Striatal Dopamine Predicts Outcome-Specific Reversal Learning and Its Sensitivity to Dopaminergic Drug Administration

Roshan Cools,1 Michael J. Frank,2,4 Sasha E. Gibbs,2,5 Asako Miyakawa,2 William Jagust,2,6 and Mark D’Esposito2

1Donders Institute for Brain, Cognition and Behaviour, Radboud University Nijmegen Medical Centre, 6500 HB Nijmegen, The Netherlands, 2Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720-3192, Departments of Cognitive and Linguistic Sciences and Psychology, Brown University, Providence, Rhode Island 02912, 3Department of Psychology, Stanford University, Stanford, California 94305, and 4Lawrence Berkeley National Laboratory, Berkeley, California 94720

Individual variability in reward-based learning has been ascribed to quantitative variation in baseline levels of striatal dopamine. However, direct evidence for this pervasive hypothesis has hitherto been unavailable. We demonstrate that individual differences in reward-based reversal learning reflect variation in baseline striatal dopamine synthesis capacity, as measured with neurochemical positron emission tomography. Subjects with high baseline dopamine synthesis in the striatum showed relatively better reversal learning from unexpected rewards than from unexpected punishments, whereas subjects with low baseline dopamine synthesis in the striatum showed the reverse pattern. In addition, baseline dopamine synthesis predicted the direction of dopaminergic drug effects. The D2 receptor agonist bromocriptine improved reward-based relative to punishment-based reversal learning in subjects with low baseline dopamine synthesis capacity in the striatum. Finally, this pattern of drug effects was outcome-specific, and driven primarily by drug effects on punishment-, but not reward-based reversal learning. These data demonstrate that the effects of D2 receptor stimulation on reversal learning in humans depend on task demands and baseline striatal dopamine synthesis capacity.

Key words: dopamine; reward; punishment; striatum; PET; learning

Introduction

Adaptation to our environment requires the anticipation of biologically relevant events by learning signals of their occurrence, i.e., reward-based learning. Models of reward-based learning use a prediction error signal, representing the difference between expected and obtained events, to update their predictions based on the environment (Sutton and Barto, 1998). A putative mechanism of the prediction error signal for reward is the phasic firing of dopamine neurons in the midbrain (Montague et al., 1996; Schultz et al., 1997). These neurons innervate large parts of the brain, including the striatum, the major input structure of the basal ganglia. In keeping with this anatomical arrangement, the striatum has often been implicated in reward-based learning and its modulation by dopamine (Cools and Robbins, 2004; Frank, 2005; Pessiglione et al., 2006; Schönberg et al., 2007) and reward-based learning is modulated by agonists of D2/D3 receptors that are abundant in the striatum (Frank and O’Reilly, 2006; Pizzagalli et al., 2008). Schönberg et al. (2007) have recently proposed that individual differences in reward-based learning may reflect differences in striatal dopamine function. However, there is no direct evidence for this hypothesis. Here we demonstrate a significant positive relationship between reward-based learning and baseline striatal dopamine synthesis capacity, as measured with uptake of the positron emission tomography (PET) tracer [18F]fluorometatyrosine (FMT).

We further establish the link between dopamine in the striatum and reward-based learning by showing that effects of dopamine D2 receptor stimulation also depend on baseline striatal dopamine synthesis capacity. Evidence from studies with experimental animals (Williams and Goldman-Rakic, 1995; Zahrt et al., 1997; Arnsten, 1998) has revealed an “inverted U” shaped relationship between D1 receptor stimulation in the prefrontal cortex and cognitive performance. This relationship has been related to baseline-dependency of drug effects, so that low baseline dopamine levels are remedied, while high baseline dopamine levels are detrimentally over-dosed by the same dopamine D1 receptor agonist (Phillips et al., 2004). Although recent studies with humans, which have made use of genetic variation in the D2...
receptor gene, have suggested that a similar mechanism might underlie contrasting effects of D<sub>2</sub> receptor stimulation in the striatum on reward-based learning (Frank and O’Reilly, 2006; Cohen et al., 2007), direct evidence for baseline-dependency of dopaminergic drug effects on reward-learning in the striatum is lacking. We combined neurochemical PET imaging with behavioral psychopharmacology to test this hypothesis. We studied the effects of a single oral dose (1.25 mg) of the dopamine D<sub>2</sub> receptor agonist bromocriptine on reversal learning in young healthy volunteers, who also, on a separate occasion, underwent a PET scan with the tracer FMT. Subjects with low synthesis capacity were predicted to benefit from D<sub>2</sub> receptor stimulation with bromocriptine, while subjects with high synthesis capacity were predicted to be detrimentally overdosed by the same drug.

We used a reversal learning paradigm that enabled the separate assessment of reward- and punishment-based reversal learning (Cools et al., 2008a). Based on prior data indicating that dopaminergic drug effects are outcome-specific (Cools et al., 2006), we anticipated contrasting effects of bromocriptine on reward- and punishment-based reversal learning.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Eleven subjects [all female; mean (SD) age = 22.2 (2.0)] underwent a single PET scan with the tracer 6-[18F]fluoro-L-m-tyrosine (FMT). The PET data from these subjects were previously reported in relation to their working memory capacity as measured with the listening span test (Cools et al., 2008b). Detailed neuropsychological characteristics of the subjects are presented in that previous paper. All subjects were screened for psychiatric and neurological disorders; exclusion criteria were any history of cardiac, hepatic, renal, pulmonary, neurological, psychiatric or gastrointestinal disorders, an episode of loss of consciousness, use of psychotropic drugs, sleeping pills and heavy marihuana use (>10 times in a lifetime).

Subjects were selected from a sample of subjects that had also participated in a psychopharmacological study on the effects of bromocriptine (Cools et al., 2007). Initial selection of subjects for this study was based on their high or low scores on the Barratt Impulsiveness Inventory (BIS-11) (Patton et al., 1995). However, there was no relationship between dopamine synthesis capacity and trait impulsivity, as reported in our previous study (Patton et al., 1995). Hence, we used the BIS-11 to select subjects for this study.

Subjects were divided into two groups based on their baseline dopamine D<sub>2</sub> receptor gene expression levels as estimated from the PET data. The PET data were used to establish a baseline dopamine synthesis capacity (K<sub>i,0</sub>) for each subject. Subjects were considered to have low synthesis capacity if their K<sub>i,0</sub> was below the mean, and high synthesis capacity if their K<sub>i,0</sub> was above the mean. The mean K<sub>i,0</sub> of the study sample was 2.5 mCi/ml.

A detailed neuropsychological profile of the subjects was obtained prior to the study. All subjects were screened for psychiatric and neurological disorders; exclusion criteria were any history of cardiac, hepatic, renal, pulmonary, neurological, psychiatric, or gastrointestinal disorders, an episode of loss of consciousness, use of psychotropic drugs, sleeping pills and heavy marihuana use (>10 times in a lifetime).

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.

Materials and Methods

General procedure. The University of California Berkeley Committee for the Protection of Human Subjects approved the procedures, which were in accord with the Helsinki Declaration of 1975.
reflect a relatively stable process (synthesis capacity) that is not particularly sensitive to small state-related changes, much like uptake of the striatal \( K_i \) is a reliable measurement, with it having a 95% chance of lying within 18% of its value within an individual normal subject. We argue that the delay does not confound our results, but rather renders them more striking. Any instability in the PET measurement across the delay would have reduced rather than enhanced the likelihood of obtaining the results, which were statistically controlled for noise by an \( \alpha \) level of 0.05. Data analyses supported this hypothesis, as the effects were stronger when effects of interest were corrected for the delay between the two sessions than when they were not. Here we report only those analyses in which we corrected for the delay, by entering it as a covariate, although all effects were also significant when they were not corrected for the delay. Furthermore, in the supplemental Results C (available at www.jneurosci.org as supplemental material), we report an additional analysis, explicitly addressing in a quantitative manner the possibility that the effect reflects noise.

**Experimental paradigm.** The task required the learning and reversal of predictions of reward and punishment. On each trial two vertically adjacent stimuli were presented: one face and one scene; location randomized; about 19 inch viewing distance; subtending \( \sim 3^\circ \) horizontally and \( 3.5^\circ \) vertically. One of the stimuli was associated with reward, while the other was associated with punishment. On each trial, one of the two stimuli was highlighted with a black border surrounding the stimulus and subjects had to predict, based on trial and error learning, whether the highlighted stimulus would lead to reward or punishment. Subjects indicated their predictions by pressing, with the index or middle finger, one of two colored buttons (corresponding to keys “b” and “n” depending on the response-outcome mapping) on a laptop keyboard. They pressed the green button for reward and the red button for punishment. The outcome-response mappings were counterbalanced between subjects. The (self-paced) response was followed by an interval of 1000 ms, after which the outcome was presented for 500 ms. Note that this outcome did not provide direct performance feedback. Reward consisted of a green smiling face, a “+ $100” sign and a high-frequency jingle tone. Punishment consisted of a red sad face, a “− $100” sign and a single low-frequency tone. After the outcome, the screen was cleared for 500 ms, after which the next two stimuli were presented. The stimulus-outcome contingencies reversed multiple times provided learning criteria were met.

Each subject performed one practice block and four experimental blocks. Each practice block consisted of one acquisition stage and one reversal stage (learning criterion was 20 [not necessarily consecutive] correct trials). Each experimental block consisted of one acquisition stage and one reversal stage, for a total of four reversal stages per block. Subjects indicated by unexpected outcome ("switch trials") reflected the degree to which subjects updated their predictions based on unexpected outcomes. The stimulus that was highlighted on the first trial of each reversal stage (on which the unexpected outcome was presented) was always highlighted again on the second trial of that stage (i.e., the switch trial on which the subject had to implement the reversed contingencies and switch their predictions).

Based on prior data (Frank et al., 2004; Cools et al., 2006; Frank and O’Reilly, 2006), we predicted that bromocriptine would have contrasting effects on reward- and punishment-based reversal learning. Following this prior work, we were specifically interested in the drug effect on the balance between (reversal) learning from reward and from punishment. Therefore, relative reversal learning scores were calculated, by which accuracy scores on punishment-based switch trials were subtracted from accuracy scores on reward-based switch trials. The additional advantage of this method is that it controls for within-subject variability due to other factors such as arousal, attention and motivation, which would have affected each measure in the same direction. Thus, general effects of the drug that were not specific to the ability to learn from reward or punishment were subtracted out. Further, we report drug effects (differences between the placebo and the bromocriptine session), because it is these drug effects that are of primary interest in the present study. Finally, we also report reward- and punishment-based reversal learning under placebo and under bromocriptine separately.

**Statistical analysis.** Mean proportions of correct responses on the learning task were calculated as reported previously (Cools et al., 2006, 2008a). Repeated measures ANOVAs were conducted using SPSS 15.0 with drug and valence as within-subject factors and dopamine synthesis capacity as a covariate. The delay between acquisition of the drug and PET data was also entered as a covariate. Pearson product-moment correlation coefficients were calculated between \( K_i \)-values extracted from ROIs and behavioral data. All correlations represent partial correlations, correcting for the delay between PET and drug data acquisition. All correlations were also significant without this correction. The distribution of none of the parameters assessed here deviated from normality as indicated by Kolmogorov–Smirnov tests (all \( p = 0.2 \)).

All reported \( p \) values are two-tailed.

**Results** In our sample of young healthy volunteers, influx constant \( K_i \) values varied between 0.018 and 0.027, falling well within the range of “normal” values observed previously (Eberling et al., 2007). Subjects performed well on the reversal learning task, with an average accuracy rate on trials after the unexpected outcomes >90% (supplemental Table 2, available at www.jneurosci.org as supplemental material).

First, we analyzed the data from the placebo session. A repeated measures ANOVA was conducted with valence as the within-subject factor and synthesis capacity and acquisition delay as covariates. Consistent with neurophysiological evidence from nonhuman primates (Hollerman and Schultz, 1998), this analysis revealed a highly significant interaction between valence and synthesis capacity (\( F(1.8) = 19.0, p = 0.002 \)) and no effects of acquisition delay. This interaction reflected a positive correlation between dopamine synthesis capacity in the striatum and reversal learning from reward relative to punishment under placebo (Fig. 2a). It was present across the entire striatum (averaged across right and left caudate nucleus and putamen; \( r_e = 0.84, p = 0.004 \)), and also within striatal subregions (bilateral caudate nucleus: \( r_e = 0.8, p = 0.007 \); bilateral putamen: \( r_e = 0.87, p = 0.001 \)). The correlation between dopamine synthesis capacity and relative performance on non-switch trials (trials requiring reward-prediction minus trials requiring punishment-prediction) was also positive, albeit non-significant (entire striatum: \( r_e = 0.54, p = 0.1 \)).

The positive correlation between synthesis capacity and relative reversal learning scores (i.e., the difference between reward and punishment) was driven by a positive correlation between synthesis capacity and reward-based reversal under placebo (accuracy: \( r_e = 0.79, p = 0.007 \)), indicating that greater dopamine synthesis capacity predicted better reward-based reversal. Conversely, punishment-based reversal under placebo did not depend on baseline dopamine synthesis capacity (accuracy: \( r_e =
Baseline dopamine predicts drug effects on reversal learning

Cools et al. • Baseline Dopamine Predicts Drug Effects on Reversal Learning

Hollerman and Schultz, 1998), given that synthesis capacity likely influences the efficacy of impulse-dependent phasic dopamine release. The finding that the effect did not extend to learning from unexpected punishment suggests that synthesis capacity did not influence the efficacy of the impulse-dependent pause in dopamine firing that accompanies unexpected reward omission (Hollerman and Schultz, 1998).

Next we assessed whether baseline striatal dopamine synthesis capacity also predicted the effects of the dopamine D2 receptor agonist bromocriptine on reversal learning. To this end, we conducted a repeated measures ANOVA with drug and valence as within-subject factors and dopamine synthesis capacity and acquisition delay as covariates. As predicted, this analysis revealed highly significant two-way drug by valence (F(1,7) = 17.4, p = 0.004) and three-way drug by valence by synthesis capacity interactions (F(1,7) = 29.4, p = 0.001). There were no main effects (acquisition delay: F(1,7) = 0.2, p = 0.7; synthesis capacity: F(1,7) = 3.0, p = 0.1; valence: F(1,7) = 0.8, p = 0.4; drug: F(1,7) = 0.03, p = 0.9) and no other interaction effects (valence by synthesis capacity: F(1,7) = 0.6, p = 0.5; valence by delay: F(1,7) = 0.3, p = 0.6; drug by synthesis capacity: F(1,7) = 0.2, p = 0.7; drug by delay: F(1,7) = 0.003, p = 0.96; drug by valence by delay: F(1,7) = 2.4, p = 0.2). The significant three-way interaction reflected a significant negative correlation between synthesis capacity and drug-induced improvement on relative reversal learning scores (rT = -0.9, p = 0.001) (Fig. 2b). Consistent with an ‘inverted u’-shaped dose–response curve, bromocriptine improved reward-based reversal relative to punishment-based reversal in subjects with low baseline levels of striatal dopamine synthesis capacity, but had the reverse effect in subjects with high baseline levels. Again the effect extended across striatal subregions (bilateral caudate nucleus: rT = -0.89, p = 0.001; bilateral putamen: rT = -0.89, p = 0.001). The correlation with (relative) performance on non-switch trials was not significant (rT = -0.45, p = 0.2).

Breakdown of the three-way interaction effect into simple interaction effects for reward and punishment separately revealed a significant interaction between drug and synthesis capacity for punishment-based reversal (F(1,8) = 14.2, p = 0.007), as well as a near-significant interaction between drug and synthesis capacity for reward-based reversal (F(1,8) = 3.4, p = 0.1). These interactions reflected a highly significant positive correlation between striatal dopamine synthesis and drug-induced improvement in punishment-based reversal (rT = 0.8, p = 0.007) (Fig. 3b,df), while the negative correlation between dopamine synthesis and drug effects on reward-based reversal was less convincing (rT = -0.55, p = 0.1) (Fig. 3a,ce).

In supplementary analyses, we aimed to disentangle two alternative hypotheses regarding dopaminergic modulation. Specifically, to establish whether the here described effects reflect a modulation of learning or switching, we applied computational reinforcement learning algorithms to fit individual subjects’ trial-by-trial sequence of choices (Sutton and Barto, 1998; Frank et al., 2007b). These algorithms allowed us to generate learning-rate parameters (separately for reward and punishment) that were not directly observable in the behavioral data. Detailed methods and results are presented in the supplemental Materials, available at www.jneurosci.org. Critically, a significant relationship was obtained between dopamine synthesis and the drug effect on reward learning rate (rT = -0.71, p = 0.02), as well as between dopamine synthesis and the drug effect on punishment learning rate (rT = 0.78, p = 0.01) (supplemental Figure and Table 3, available at www.jneurosci.org as supplemental material).

In summary, higher dopamine synthesis capacity in the striatum was associated with better reward-based reversal learning under placebo. Furthermore, bromocriptine improved reward-based reversal learning in subjects with low synthesis capacity, while impairing it in subjects with high synthesis capacity. Conversely, the same drug dose impaired punishment-based reversal learning in subjects with low synthesis capacity, while improving it in subjects with high synthesis capacity.

Discussion

Baseline dopamine measures predicted reversal learning due to reward prediction errors relative to punishment prediction errors. The result provides the first empirical evidence for the pervasive, but hitherto untested hypothesis that individual variation in reward-based learning reflects quantitative variation in baseline levels of striatal dopamine function, as indexed by uptake of a PET dopamine synthesis tracer. Critically, the effect was outcome-specific, so that high dopamine synthesis was associated with a bias toward reward relative to punishment-based reversal learning. This observation concurs with recent theoretical mod-
baseline dopamine synthesis capacity and punishment-based reversal learning under placebo. C. Nonsignificant correlation between striatal dopamine synthesis capacity and reward-based reversal learning under bromocriptine ($r = -0.15$). D. Significant positive correlation between striatal dopamine synthesis capacity and punishment-based reversal learning under bromocriptine ($r = 0.78, p = 0.015$). E. Nonsignificant negative correlation between striatal dopamine synthesis capacity and the effect of bromocriptine on reward-based reversal learning (bromocriptine minus placebo). F. Significant positive correlation between striatal dopamine synthesis capacity and the effect of bromocriptine on punishment-based reversal learning (bromocriptine minus placebo). For statistics, see Results.

Figure 3. Baseline dependency of absolute reversal learning scores and their sensitivity to D2 receptor stimulation. A. Significant positive correlation between striatal dopamine synthesis capacity ($K$-values) and reward-based reversal learning under placebo. B. Nonsignificant correlation between striatal dopamine synthesis capacity and punishment-based reversal learning under placebo.
and remained highly significant after statistical correction for the acquisition delay. Finally, any instability in the PET measurement across the delay would have reduced rather than enhanced the likelihood of obtaining the result. There is a possibility that, if there had been no delay, the correlation between synthesis capacity and behavioral data might have been even stronger. Therefore, the reported correlations might represent noisy versions of the true correlations.

Our data elucidate not only a fundamental mechanism underlying the behavioral role of striatal dopamine, but also identify an important neurobiological factor, i.e., baseline striatal dopamine synthesis, that contributes to the large variability in dopaminergic drug efficacy. This finding should have far-reaching implications for individualized drug development in neuropsychiatry, where variable drug efficacy provides a major problem for the treatment of patients with heterogeneous spectrum disorders like schizophrenia, attention deficit/hyperactivity disorder and drug addiction.

References