5 h in a room with relatively neutral videos and reading material available to them. Four hours after AA administration, subjects performed the PST as described below, and a blood sample was taken directly after. Subjects who failed to show a reduction in plasma tyrosine at 4 h were excluded from analysis.

Probabilistic Selection Task

We used the PST to assess the ability to learn from positive (Go learning) and negative (NoGo Learning) outcomes (Frank et al., 2004; Frank and Hutchison, 2009). The paradigm consists of a training and a test phase. During training, three different stimulus pairs were presented (A/B, C/D and E/F) in random order and participants had to learn to select the "correct" stimulus. Visual feedback was provided ("Correct!" or "Incorrect"), but this feedback was probabilistic. Choosing stimulus A led to correct feedback on 80% of the AB trials, whereas choosing stimulus B led to correct feedback on 20% of trials. The reward probabilities were 70/30 for stimulus pair C/D and 60/40 for E/F respectively. Once subjects reached the performance criterion (65% A in AB, 60% C in CD and 50% E in EF), they advanced to the test phase.

During the test phase subjects had to select among all possible combinations of stimulus pairings, without further feedback, and were instructed to choose the symbol that was more likely to be "correct" based on what they learned during training. Each stimulus was paired with every other stimulus for a total of 60 trials. Positive and negative feedback learning was assessed by calculating the percentage of trials where patterns A and C were chosen (A > CDEF + C > EF) and B and D

were avoided (B < CDEF + D < EF), respectively, when presented in novel combinations. This measure is very similar to the "choose-A"/"avoid-B" distinction often used in this task, but is somewhat more sensitive and includes more trials. Results were qualitatively similar with the standard measures, but there were some participants who performed at ceiling on "choose-A"/"avoid-B", making the data less appropriate for statistical inference. Note that, for evaluation of learning performance in the test phase, A (and C) and B (and D) are always compared to a set of stimuli that are on average neutral (mean value of 50%).

Classical reinforcement learning models assume that each action has a single value that reflects the integrated history of both positive and negative contingencies and that the agent then makes choices based on the relative difference in values among available actions. These models cannot predict any difference in choose-A and avoid-B performance (because the difference between A and neutral is the same as the difference between neutral and B). However, when these classical models are modified to incorporate opponent valuation systems that differentially represent positive and negative values (summarizing the D1 and D2 cortico-striatal pathways in our model), they exhibit differences in simulated choose-A vs. avoid-B performance when there is an asymmetry in the degree of learning in these two systems (Frank et al., 2007a; Collins and Frank, 2014).

Image acquisition and processing

Subjects were scanned twice on a Siemens ECAT high-resolution research tomograph (HRRT) PET camera (207 slice-coverage with a spatial resolution range between 2.3–3.4 mm full width at half maximum). At the beginning of each PET session a 6 minute ^{137}Cs transmission scan for attenuation correction was acquired followed by a bolus injection of 8–10 mCi of [^{11}C]SCH23390 or [^{11}C]raclopride. For each scan emission data were collected over 60 min in 26 time frames of progressively longer duration. No task was administered during scanning; subjects were instructed to remain awake and rest quietly.

The PET images were reconstructed using the Ordinary Poisson Ordered Subset Expectation Maximization (OP-OSEM) reconstruction algorithm with 10 iterations and 16 subsets (Comtat et al., 2004), which included correction for scatter, randoms, attenuation, normalization, resolution degradation and head motion (Costes et al., 2009). The reconstructed image frames were composed of 256 × 256 × 207 voxels (voxel side length = 1.21875 mm).

For anatomical coregistration, high-resolution (1 mm^3) T1-weighted magnetic resonance images (MRI) were obtained for all participants on a Siemens Sonata 1.5 T system, using a gradient echo pulse sequence (repetition time (TR) = 22 ms, echo time (TE) = 9.2 ms, flip angle 30° and matrix size 176 × 256 × 256). Each MR image was first pre-processed with the CIVET pipeline (version 1.1.9) (wiki.bic.mni.mcgill.ca/index.php/CIVET) developed at the Montreal Neurological Institute (MNI) for fully automated structural image analysis (Ad-Dab'bagh et al., 2006). The native MR volumes were corrected for image intensity non-uniformity (Sled et al., 1998), and linearly and non-linearly transformed into standardized MNI space using automated feature matching to the ICBM152 template (Collins et al., 1994). The MR image in MNI space was classified into white matter, gray matter and CSF (Sled et al., 1998), and was automatically segmented using a probabilistic atlas based approach (Collins and Evans, 1997). Regions of interest (ROI), including the ventral striatum, caudate and putamen, were defined on each individual's MRI in MNI space using a high resolution template (Fonov et al., 2009).

The spatial rigid-body transformation between the summed PET volume and the native MR image was estimated with normalized mutual information, and was used to position the ROI masks into the native PET space. The resulting registration was visually checked for the whole brain and at the level of basal ganglia. Since [^{11}C]SCH23390 has a higher affinity for binding to extrastriatal dopaminergic receptors than [^{11}C]raclopride, a more accurate MRI/PET coregistration was obtained with the former tracer. In order to improve the MRI/PET coregistration for the raclopride images, we linearly transformed the [^{11}C]raclopride volume onto the [^{11}C]SCH23390 volume. We then combined these transformation parameters with those of the [^{11}C]SCH23390 to MRI transform to position the ROI masks into native PET [^{11}C]raclopride space.

For both [^{11}C]raclopride and [^{11}C]SCH23390 scans, time–activity curves were extracted from the original (non-smoothed) dynamic image by masking the ROIs onto the native PET space. The ROIs were eroded in order to reduce partial volume effects on the PET images. Regional [^{11}C]raclopride and [^{11}C]SCH23390 non-displaceable binding potential values (BP_{ND}) were then computed for each ROI using tools developed by Turku PET center (http://www.turkupetcentre.net/). For each ROI, BP_{ND} values were calculated using the Simplified Reference Tissue Model with the cerebellum as the reference region, which is devoid of D1/D2/D3 receptors, to describe the kinetics of the free and specifically bound ligand (Lammertsma and Hume, 1996). BP_{ND} expresses the relationship between the estimated density of available dopamine receptors (B_{avail}), the dissociation constant of its target dopamine receptor (K_D) and the free fraction of non-specifically bound tracer in the brain (F_{ND}) (Mintun et al., 1984):

$$BP_{ND} = F_{ND} \cdot \left({}^{B_{avail}}\!/\!_{K_D} \right). \quad (1)$$

Additionally, voxel-wise BP_{ND} images were generated for each subject for each tracer, using the time–activity curve at each voxel and the same Simplified Reference Tissue Model as for the ROI. These maps were used to generate statistical parametric maps of the correlations between BP_{ND} and task performance (see below).

BP_{ND} is proportional to the density of available receptors (B_{avail}), which is itself a function of total receptor density (B_{max}). However, [^{11}C]Raclopride and [^{11}C]SCH23390 BP_{ND} are related to receptor density in different ways (Marcellino et al., 2012). [^{11}C]SCH23390 BP_{ND} is linearly proportional to D1R density or B_{max}, and unaffected by variations in dopamine levels, likely due to the low affinity of dopamine for the D1R, and to the fact that the ligand binds to the receptor equally well in either affinity state (high or low), or when it is internalized. Several studies in primates (Chou et al., 1999) and mice (Thibaut et al., 1996) confirm that SCH23390 binding to dopamine receptors is unaffected by acute changes in extracellular dopamine. We thus predicted a linear relationship between [^{11}C]SCH23390 BP_{ND} and learning from positive outcomes.

[^{11}C]Raclopride BP_{ND}, on the other hand, depends on both D2R density and endogenous dopamine levels (Laruelle et al., 1997). Thus [^{11}C]raclopride BP_{ND} can be expressed as a function of receptor occupancy (σ, ranging from 0 to 1), which depends on tonic dopamine levels, and the theoretical number of receptor sites available for binding (BP_0, the density of available receptors if there were no dopamine present) represented by the following equation (Gjedde et al., 2010).

$$BP_{ND} = (1-\sigma) \cdot BP_0 \quad (2)$$

Since receptor occupancy and the number of available receptor sites have opposing effects on [^{11}C]raclopride BP_{ND}, we predicted a non-linear, quadratic relationship between the [^{11}C]raclopride BP_{ND} and learning from negative outcomes (see below).

Statistical analyses

Study 1

To assess the relationship between BP_{ND} of the D1R and D2R tracers and learning from positive and negative outcomes (study 1) linear (Pearson correlation) and non-linear (quadratic) regression models were used, respectively. The significance of these regression effects was tested at the regional level using the BP_{ND} values for each ROI, as

well as at the level of individual voxels. SPSS Version 20 was used to perform statistical analysis of the ROI data, with significance set at $p < 0.05$. Additionally, a T-map of voxel-wise linear regression between [^{11}C]SCH23390 BP_{ND} and Go learning was generated using Eq. (3) and an F-map of the voxel-wise quadratic regression between BP_{ND} [^{11}C]raclopride and NoGo learning was created using Eq. (4).

$$BP_{ND}\left[^{11}C\right]\text{SCH23390} = \alpha_0 + \alpha_1 \text{Go} \quad \alpha_1 \neq 0 \tag{3}$$

$$BP_{ND}\left[^{11}C\right]\text{raclopride} = \beta_0 + \beta_1 \text{NoGo} + \beta_2 \text{NoGo}^2 \quad \beta_2 \neq 0 \tag{4}$$

Both statistical parametric maps were generated in MNI space, thresholded at $p = 0.05$ corrected for multiple comparisons using the striatum as search volume (Worsley et al., 1996), and superimposed on the average T1-weighted MRI of all participants.

Study 2

Amino acid levels in blood were measured using gradient reverse-phase high-performance liquid chromatography with fluorometric detection. The availability of dopamine precursors in the brain was assessed by calculating the ratio of tyrosine and phenylalanine to all other large neutral amino acids (LNAAs) (including valine, methionine, isoleucine, leucine, tryptophan, tyrosine and phenylalanine in the denominator) at two time points: at baseline, prior to AA mixture ingestion and 4 h post ingestion. The effect of AA mixture on blood plasma levels was measured using a 2 × 2 (time × AA condition) repeated measures ANOVA. A paired t-test across sessions was used to measure the effect of APTD on learning from negative outcomes. Significance was set at $p < 0.05$.

Results

Study 1: relationship between D1R and D2R availability and learning

Learning from positive outcomes correlated linearly with D1R binding as assessed by [^{11}C]SCH23390 BP_{ND} in the caudate ($r = .57$, $p = .002$), putamen ($r = .52$, $p = .005$) and the striatum as a whole ($r = .55$, $p = .003$) (Fig. 2A). Since D1R and D2R binding significantly correlated with each other in the striatum ($r^2 = 0.23$, $p = 0.012$), a linear regression model including both D1R and D2R binding in the striatum as predictors of learning from positive outcomes was applied to assess the selectivity of this relationship. We observed a significant linear relationship (caudate: $F(2, 24) = 5.78$, $p = .009$; putamen: $F(2,24) = 4.58$, $p = .021$; striatum whole: $F(2, 24) = 5.33$, $p = .012$), with a selective effect of D1R binding on learning from positive feedback (caudate: $t = 2.6$, $p = .016$; putamen: $t = 2.5$, $p = .019$; striatum whole: $t = 2.7$, $p = .013$), but no effect of D2R binding (caudate: $t = .46$, $p = .65$; putamen: $t = .42$, $p = .68$; striatum whole: $t = .34$, $p = .74$). A voxel-wise linear regression model further confirmed the significance of these findings, demonstrating a linear relationship between D1R binding and learning from positive outcomes ($t = 6.2$, $p < 0.05$, corrected, Fig. 2B). No relationship was observed between D1R and learning from negative outcomes ($r^2 \leq 0.01$, $p > 0.6$). The correlation between learning from positive outcomes and ventral striatum [^{11}C]SCH23390 BP_{ND} was non-significant ($p = 0.12$).

Conversely, in line with the model prediction, we observed a quadratic, inverted U-shape relationship between [^{11}C]raclopride BP_{ND} and learning from negative feedback in the putamen ($r^2 = .19$, $p = 0.03$, Fig. 2C). A trend towards significance was found in the caudate nucleus ($p = 0.054$) and striatum as a whole ($p = 0.058$) but not the ventral striatum ($p = 0.6$). The quadratic effect was confirmed by a voxel-wise non-linear regression F-map of [^{11}C]raclopride BP_{ND} in the striatum and learning from negative outcomes, which shows significant peaks in bilateral putamen (Fig. 2E). No relationship was observed,

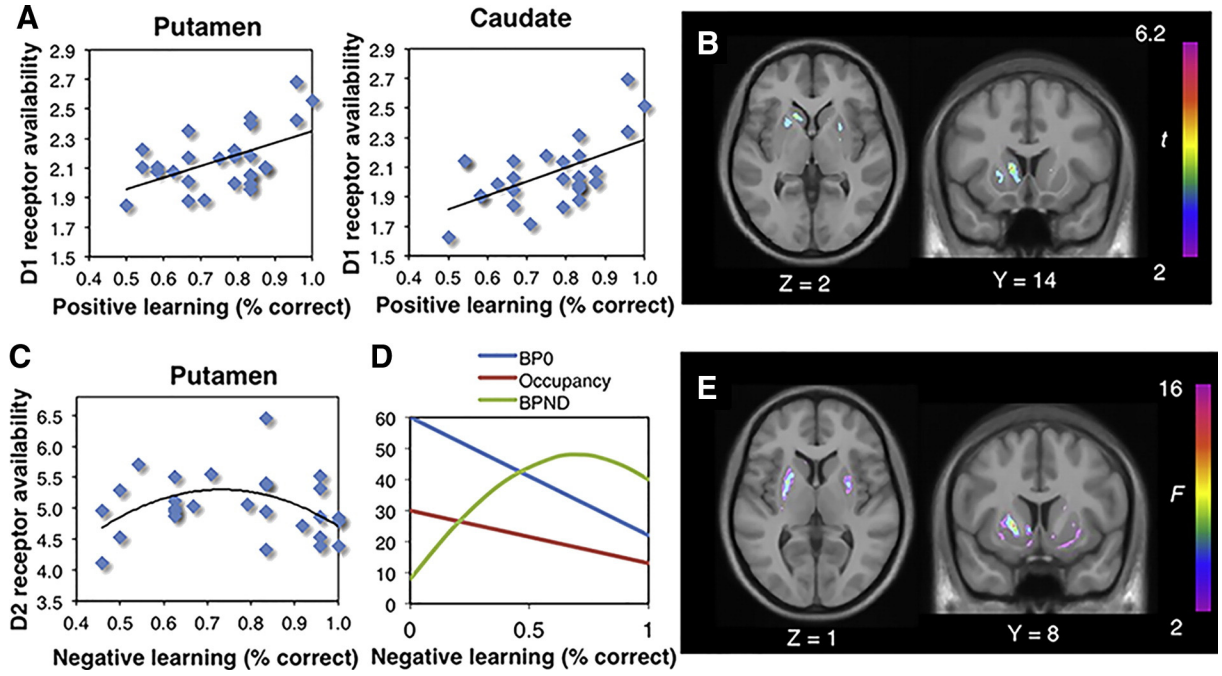

Fig. 2. D1R and D2R availability differentially predict learning from positive and negative outcomes. (A) [^{11}C]SCH23390 BP_{ND}, indicative of D1 receptor density, in the caudate ($r = 0.57$, $p = 0.002$) and putamen ($r = 0.52$, $p = 0.005$) predicts individual differences in learning from positive outcomes. (B) T-map of voxel-wise linear regression between [^{11}C]SCH23390 BP_{ND} in the striatum and reward learning ($p = 0.05$, corrected). (C) Inverted U relationship between [^{11}C]raclopride BP_{ND} in the putamen and learning from negative outcomes ($r^2 = 0.19$, $p = 0.03$). (D) Theoretical model of the relationship between NoGo learning and D2 binding as measured by PET, adapted from Gjedde et al. (2010). $BP_{ND} = (1 - \sigma) * BP_0$. The occupancy ($\sigma$) is the occupancy of D2R by dopamine. BP_0 is the total number of D2R available for binding. BP_{ND} is the measured raclopride BP by PET. (E) F-map of voxel-wise non-linear regression between [^{11}C]raclopride BP_{ND} in the striatum and avoidance learning ($p = 0.05$ corrected).

either linear or quadratic, between [^{11}C]raclopride BP_{ND} and learning from positive outcomes after correcting for the effects of D1R binding ($r^2 = 0.005$, $p > 0.7$).

Study 2: the effect of dopamine depletion on learning

Subjects performed the PST on two separate occasions, once after APTD, and once after drinking a balanced amino acid solution. APTD caused a significant reduction in the ratio of tyrosine ($F(1,9) = 28.2$, $p < 0.0001$) and phenylalanine ($F(1,9) = 111.9$, $p < 0.0001$) to large neutral amino acids as reflected by a AA mixture × time interaction, confirming that the APTD mixture successfully reduced dopamine precursor levels (Fig. 3A), both compared to baseline, and compared to the BAL session. The mean reduction in plasma tyrosine was 75% (SD: 6.6%). Two subjects failed to show a reduction in plasma tyrosine and phenylalanine and were excluded from the analysis. This reduction in dopamine precursor levels significantly improved learning from negative outcomes ($t(9) = 2.3$, $p = .046$), without affecting learning from positive feedback ($t(9) = -.23$, $p = .556$, Fig. 3B).

Discussion

Together, these findings demonstrate that individual differences in learning from positive and negative outcomes are related to striatal dopamine D1R and D2R function, respectively. The linear relationship between [^{11}C]SCH23390 BP_{ND} and learning from positive outcomes suggests that D1R signaling modulates the sensitivity to phasic dopamine bursts in response to RPEs, consistent with predictions from animal research (Montague et al., 1996; Schultz et al., 1997). As mentioned above, [^{11}C]SCH23390 BP_{ND} is an index of D1R density and is independent of tonic dopamine levels. D1R-modulated signals modify cortico-striatal synaptic plasticity in response to phasic dopamine signals, thereby facilitating adaptive learning to positive outcomes (Reynolds et al., 2001; Reynolds and Wickens, 2002; Shen et al., 2008), consistent with optogenetic manipulations demonstrating positive instrumental conditioning following stimulation of D1R-bearing striatal medium spiny neurons (Kravitz et al., 2012).

We also demonstrate an inverted U-shaped relationship between D2R binding and learning from negative outcomes. According to the model (Frank, 2005), NoGo learning occurs when dopamine levels are sufficiently low to allow negative RPEs to disinhibit D2R-bearing medium spiny neurons. This is supported by our findings that tyrosine and phenylalanine depletion improved learning from negative outcomes. While previous studies have shown enhanced NoGo learning in unmedicated Parkinson's disease patients (Frank et al., 2004), here we demonstrate that reductions in dopamine per se can selectively improve negative feedback learning in healthy humans. At a cellular level, long-term potentiation within the indirect pathway occurs following a lack of D2R stimulation (Shen et al., 2008) and is enhanced by D2R blockade (Beeler et al., 2012), while direct optogenetic stimulation of striatal D2R neurons promotes avoidance behavior (Kravitz et al., 2012).

The direct and indirect cortico-striatal pathways support the learning of adaptive and maladaptive action values via differential effects of dopamine bursts and dips on D1R and D2R, respectively. D1R have low affinity for dopamine (Marcellino et al., 2012) and only respond to large increases in synaptic dopamine released during phasic dopamine neuron bursts that reflect positive RPEs, supporting learning to approach rewarding stimuli (Frank, 2005). Conversely, the higher affinity D2R in the indirect pathway can detect transient reductions in tonic dopamine levels that follow pauses in dopamine neuron firing during negative RPEs, thereby supporting learning to avoid negative outcomes (Frank, 2005). Note that D2R stimulation reduces excitability of medium spiny neuron of the indirect pathway (Hernández-López et al., 2000); therefore, reductions in D2R signaling have the effect of potentiating the inhibitory indirect pathway.

The observed inverted U relationship between raclopride binding and learning from negative outcomes is consistent with the fact that [^{11}C]raclopride BP_{ND} reflects not only D2R density, but also tonic dopamine levels (Laruelle et al., 1997). Accordingly, the ability to learn from negative outcomes will be greater either when tonic dopamine levels are low, leading to low occupancy (and higher BP_{ND}), or when D2R density is low (leading to lower BP_{ND}). Both of these conditions are associated with reduced D2 signaling, which facilitates neuronal firing and long-term potentiation within D2R-bearing medium spiny neurons of the indirect pathway during negative RPEs. This explains the quadratic relationship between [^{11}C]raclopride BP_{ND} and NoGo learning as per Eq. (2) and Fig. 2D (also see Supplementary materials). This interpretation is further supported by genetic studies showing that the C allele of the promoter polymorphism rs12364283 of *DRD2*, which is associated with higher transcriptional activity and D2R density (Zhang et al., 2007) in post-mortem studies, was associated with poorer learning from negative outcomes using the same task (Frank and Hutchison, 2009). Note that the foregoing argument assumes that dopamine receptor density determines the signaling strength of dopamine, for both D1R and D2R. In other words, we suggest that the intensity of the dopamine signal in the striatum (i.e. its effect on learning) depends on the absolute number of interactions between dopamine and its post-synaptic receptors, rather than the percent of receptors occupied. This hypothesis cannot be tested here, but the linear relationship between D1R density and positive learning shown here, and the findings related to the D2R promoter and learning from negative outcomes both support this view.

The APTD findings further support the interpretation that NoGo learning is facilitated by reduced D2R occupancy by dopamine, which makes it easier for indirect pathway neurons to become disinhibited by negative RPEs. Note that acute changes in tonic dopamine do not lead to changes in dopamine receptor expression (Laruelle et al., 1997), therefore the effect of APTD is assumed to be limited to reduced

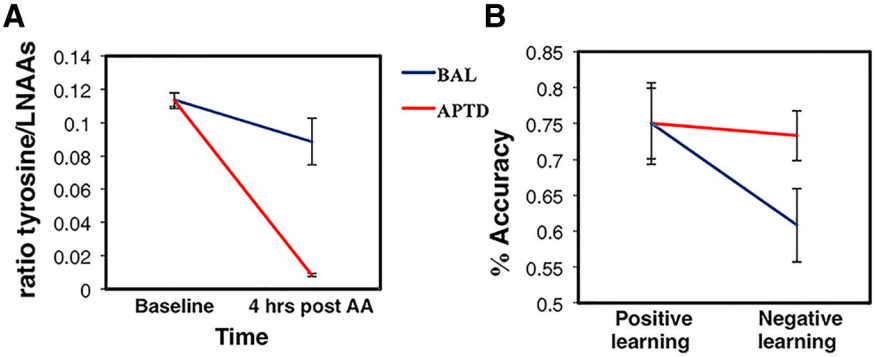

Fig. 3. APTD reduces dopamine precursor tyrosine levels in blood and improves learning from negative outcomes. (A) APTD compared to the nutritionally balanced mixture (BAL) significantly reduced blood tyrosine/LNAAs ratio ($p < 0.001$). The result for phenylalanine/LNAAs ratio was the same ($p < 0.001$). (B) APTD significantly improved learning from negative outcomes ($t = 2.3$, $p = 0.046$) without affecting learning from positive feedback. These data, from different sessions in the same individuals, are represented as mean ± SEM.

receptor occupancy. APTD did not affect Go learning. According to the model (Frank, 2005), NoGo learning is quite sensitive to changes in baseline tonic dopamine levels, influencing the ability of D2R-bearing medium spiny neurons to respond to pauses in dopamine firing in response to negative RPEs. According to recent evidence, APTD affects phasic as well as tonic dopamine release (Le Masurier et al., 2013). However, modeling of synaptic dopamine signaling suggests that depleting vesicular dopamine may only affect tonic signaling at D2 receptors and have little effect on phasic signaling (Dreyer et al., 2010). Please see Supplementary materials for details.

These results have implications for our understanding of striatal dopamine signaling and its role in pathologies of motivated behavior. First, they support the view that dopamine is not only implicated in positive reinforcement, but that reductions in dopamine are also relevant for learning from negative outcomes via D2 signaling. Rodent studies have demonstrated that D2R blockade not only induces motor skill deficits, but also leads to persistently impaired performance even after normalization of D2 signaling, implicating aberrant learning (Beeler et al., 2012). The present findings suggest that these same mechanisms are relevant to avoidance learning from negative outcomes in humans.

Our finding that low dopamine facilitates rather than impairs learning from negative outcomes specifically supports the theory that dopamine acts as a RPE signal rather than a saliency signal. According to the saliency hypothesis low dopamine would be expected to cause a reduction in both Go and NoGo learning (Bromberg-Martin et al., 2010). The bidirectional effect of dopamine on processing positive and negative outcomes, as observed here, forms one of the key distinctions between the RPE and saliency hypotheses, and thus emphasizes the role of dopamine in coding RPE, although this does not exclude the existence of an additional dopaminergic saliency signal.

The model of basal ganglia processing supported by these results provides a mechanism, at the computational level, for impulse control disorders. Persistent over-stimulation of striatal D2R should reduce the impact of negative outcomes. This may explain the phenomenon of impulsivity induced by dopamine agonist medications. The dopamine agonists that cause impulse control disorders such as pathological gambling and compulsive shopping preferentially stimulate the D2R family (Potenza et al., 2007; Voon et al., 2011). In many cases the impulsive behavior is time-locked to drug administration. Impulsivity, in this case, would result from an inability to consider the impact of negative outcomes. Further support for this notion comes from two other reports showing an inverted U-shaped relationship between [^{11}C]raclopride BP_{ND} and measures of impulsive personality, namely the personality construct of sensation seeking (Gjedde et al., 2010), which is itself associated with impulsive and risky behavior (Zuckerman and Kuhlman, 2000), and Negative Urgency (Clark et al., 2012), which is associated with problem gambling. Taken together with these findings, our results suggest that inherent differences in dopamine D2R signaling may predispose individuals to addictive and impulsive disorders. This may require re-thinking the labeling of this dopaminergic vulnerability as reflecting "reward deficiency" (Dagher and Robbins, 2009), which implies that increased vulnerability results from reduced dopamine signaling. In contrast, our findings suggest that increased D2R signaling may be associated with increased vulnerability to addictive and impulsive disorders, which is consistent with other studies that have shown that decreased activity of the indirect pathway (or increased D2R stimulation) predisposes to addictive behaviors (Collins and Woods, 2009; Lobo et al., 2010; Maia and Frank, 2011). Future studies could test whether impaired NoGo learning and impulsivity correlate across individuals or following dopaminergic manipulations, and investigate the relationship between negative reinforcement and the urge to engage in reward seeking behavior.

Our results need to be interpreted in light of the following limitations. First, the sample size of the APTD study ($n = 10$) was modest, but sufficient to detect a significant effect within-subjects, and within the commonly accepted range for assessing pharmacological challenges. Second, our interpretation of the D2 results rests on several assumptions about D2 signaling and APTD. As highlighted above, measures of [^{11}C]raclopride BP_{ND} are unable to differentiate between D2R density (B_{max}) and receptor occupancy by dopamine. These two measures have opposing effects on D2 signaling and thus on learning from negative outcomes. Although this accounts for the observed inverted U-shape function, this method does not allow us to disentangle the contribution of endogenous dopamine levels versus receptor density, nor interactions between the two. Other methods will be needed to address this issue. Nevertheless, our findings do show selective modulation of positive and negative learning by D1 and D2 signaling, respectively. Third, the PST, as employed here, is unable to tease apart the expression of learned associations versus the learning itself. Therefore, an alternative explanation for our findings is that dopamine signaling mediates value-based choice performance rather than learning per se (Smittenaar et al., 2012), an interpretation that is also consistent with recent refinements to the computational model used here (Collins and Frank, 2014).

Conclusion

Our findings support a modulatory role for striatal dopamine in reward and avoidance based learning via segregated striatal D1R and D2R pathways. Individual differences in D1R and D2R binding predicted learning from positive and negative outcomes of decisions, respectively. This variability in D1R and D2R signaling may be responsible for individual differences in the response to reward and punishment related signals, which may underlie differences in vulnerability to drug addiction, obesity and other impulse control disorders. We further suggest that the often-noted association between genetic or PET measures of D2R and drug addiction or pathological gambling implicates impaired punishment learning in these disorders.

Acknowledgments

We thank Gabriel Wolf for his help with programming the cognitive task and Julia Wagner, Lauren Templeton, Helen Hsu, Vera Khramova and Allison Goodwin-Wilson for their help with data collection and entry. This work was supported by CIHR (MOP 144079) and NSERC (436259) grants to AD and NIH grant (R21DA022630) to LKF.

Conflict of interest

The authors declare no competing financial interests.

Appendix A. Supplementary data

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.neuroimage.2014.12.070.

References

Ad-Dab'bagh, Y., Einarson, D., Lyttelton, O., Muehlboeck, J.-S., Mok, K., Ivanov, O., Vincent, R.D., Lepage, C., Lerch, J., Fombonne, E., Evans, A.C., 2006. The CIVET image-processing environment: a fully automated comprehensive pipeline for anatomical neuroimaging research. Organization for Human Brain Mapping (Florence).

Bayer, H.M., Glimcher, P.W., 2005. Midbrain dopamine neurons encode a quantitative reward prediction error signal. Neuron 47, 129–141.

Beeler, J.A., Frank, M.J., McDaid, J., Alexander, E., Turkson, S., Bernandez, M.S., McGehee, D.S., Zhuang, X., 2012. A role for dopamine-mediated learning in the pathophysiology and treatment of Parkinson's disease. Cell Rep. 2, 1747–1761.

Brodnik, Z., Double, M., Jaskiw, G.E., 2013. Presynaptic regulation of extracellular dopamine levels in the medial prefrontal cortex and striatum during tyrosine depletion. Psychopharmacology 227, 363–371.

Bromberg-Martin, E.S., Matsumoto, M., Hikosaka, O., 2010. Dopamine in motivational control: rewarding, aversive, and alerting. Neuron 68, 815–834.

Chou, Y.H., Karlsson, P., Halldin, C., Olsson, H., Farde, L., 1999. A PET study of D1-like dopamine receptor ligand binding. Psychopharmacology 146, 220–227.

Clark, L., Stokes, P.R., Wu, K., Michalczuk, R., Benecke, A., Watson, B.J., Egerton, A., Piccini, P., Nutt, D.J., Bowden-Jones, H., Lingford-Hughes, A.R., 2012. Striatal dopamine D2/D3 receptor binding in pathological gambling is correlated with mood-related impulsivity. NeuroImage 63, 40–46.

Collins, D.L., Evans, A.C., 1997. Animal: validation and applications of nonlinear registration-based segmentation. Int. J. Pattern Recognit. Artif. Intell. 11.
Collins, A.G., Frank, M.J., 2014. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol. Rev. 121, 337–366.
Collins, G.T., Woods, J.H., 2009. Influence of conditioned reinforcement on the response-maintaining effects of quinpirole in rats. Behav. Pharmacol. 20, 492–504.
Collins, D.L., Neelin, P., Peters, T.M., Evans, A.C., 1994. Automatic 3D intersubject registration of MR volumetric data in standardized Talairach space. J. Comput. Assist. Tomogr. 18, 192–205.
Comtat, C., Bataille, F., Michel, C., Jones, J.P., Sibomana, M., Janeiro, L., Trebossen, R., 2004. OSEM-3D reconstruction strategies for the ECAT HRRT. IEEE Nucleur Science Symposium Conference Record 6, pp. 3492–3496.
Cools, R., Altamirano, L., D'Esposito, M., 2006. Reversal learning in Parkinson's disease depends on medication status and outcome valence. Neuropsychologia 44, 1663–1673.
Costes, N., Dagher, A., Larcher, K., Evans, A.C., Collins, D.L., Reilhac, A., 2009. Motion correction of multi-frame PET data in neuroreceptor mapping: simulation based validation. NeuroImage 47, 1496–1505.
Dagher, A., Robbins, T., 2009. Personality, addiction, dopamine: insights from Parkinson's disease. Neuron 61, 502–510.
Daw, N.D., Kakade, S., Dayan, P., 2002. Opponent interactions between serotonin and dopamine. Neural Netw. 15, 603–616.
Doll, B.B., Hutchison, K.E., Frank, M.J., 2011. Dopaminergic genes predict individual differences in susceptibility to confirmation bias. J. Neurosci. 31, 6188–6198.
Dreyer, J.K., Herrik, K.F., Berg, R.W., Hounsgaard, J.D., 2010. Influence of phasic and tonic dopamine release on receptor activation. J. Neurosci. 30, 14273–14283.
First, M.B., Spitzer, R.L., Gibbon, M., Williams, J.B.W., 2002. Structural Clinical Interview for DSM-IV-TR Axis I Disorders, Research Version, Non-patient Edition (SCID-I/NP). Biometrics Research, New York State Psychiatric Institute, New York (November).
Fonov, V.S., Evans, A.C., McKinstry, R.C., Almli, C.R., Collins, D.L., 2009. Unbiased nonlinear average age-appropriate brain templates from birth to adulthood. Organization for Human Brain Mapping, p. S102.
Frank, M.J., 2005. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J. Cogn. Neurosci. 17, 51–72.
Frank, M.J., Hutchison, K., 2009. Genetic contributions to avoidance-based decisions: striatal D2 receptor polymorphisms. Neuroscience 164, 131–140.
Frank, M.J., O'Reilly, R.C., 2006. A mechanistic account of striatal dopamine function in human cognition: psychopharmacological studies with cabergoline and haloperidol. Behav. Neurosci. 120, 497–517.
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., Scheres, A., Sherman, S.J., 2007a. Understanding decision-making deficits in neurological conditions: insights from models of natural action selection. Philos. Trans. R. Soc. Lond. B Biol. Sci. 362, 1641–1654.
Frank, M.J., Moustafa, A.A., Haughey, H.M., Curran, T., Hutchison, K.E., 2007b. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. Proc. Natl. Acad. Sci. U. S. A. 104, 16311–16316.
Frank, M.J., Doll, B.B., Oas-Terpstra, J., Moreno, F., 2009. Prefrontal and striatal dopaminergic genes predict individual differences in exploration and exploitation. Nat. Neurosci. 12, 1062–1068.
Gjedde, A., Kumakura, Y., Cumming, P., Linnet, J., Moller, A., 2010. Inverted-U-shaped correlation between dopamine receptor availability in striatum and sensation seeking. Proc. Natl. Acad. Sci. U. S. A. 107, 3870–3875.
Hernández-López, S., Tkatch, T., Perez-Garci, E., Galarraga, E., Bargas, J., Hamm, H., Surmeier, D.J., 2000. D2 dopamine receptors in striatal medium spiny neurons reduce L-type Ca^{2+} currents and excitability via a novel PLCβ1–IP3–calcineurin-signaling cascade. J. Neurosci. 20, 8987–8995.
Jocham, G., Klein, T., Neumann, J., von Cramon, D., Reuter, M., Ullsperger, M., 2009. Dopamine DRD2 polymorphism alters reversal learning and associated neural activity. J. Neurosci. 29, 3695–3704.
Kravitz, A., Tye, L., Kreitzer, A., 2012. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat. Neurosci. 15, 816–818.
Lammertsma, A.A., Hume, S.P., 1996. Simplified reference tissue model for PET receptor studies. NeuroImage 4, 153–158.
Laruelle, M., D'Souza, C., Baldwin, R., Abi-Dargham, A., Kanes, S., Fingado, C., Seibyl, J., Zoghbi, S., Bowers, M., Jatlow, P., Charney, D., Innis, R., 1997. Imaging D2 receptor occupancy by endogenous dopamine in humans. Neuropsychopharmacology 17, 162–174.
Le Masurier, M., Zetterström, T., Cowen, P., Sharp, T., 2013. Tyrosine-free amino acid mixtures reduce physiologically-evoked release of dopamine in a selective and activity-dependent manner. J. Psychopharmacol. 28, 561–569.
Leyton, M., Young, S., Pihl, R., Etezadi, S., Lauze, C., Blier, P., Baker, G., Benkelfat, C., 2000. Effects on mood of acute phenylalanine/tyrosine depletion in healthy women. Neuropsychopharmacology 22, 52–63.
Leyton, M., Dagher, A., Boileau, I., Casey, K., Baker, G., Diksic, M., Gunn, R.N., Young, S., Benkelfat, C., 2004. Decreasing amphetamine-induced dopamine release by acute phenylalanine/tyrosine depletion: a PET/[11C]raclopride study in healthy men. Neuropsychopharmacology 29, 427–432.
Lobo, M.K., Covington III, H.E., Chaudhury, D., Friedman, A.K., Sun, H., Damez-Werno, D., Dietz, D.M., Zaman, S., Koo, J.W., Kennedy, P.J., Mouzon, E., Mogri, M., Neve, R.L., Deisseroth, K., Han, M.H., Nestler, E.J., 2010. Cell type-specific loss of BDNF signaling mimics optogenetic control of cocaine reward. Science 330, 385–390.
Lorr, M., McNair, D.M., Fisher, S.U., 1982. Evidence for bipolar mood states. J. Pers. Assess. 46, 432–436.
Maia, T.V., Frank, M.J., 2011. From reinforcement learning models to psychiatric and neurological disorders. Nat. Neurosci. 14, 154–162.
Marcellino, D., Kehr, J., Agnati, L.F., Fuxe, K., 2012. Increased affinity of dopamine for D(2)-like versus D(1)-like receptors. Relevance for volume transmission in interpreting PET findings. Synapse 66, 196–203.
McTavish, S., Cowen, P., Sharp, T., 1999. Effect of a tyrosine-free amino acid mixture on regional brain catecholamine synthesis and release. Psychopharmacology 141, 182–188.
Mintun, M.A., Raichle, M.E., Kilbourn, M.R., Wooten, G.F., Welch, M.J., 1984. A quantitative model for the in vivo assessment of drug binding sites with positron emission tomography. Ann. Neurol. 15, 217–227.
Montague, P., Dayan, P., Sejnowski, T., 1996. A framework for mesencephalic dopamine systems based on predictive Hebbian learning. J. Neurosci. 16, 1936–1947.
Montgomery, A.J., McTavish, S.F., Cowen, P.J., Grasby, P.M., 2003. Reduction of brain dopamine concentration with dietary tyrosine plus phenylalanine depletion: an [11C] raclopride PET study. Am. J. Psychiatry 160, 1887–1889.
Moustafa, A.A., Cohen, M.X., Sherman, S.J., Frank, M.J., 2008. A role for dopamine in temporal decision making and reward maximization in parkinsonism. J. Neurosci. 28, 12294–12304.
Palmour, R.M., Ervin, F.R., Baker, G.B., Young, S.N., 1998. An amino acid mixture deficient in phenylalanine and tyrosine reduces cerebrospinal fluid catecholamine metabolites and alcohol consumption in vervet monkeys. Psychopharmacology 136, 1–7.
Pessiglione, M., Seymour, B., Flandin, G., Dolan, R., Frith, C., 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. Nature 442, 1042–1045.
Potenza, M., Voon, V., Weintraub, D., 2007. Drug insight: impulse control disorders and dopamine therapies in Parkinson's disease. Nat. Clin. Pract. Neurol. 3, 664–672.
Reynolds, J., Wickens, J., 2002. Dopamine-dependent plasticity of corticostriatal synapses. Neural Netw. 15, 507–521.
Reynolds, J., Hyland, B., Wickens, J., 2001. A cellular mechanism of reward-related learning. Nature 413, 67–70.
Schultz, W., 2002. Getting formal with dopamine and reward. Neuron 36, 241–263.
Schultz, W., Dayan, P., Montague, P.R., 1997. A neural substrate of prediction and reward. Science 275, 1593–1599.
Shen, W., Flajolet, M., Greengard, P., Surmeier, D.J., 2008. Dichotomous dopaminergic control of striatal synaptic plasticity. Science 321, 848–851.
Sled, J.G., Zijdenbos, A.P., Evans, A.C., 1998. A nonparametric method for automatic correction of intensity nonuniformity in MRI data. IEEE Trans. Med. Imaging 17, 87–97.
Smittenaar, P., Chase, H., Aarts, E., Nusselein, B., Bloem, B., Cools, R., 2012. Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection—learning or performance? Eur. J. Neurosci. 35, 1144–1151.
Spielberger, C.D., Gorsuch, R.L., Lushene, R., Vagg, P.R., Jacobs, G.A., 1983. Manual for the State-Trait Anxiety Inventory. Consulting Psychologists Press, Palo Alto, CA.
Surmeier, D., Carrillo-Reid, L., Bargas, J., 2011. Dopaminergic modulation of striatal neurons, circuits, and assemblies. Neuroscience 198, 3–18.
Thibaut, F., Vaugeois, J.-M., Bonnet, J.-J., Costentin, J., 1996. In vivo striatal binding of the D1 antagonist SCH 23390 is not modified by changes in dopaminergic transmission. Neuropharmacology 35, 267–272.
Voon, V., Schoerling, A., Wenzel, S., Ekanayake, V., Reiff, J., Trenkwalder, C., Sixel-D ring, F., 2011. Frequency of impulse control behaviours associated with dopaminergic therapy in restless legs syndrome. BMC Neurol. 11, 117.
Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Hum. Brain Mapp. 4, 58–73.
Zhang, Y., Bertolino, A., Fazio, L., Blasi, G., Rampino, A., Romano, R., Lee, M.L., Xiao, T., Papp, A., Wang, D., Sadee, W., 2007. Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. Proc. Natl. Acad. Sci. U. S. A. 104, 20552–20557.
Zuckerman, M., Kuhlman, D., 2000. Personality and risk-taking: common biosocial factors. J. Pers. 68, 999–1029.