



Single Dose of a Dopamine Agonist Impairs Reinforcement Learning in Humans: Evidence from Event-related Potentials and Computational Modeling of Striatal-Cortical Function

Citation

Santesso, Diane L., A. Eden Evins, Michael J. Frank, Erika M. Cowman Schetter, Ryan Bogdan, and Diego A. Pizzagalli. 2009. Single dose of a dopamine agonist impairs reinforcement learning in humans: Evidence from event-related potentials and computational modeling of striatal-cortical function. *Human Brain Mapping* 30(7): 1963-1976.

Published Version

doi:10.1002/hbm.20642

Permanent link

<http://nrs.harvard.edu/urn-3:HUL.InstRepos:4215080>

Terms of Use

This article was downloaded from Harvard University's DASH repository, and is made available under the terms and conditions applicable to Open Access Policy Articles, as set forth at <http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#OAP>

Share Your Story

The Harvard community has made this article openly available.
Please share how this access benefits you. [Submit a story](#).

[Accessibility](#)

IN PRESS: HUMAN BRAIN MAPPING

Running Head: Dopamine Agonist and Reinforcement Learning

**Single Dose of a Dopamine Agonist Impairs Reinforcement Learning in Humans:
Evidence from Event-related Potentials and Computational Modeling of Striatal-Cortical
Function**

Diane L. Santesso¹, A. Eden Evins², Michael J. Frank³, Erika M. Cowman Schetter¹,
Ryan Bogdan¹, & Diego A. Pizzagalli¹

¹Department of Psychology, Harvard University, Cambridge, Massachusetts

²Addiction and Schizophrenia Clinical and Research Program, Massachusetts General Hospital,
Boston, Massachusetts

³Department of Psychology, University of Arizona, Tucson, Arizona

Please address all correspondence to:

Diego A. Pizzagalli, Ph.D.
Department of Psychology
Harvard University
1220 William James Hall
33 Kirkland Street
Cambridge, MA 02138, USA

Phone: +1-617-496-8896
Fax: +1-617-495-3728
Email: dap@wjh.harvard.edu

Abstract

Animal findings have highlighted the modulatory role of phasic dopamine (DA) signaling in incentive learning, particularly in the acquisition of reward-related behavior. In humans, these processes remain largely unknown. In a recent study we demonstrated that a single low dose of a D2/D3 agonist (pramipexole) – assumed to activate DA autoreceptors and thus reduce phasic DA bursts – impaired reward learning in healthy subjects performing a probabilistic reward task. The purpose of the present study was to extend these behavioral findings using event-related potentials and computational modeling. Compared to the placebo group, participants receiving pramipexole showed increased feedback-related negativity to probabilistic rewards and decreased activation in dorsal anterior cingulate regions previously implicated in integrating reinforcement history over time. Additionally, findings of blunted reward learning in participants receiving pramipexole were simulated by reduced presynaptic DA signaling in response to reward in a neural network model of striatal-cortical function. These preliminary findings offer important insights on the role of phasic DA signals on reinforcement learning in humans, and provide initial evidence regarding the spatio-temporal dynamics of brain mechanisms underlying these processes.

In recent years, the role of dopamine (DA) in reinforcement learning has been strongly emphasized. In particular, electrophysiological studies in non-human primates have shown that midbrain DA neurons code reward-related prediction errors: unpredicted rewards elicit phasic increases in DA neurons as well as phasic DA release (positive-prediction error), whereas omission of a predicted reward elicits phasic DA decreases (negative-prediction error) [Fiorillo et al., 2003; Schultz, 2007]. These phasic DA responses have been assumed to reflect a teaching signal for regions implicated in reward-related learning, including the anterior cingulate cortex (ACC) and basal ganglia [Holroyd and Coles, 2002]. Accordingly, when a positive prediction error occurs, learning about the consequences of the behavior that led to reward takes place; when a negative prediction error occurs, behaviors that led to lack of reward is extinguished ([Bayer and Glimacher, 2005; Garris et al., 1999; Montague et al., 1996; Schultz et al., 1997]; for findings highlighting the role of DA signaling in instrumental learning, see [Cheng and Feenstra, 2006; Reynolds et al., 2001; Robinson et al., 2007; Schwabe and Koch, 2007]). Based on these findings, disruption of phasic DA responses is expected to negatively impact prediction error and, thus, reduce reinforcement learning [Montague et al., 2004; Schultz, 2002].

Evidence from the animal literature indicates that single low doses of D2 agonists suppress DA cell firing rates through autoreceptor stimulation [Fuller et al., 1982; Martres et al., 1977; Piercey et al., 1996; Sumners et al., 1981; Tissari et al., 1983]. In humans, Frank and O'Reilly [2006] reported that low doses of the D2 agonist cabergoline impaired the ability to optimize responding based on probabilistic reward values, without affecting negative feedback learning. More recently, Pizzagalli et al. [2008] reported that administration of a single dose (0.5 mg) of a D2/D3 agonist (pramipexole) to healthy subjects blunted reinforcement learning during a probabilistic signal detection task in which correct responses to two stimuli were differentially

rewarded. The goal of the present study was two-fold. First, we aimed to extend our recent behavioral findings [Pizzagalli et al., 2008] by examining the effects of a single dose of pramipexole on electrophysiological correlates of reward learning in a sub-group of these participants with reliable event-related potential (ERP) data. To this end, the feedback-related negativity (FRN) and current source density underlying the FRN were used as indices of learning from positive feedback. The second goal was to apply a computational modeling of striatal-cortical function [Frank, 2005] on the behavioral findings described by Pizzagalli et al. [2008] and evaluate whether blunted reward learning could be explained by reduction of presynaptic DA bursts (i.e., reduced positive prediction error), as originally postulated.

Emerging evidence implicates various prefrontal cortex (PFC) regions in adaptive reward-related decision-making, but also highlights important functional dissociations. In functional neuroimaging studies, medial PFC regions spanning into the rostral ACC (i.e., Brodmann areas 10/32) have been implicated in response to immediate, but not delayed, reward [Knutson et al., 2003; McClure et al., 2004] and has been found to track the value of reward [Daw et al., 2006; Marsh et al., 2007]. The dorsal ACC (dACC), on the other hand, has been implicated in experimental tasks requiring representation of both gains and losses, and in integrating reinforcement history across several trials [Akitsuki et al., 2003; Ernst et al., 2004; Rogers et al., 2004]. Similar findings have emerged from animal studies. Hadland et al. [2003], for example, found that ACC lesions impaired monkeys' ability to select actions based on prior reinforcers, but did not impair stimulus-reward associations. In a critical extension, Kennerley et al. [2006] showed that lesions of the ACC impaired performance on a task requiring integration of reinforcement history over several trials. Thus, whereas animals with ACC lesions responded similarly to control animals on single error trials, they failed to integrate reinforcement history

over time and were thus unable to learn which response was more advantageous. Similarly, Amiez et al. [2006] found that activity in macaque ACC neurons encoded the weighted probabilistic value of available rewards. Collectively, these findings emphasize a role of the dACC in representing the reinforcement history and integrating action-outcome patterns over time to guide goal-directed behavior [Rushworth et al., 2007].

Whereas hemodynamic neuroimaging approaches provide valuable information about brain circuitries implicated in reward-based decision-making, their limited temporal resolution precludes investigation of the temporal unfolding of underlying brain mechanisms. High temporal resolution is particularly important when considering that phasic activation of DA operate on a time course of tens of milliseconds [Schultz, 2002]. The FRN – a negative ERP deflection peaking approximately 200-400 ms following feedback with a frontocentral scalp distribution – offers a non-invasive index of activity in the medial PFC implicated in reward learning. The generator of the FRN has been localized to the ACC: early dipole localization studies implicate the dACC [Miltner et al., 1997; Gehring and Willoughby 2002], whereas more recent ERP/fMRI studies implicate medial PFC regions [Muller et al., 2005; Nieuwenhuis et al., 2005; Van Veen et al., 2004].

The functional significance of the FRN is unclear. Initial observations that the FRN is increased by negative feedback or when outcomes are worse than expected led to the assumption that the FRN reflects a reward prediction error signal [Gehring and Willoughby, 2002; Holroyd and Coles 2002; Miltner et al., 1997]. Recent findings from probabilistic selection [Hajcak et al., 2005; Muller et al., 2005], gambling [Yeung and Sanfey, 2004; Donkers et al., 2005], and time estimation [Nieuwenhuis et al., 2005] tasks indicate, however, that the FRN is also modulated by positive feedback or experimental settings in which outcomes are better than expected.

Importantly, recent evidence suggests that the key variable may not be the valence of the feedback but rather its predictability. Using an anticipating timing task, Oliveira et al. [2007], for example, demonstrated that the FRN was elicited by *unexpected* positive as well as negative feedback. Critically, a large FRN emerged only when the feedback did not match participants' estimation of task performance, including when participants received positive feedback after having estimated their performance to be incorrect. Along similar lines, Muller et al. [2005] reported that equivocal (unexpected) feedback elicited a larger (i.e., more negative) FRN compared to negative feedback ($-7.5 \mu\text{V}$ vs. $-2.4 \mu\text{V}$). Based on these findings, Muller and coworkers argued that the FRN might reflect the rapid evaluation of behavior from external cues (whether it is positive, negative, or uninformative feedback), and that the FRN is enhanced under conditions in which feedback serves to guide performance on stimulus-response mapping tasks. Of interest, source localization analyses of the FRN finding revealed that a number of regions in the mPFC/ACC were involved in processing the information conveyed by feedback stimuli throughout learning [Muller et al., 2005]. Finally, Holroyd and Coles [2008] recently showed that the FRN can be modulated by positive prediction errors. Specifically, the authors used a two-choice response task in which the correct response was not clearly defined – participants had to infer the optimal response strategy by trial and error. When participants who had adopted a disadvantageous strategy occasionally choose the better option and thus earned more reward, a more positive FRN was observed. According to these authors, positive and negative prediction errors can decrease and increase, respectively, the size of the FRN. Taken together, results from recent ERP studies indicate that positive prediction errors can affect the FRN (specifically, reduce its negativity); more generally, these findings are consistent with the account that unpredicted rewards, supported by phasic DA increases [Fiorillo et al., 2003; Schultz, 2007], also

serve as teaching signal for ACC and basal ganglia to optimize goal-directed behaviors [Holroyd and Coles, 2008].

The FRN has been compared to another ACC-generated negativity - the error-related negativity (ERN), which occurs after error commission and is thought to reflect internally driven error detection, conflict monitoring, and affective reactions to errors [Yeung et al., 2004; Luu et al., 2000]). In a recent study, Zirnheld et al [2004] assessed the effect of the D2/D3 receptor antagonist haloperidol on ENR amplitude, and observed that haloperidol impaired learning and diminished the ERN on a time estimation task; these findings were later replicated by De Bruijn and colleagues [2006] using a flanker task. These authors suggested that haloperidol impaired DA signaling such that the phasic DA dip following negative outcomes (i.e., errors) was reduced. Collectively, these studies suggest that reward prediction errors from both internal (errors) and external (feedback) cues may be similarly sensitive to DA manipulation.

The first goal of the present study was to extend these ERP findings by investigating the effects of a single dose of pramipexole on the FRN. As in recent studies in non-human primates [Amiez et al., 2006; Kennerley et al., 2006], participants in the present study were confronted with a choice between two responses associated with different probabilities of reward. Owing to the probabilistic nature of this task, subjects were not able to infer which stimulus was more advantageous based on the outcome of single trials but needed to consider the reinforcement history to optimize their behavioral choices. As mention above, we recently demonstrated that a single dose of pramipexole led to blunted reward learning and reduced “win-stay” strategy (i.e., a reduced propensity to select a more advantageous stimulus after it had been rewarded in the preceding trial) [Pizzagalli et al., 2008]. Based on prior animal [Fuller et al., 1982; Martres et al., 1977; Piercey et al., 1996; Sumners et al., 1981; Tissari et al., 1983] and limited human [Frank

and O'Reilly, 2006], we postulated that these impairments were due to decreased phasic DA bursts to unpredicted reward (i.e., reduced positive prediction errors) leading to reduced ability to learn about the consequences of the behavior leading to the positive outcome [Schultz, 2007]. The second goal of this study was to test this hypothesis by investigating whether blunted reward learning in participants receiving pramipexole could be simulated by reduced presynaptic DA signaling in response to reward in a neural network model of striatal-cortical function [Frank, 2005]. Based on prior findings, we hypothesized that, compared to subjects receiving placebo, those receiving pramipexole will display *larger* (i.e., more negative) FRNs due to (1) blunted reward learning resulting in greater reward expectancy violations [Oliveira et al., 2007]; (2) reduced positive prediction error [Holroyd and Coles, 2008]; and/or (3) over-reliance of feedback information to guide performance [Muller et al., 2005]. Moreover, participants receiving pramipexole were expected to show decreased activation in brain regions that integrate reinforcement histories and action-outcome patterns across time, particularly the dACC [Amiez et al., 2006; Akitsuki et al., 2003; Ernst et al., 2004; Holroyd and Coles, 2008; Kennerley et al., 2006; Rogers et al., 2004]. Finally, we expected that blunted reward learning after pramipexole administration could be modeled through reduced phasic DA bursts in response to reward.

Method

Participants

Thirty-two participants were recruited from the community for a larger study investigating the effects of a D2 agonist on reward, motor, and attentional processes as well as mood [Pizzagalli et al., 2008]. After an initial phone screening, subjects were invited to the laboratory for a Structured Clinical Interview for the DSM-IV (SCID) [First et al., 2002], which was conducted by a research psychiatrist or master-level interviewer. Subjects meeting the following

exclusionary criteria were excluded: current unstable medical illness; pramipexole contraindications; any past or current Axis I psychiatric disorders; presence of any neurological disorder or dopaminergic abnormality; pregnancy or breast-feeding; use of prescription or over-the-counter medications in the past week that may interact with the metabolism of pramipexole; use of dopamine antagonists in the past month; use of any CNS depressant in the past 24 hours that might affect reward responsiveness, including anti-histamines and alcohol; and history in first-degree relatives of psychological disorders involving dopaminergic abnormalities (schizophrenia, psychosis, schizotypal personality disorder, bipolar disorder, major depression, substance dependence). To minimize side effects, a Body Mass Index (BMI) of at least 20 was used. In light of potential changes in dopaminergic sensitivity during the menstrual cycle, all female participants performed the experimental session during days 1-14 of their menstrual cycle [Myers et al., 2003].

From the 32 enrolled participants, 20 had usable ERP data (13 men; mean age = 25.20, $SD = 3.38$); data from remaining participants were lost due to insufficient number of artifact-free ERP data, equipment failure, and/or emergence of adverse drug effects (see [Pizzagalli et al., 2008] for more detail). Among these 20 participants, 13 received placebo and seven received pramipexole. All participants were right-handed [Chapman and Chapman, 1987]. The placebo and pramipexole groups did not differ with respect to gender ratio (8 males/5 females vs. 5 males/2 females; Fisher's Exact test: $P > 0.30$), age (24.85 ± 3.24 vs. 25.86 ± 3.80 years; $t_{18} = -0.63$, $P > 0.52$, two-tailed). In addition, the placebo and pramipexole groups did not differ on self-reported depressive symptoms (1.15 ± 1.67 vs. 0.43 ± 0.79 , $P > .29$) or trait anxiety (32.22 ± 5.91 vs. 26.40 ± 3.05 , $P > .06$), as assessed by the Beck Depression Inventory-II (BDI-II)

[Beck et al, 1996] and the trait form of the Spielberger State-Trait Anxiety Inventory (STAI) [Spielberger et al., 1970], respectively.

Participants received \$10/hour for the SCID session, \$40 for the experimental session, and \$24.60 in earnings in the probabilistic reward task. All participants provided written informed consent after a psychiatrist fully explained the experimental protocol, which had been approved by the Committee on the Use of Human Subjects at Harvard University as well as the Partners-Massachusetts General Hospital Internal Review Board.

Pharmacological manipulation

Pramipexole dihydrochloride and placebo were administered in a randomized, double-blind design. Participants in the pramipexole group were administered 0.5 mg pramipexole in capsule form, while those in the placebo group were administered an identical capsule. ERP recording was conducted approximately 2 hours after drug administration when pramipexole reaches peak concentration [Wright et al., 1997].

Data collection and reduction

Probabilistic reward task. Participants completed a task [Pizzagalli et al., 2005] that consisted of 300 trials, divided into 3 blocks of 100 trials, separated by 30-sec breaks. Each trial started with the presentation of a fixation point for 1450 ms in the middle of the screen. A mouthless cartoon face was presented for 500 ms followed by the presentation of this face with either a short mouth or a long mouth for 100 ms. Participants were asked to identify whether a short or long mouth was presented by pressing either the “z” key or the “/” key on the keyboard (counter-balanced across participants). In each block, an equal number of short and long months

were presented within a pseudorandomized sequence. For each block, only 40 correct responses were followed by positive feedback (“Correct! You won 20 cents”). To induce a response bias, an asymmetrical reinforcer ratio was used: correct responses for one stimulus (“rich stimulus”) were rewarded three times (30:10) more frequently than correct responses for the other stimulus (“lean stimulus”). Positive feedback was presented for 1500 ms followed by a blank screen for 250 ms, and participants were instructed that not all correct responses would receive reward feedback. Trials without feedback were timed identically (i.e., mouth onset to the next trials’ fixation) to those with feedback.

Following prior studies [Davison and Tustin, 1978; Pizzagalli et al., 2005], response bias ($\log b$) and discriminability ($\log d$) were computed as:

$$\log b = \frac{1}{2} \log \left(\frac{Rich_{correct} * Lean_{incorrect}}{Rich_{incorrect} * Lean_{correct}} \right)$$

$$\log d = \frac{1}{2} \log \left(\frac{Rich_{correct} * Lean_{correct}}{Rich_{incorrect} * Lean_{incorrect}} \right)$$

As evident from the formula, response bias incorporates responses to both the rich and lean stimulus, and increases if participants tend to (1) correctly identify the rich stimulus, and/or (2) misclassify the lean stimulus as the rich stimulus.

Scalp event-related potentials. EEG was recorded continuously using a 128-channel Electrical Geodesics system (EGI Inc., Eugene, OR) at 250 Hz with 0.1-100 Hz analog filtering referenced to the vertex. Impedance of all channels was kept below 50 k Ω . Data were segmented and re-referenced off-line to an average reference. EEG epochs were extracted beginning 200 ms before and ending 800 ms after feedback presentation on correct trials during blocks 2 and 3 for the midline sites Fz, FCz, Cz, Pz (sensors 11, 6, 129, 62). Only ERP data from blocks 2 and 3

were used to allow participants to be exposed to the differential reinforcement schedule. Data was processed using Brain Vision Analyzer (Brain Products GmbH, Germany). Each trial was visually inspected for movement artifacts and automatically removed with a $\pm 75 \mu\text{V}$ criterion. Eye-movement artifacts were corrected by Independent Component Analysis. The amplitude of the ERP was derived from each individual's average waveform and filtered at 1-30 Hz. The FRN was scored manually for each subject at each site using a pre-stimulus baseline between -200-0 ms and a base-to-peak approach (see e.g., [Hajcak et al., 2007]). The FRN was defined as the most negative peak 200-400 ms after feedback presentation. In addition, to evaluate potential group differences in other stages of the information processing flow, EEG epochs beginning 200 ms before and ending 800 ms after stimulus presentation (short or long mouth) were extracted from blocks 2 and 3. N1 amplitude was defined as the most negative peak 70-130 ms after stimulus onset and P3 amplitude was defined as the most positive peak 300-500 ms after stimulus onset. A pre-stimulus baseline between -200-0 ms was used.

Source localization analyses. Low Resolution Electromagnetic Tomography (LORETA; [Pascual-Marqui et al., 1999]) was used to estimate intracerebral current density underlying the FRN. The LORETA algorithm is a form of Laplacian-weighted minimal norm solution that solves the inverse problem by assuming that: (1) neighboring neurons are synchronously activated and display only gradually changing orientations; and (2) the scalp-recorded signal originates mostly from cortical gray matter. Unlike other source localization techniques (e.g., dipole modeling), the LORETA algorithm does not assume an a priori number of underlying sources to solve the inverse problem. Independent validation for the algorithm has been derived from studies combining LORETA with fMRI [Mulert et al., 2004; Vitacco et al., 2002], PET ([Pizzagalli et al., 2004]; but see [Gamma et al., 2004]), and intracranial recordings [Zumsteg et

al., 2005]. In two recent studies, LORETA localizations were, on average, 16 mm [Mulert et al., 2004] and 14.5 mm [Vitacco et al., 2002] from fMRI activation loci, a discrepancy within the range of the LORETA's estimated spatial resolution (~1-2 cm).

For the present study, a three-shell spherical head model registered to the Talairach brain atlas (available as digitized MRI from the Brain Imaging Centre, Montreal Neurological Institute) and EEG electrode coordinates derived from cross-registrations between spherical and realistic head geometry [Towle et al., 1993] were used. The solution space (2,394 voxels; voxel resolution: 7 mm³) was constrained to cortical gray matter and hippocampi, which were defined according to a digitized probability atlas provided by the MNI (i.e., coordinates reported in main text are in MNI space). Based on this probability atlas, a voxel was labelled as gray matter if its probability of being gray matter was higher than 33% and higher than the probability of being white matter or cerebrospinal fluid. After converting MNI coordinates into Talairach space [Brett et al., 2002], the Structure-Probability Maps atlas [Lancaster et al., 1997] was used to identify gyri and Brodmann area(s).

In the present analyses, current density was computed within a 140-276 ms post-feedback time window, which captured the mean peak latency of the FRN at Cz (232 ms) on correct trials. At each voxel, current density was computed as the linear, weighted sum of the scalp electric potentials (units are scaled to amperes per square meter, A/m²). For each subject, LORETA values were normalized to a total power of 1 and then log-transformed before statistical analyses.

Physical adverse effects. Throughout the session, participants were asked to indicate the extent to which they experienced 12 physical symptoms using a 5-point Likert scale. The symptoms assessed were headache, cold or chilled, hot or flushed, dizziness, sleepiness, sweating, blurred vision, nausea, fast heartbeat, dry mouth, abdominal pain, and diarrhea.

A total adverse effect score was obtained by subtracting the pre-administration score from the maximal adverse effect score (see [Pizzagalli et al., 2008] for more detail).

Statistical analyses

The FRN data were analyzed using a mixed ANOVA with *Group* as between-subject factor and *Site* (Fz, FCz, Cz) as repeated measure. When applicable, the Greenhouse-Geisser correction was used. Follow-up independent t-tests (two-tailed) were performed to decompose significant effects. Pearson correlations were performed among the variables. For the LORETA data, the groups were contrasted on a voxel-wise basis using unpaired t-tests. Based on prior studies using permutation procedures to determine an experiment-wide alpha level protecting against Type I error, statistical maps were thresholded at $P < 0.005$ and displayed on a standard MRI template [Pizzagalli et al., 2001].

Control analyses. Separate multiple regression analyses were conducted to ensure that physical adverse effects were not contributing to significant findings. Total adverse effect score was entered in the first step followed by group (dummy-coded) in the second step in analyses predicting FRN or LORETA data. Finally, to evaluate whether the two groups differed in other steps of information processing, separate ANOVAs with *Site* (Fz, FCz, Cz, Pz) and *Group* (2) as factors were conducted on the stimulus-locked N1 and P3 data.

Computational modeling

A computational neural network model of striato-cortical function [Frank, 2005] was used to simulate the behavioral results reported in our recent study [Pizzagalli et al., 2008]. The model, which includes “Go” and “NoGo” striatal populations for learning to facilitate rewarding

responses and suppress others (Figure 1), has been applied to several other tasks and has been corroborated by empirical and pharmacological data [Frank and O'Reilly, 2006; Frank et al., 2004]. The core network parameters were left unchanged to maintain consistency with prior work, but simulations were conducted with a more recent model that includes the subthalamic nucleus [Frank, 2006]. The currently implemented model included explicit inhibitory interneurons to regulate overall striatal activity, instead of the “k winners take all” mathematical approximation to inhibitory effects.

Simulating the probabilistic signal detection task. Networks were trained on an analogous reward responsiveness task as the one used in humans. These new simulations involved presenting two overlapping stimuli, labelled rich and lean, respectively, to the input layer. Each input stimulus consisted of 4 units, and the two stimuli overlapped by 3 out of 4 units, so that each stimulus had just one unique unit of activation. The purpose of this overlap was to approximate the visual similarity of the rich and lean stimuli. Further, human subjects come in to the task with the ability to perceptually discriminate these stimuli, and are provided with explicit task instructions for selecting rich and lean prior to learning. Because the computational model does not simulate the perceptual/object recognition system, and is primarily focused with reward learning and respond selection, we simulated this pre-task perceptual knowledge and discrimination ability by setting the weights from the unique identifying input unit corresponding to the rich and lean stimulus to the appropriate premotor output unit (R1 for rich, R2 for lean). These pre-set cortico-cortical weights cause the model to be more likely at the outset to activate the rich response for the rich stimulus and the lean response for the lean stimulus. But note that these input to premotor weights do not guarantee that the model selects the appropriate responses in each trial, because: (i) the input stimuli are still highly overlapping, and the overlapping units

begin with random weights to motor and striatal units; (ii) both premotor and striatal unit activity is noisy; and (iii) the input to premotor connections only affect the degree to which one or the other response is initially more biased in premotor cortex – a given response is not reliably executed unless it also receives facilitation from striatal Go signals. The weights from the input layer to the striatum are all randomly initialized, and are modified by subsequent phasic changes in DA [Frank, 2005].

Training. As in the behavioral experiments, three blocks of 100 trials were run in the model. Networks were rewarded (given a DA burst) on 30% of correct responses on rich trials, and on 10% of trials for correct lean responses. During these DA bursts, which involve maximal DA unit firing in intact networks, the Go units that participated in selecting the associated response become transiently more active, while their NoGo counterparts become less active. These transient changes in Go/NoGo activity are accompanied by changes in synaptic plasticity using contrastive Hebbian learning [Frank, 2005], such that Go representations become stronger for responses that are rewarded more frequently as training progresses. Because rewarded trials are very infrequent in this task, a higher learning rate was applied to rewarded trials (0.003; three times that of non-rewarded trials), enabling weights to change by a greater degree in these trials, and to simulate DA effects on synaptic plasticity. Furthermore, the infrequent presentation of rewards was assumed to produce a low reward expectation, and therefore it was assumed that DA “dips” do not occur during reward omissions. However, synaptic plasticity is not “turned off” during reward omissions, and model neurons continue to adjust their connection weights after each experience. Because the cortico-striatal projections and plasticity are somewhat stronger in strength from cortex to the NoGo pathway, as supported by several physiological findings [Berretta et al., 1997, 1999; Kreitzer and Malenka, 2007; Lei et al., 2004], a non-reward

still leads to a small degree of NoGo learning “by default” (even without an explicit DA dip). This mechanism effectively allows the model to learn that lean responses are more often associated with NoGo than rich responses, and is not different between intact and pramipexole simulations.

Simulating presynaptic effects of pramipexole. As discussed above, we posited that the mechanism by which low doses of pramipexole reduced response bias in this task is by stimulating presynaptic D2 autoreceptors and reducing phasic DA firing and release. We therefore simulated pramipexole effects in the model by reducing the magnitude of DA bursts such that firing in DA cells reached only 60% maximal activation, compared with 100% in the intact case. Accordingly, these presynaptic simulations differ from previous simulations of reduced DA in Parkinson's disease [Frank et al., 2004, 2005], in which a proportion of DA units were removed altogether from processing to simulate DA cell damage. Thus, pramipexole networks had a full set of intact DA units, but firing during rewards was simply reduced. The above-described increase in learning rate during rewarded trials in intact networks was also maintained in pramipexole simulations, because it is assumed that increases in DA (and other neuromodulatory signals) during rewarded trials would still enhance learning, allowing us to specifically investigate the effects solely due to weakened effects of Go/NoGo modulation during rewards. (Note that if we also reduced the learning rate in pramipexole networks, the resulting response bias effects would be even stronger, as networks with lower learning rates necessarily learn slower. Thus the current implementation shows that the presynaptic simulation accounts for the impaired response bias effects even without reducing the learning rate, and is therefore a more “fair” test of the proposed mechanism.) Finally, tonic levels remained

unaffected by presynaptic simulations, in keeping with suggestions that only phasic DA is modulated by presynaptic autoreceptor stimulation [Grace, 1995].

Results

Behavioral data

Findings concerning behavioral performance in the probabilistic reward task have been reported in detail in Pizzagalli et al. [2008]. Briefly, response bias was used to measure the systematic preference for the response associated with more frequent rewarded (rich) stimulus, and thus to assess the extent to which behavior was modulated by reinforcement history. Reward learning was calculated by subtracting the response bias for Block 1 from Block 3.

Discriminability provided a measure of the participants' ability to distinguish between the two stimuli. As shown in the left panel of Figure 2, compared with the placebo group, participants receiving pramipexole showed significantly (1) reduced response bias in Block 2 and Block 3 and overall reduced reward learning across blocks; (2) lower accuracy for rich stimuli in Block 3 and higher accuracy for lean stimuli in Blocks 2 and 3; and (3) reduced probability of choosing rich following a rewarded rich stimulus (i.e., "win-stay" strategy). No significant effects emerged for discriminability, suggesting that the two groups did not differ in task difficulty. Significantly reduced response bias and accuracy for the rich stimulus were replicated in the reduced sample size used in the present study (all P 's < .03).

Scalp ERP data

As hypothesized, a main effect for *Group* emerged, due to larger FRN for the pramipexole group than placebo group across sites ($F_{1,18} = 5.47, P = .031$) (Figure 3a,b). Follow-up t-tests indicated that the pramipexole group had larger (i.e., more negative) FRNs compared

with the placebo group at Fz ($t_{18} = 3.37$, $P = .003$) and FCz ($t_{18} = 2.19$, $P = .042$) only (see Table 1).

Source localization data

LORETA was used to estimate intracerebral current density underlying the FRN. As hypothesized, the pramipexole group showed relatively lower activity to the reward feedback stimulus than the placebo group in the dACC (BA 24), a region previously implicated in representing reinforcement histories and integrating action-outcome patterns [Amiez et al., 2006; Kennerley et al., 2006] (see Table 2, Figure 3c). In direct contrast, the pramipexole group showed relatively *higher* activity in mPFC regions (BA 10/11/32) previously implicated in responding to single reinforcements.

Control analyses

Hierarchical regression analyses confirmed that the behavioral, FRN, and LORETA results remained after adjusting for adverse drug side effects (all ΔR^2 s $> .39$, all ΔF s > 11.49 , all P s $< .003$). Moreover, no significant differences between groups were found for the N1 and P3 time-locked to the presentation of the stimulus (all F s > 2.19 , all P s $> .16$; see Table 1). These results provide support that there was no general effect of sedation and stimulus information processing and categorization.

Computational Modeling

The right panel of Figure 2 shows the results of the neural network simulating performance in the probabilistic reward task in an intact model of striato-cortical function

(“placebo group”) and a model incorporating reduced pre-synaptic DA signals in response to rewards (“pramipexole group”). To generate these data, we first approximated the network's discriminability to those of the participants described in Pizzagalli et al. [2008] by tuning the weights from sensory input representations to the corresponding cortical response units (see modeling methods). Specifically, the weight from the unique unit representing rich or lean stimulus to the corresponding cortical response unit was initialized to 0.38. This produced discriminability results that roughly matched those of human subjects (Figure 2b).

Next, we examined corresponding response bias effects in these networks. Intact networks developed rapid increases in response bias even in the very first block (Figure 2b), which continued to increase across blocks due to asymmetric phasic DA burst probabilities for the rich and lean stimuli. This bias was primarily associated with increased accuracy to the rich stimulus, but also with relatively decreased accuracy to the lean stimulus, as training progressed (Figure 2c,d). In contrast, networks with simulated reductions in presynaptic DA bursts during rewards showed overall reduced response bias, and the simulated response bias in blocks 1 and 2 mirrored performance in the pramipexole group. In sum, the current simulations reveal that an *a priori* model of corticostriatal function can capture DA-dependent reward learning biases in the signal detection task, thereby providing an explicit account for reward blunting induced by pramipexole.

Discussion

A large body of animal data has emphasized the modulatory role of reward-related DA signaling in incentive learning, particularly in the acquisition of reward-related behavior [Garris et al., 1999; Reynolds et al., 2001] and expectation of reward [Fiorillo et al., 2003]. Using a probabilistic reward task in conjunction with a pharmacological challenge, we previously

showed that a single, low dose of a D2/3 agonist led to diminished reward responsiveness toward a more frequently reinforced stimulus [Pizzagalli et al., 2008]. Based on the observations that (1) presynaptic D2 receptors have higher affinity for DA than postsynaptic receptors [Cooper et al., 2003]; (2) low doses of D2 agonists stimulate autoreceptors and thus reduce phasic DA releases [Fuller et al., 1982; Martres et al., 1977; Tissari et al., 1983]; and (3) low doses of D2 agonists suppress DA cell firing rates in the ventral tegmental area [Piercy et al., 1996], we originally suggested that blunted reward learning in the pramipexole group might have been due to reduced phasic signaling to the positive feedback ([Pizzagalli et al., 2008]; see also [Frank and O'Reilly, 2006]). This interpretation was supported by the present simulations derived from an *a priori* computational modeling of striatal-cortical function [Frank, 2005], which showed that diminished dopamine burst in the Go learning pathway impaired the ability to learn from positive feedback. In addition, high-density event-related potentials revealed that blunted reward learning was associated with disrupted activation within frontocingulate pathways implicated in integrating reinforcement history over time.

Importantly, sedative effects cannot explain the effect of pramipexole on response bias since no group differences emerged for response time [Pizzagalli et al., 2008], and group differences remained when controlling for adverse effects. In addition, modulation of the FRN did not reflect global DA-induced attenuation in brain activity since the N1 and P3 amplitudes were not affected by pramipexole. Rather, the FRN – an ERP component assumed to generate from DA-mediated prediction error [Holroyd and Coles, 2002, 2008] – was uniquely affected.

Consistent with the behavioral findings of impaired reward learning, the pramipexole group displayed larger (i.e., more negative) FRNs compared to the placebo group. Although the FRN is typically reported following errors and poor performance, the FRN can be elicited by

positive (particularly unexpected) feedback [Hajcak et al., 2005; Holroyd and Coles, 2008; Muller et al., 2005; Nieuwenhuis et al., 2005; Oliveira et al., 2007; Yeung et al., 2004]. There are a few possible explanations for these FRN results. First, Muller et al. [2005] reported that the size of the FRN decreased over time as participants learned a stimulus-response association. They interpreted their finding as suggesting that, as learning progressed, externally-driven feedback was no longer needed to guide performance. Since learning was impaired for participants receiving pramipexole, perhaps they continued to rely on external feedback as indexed by larger (i.e., more negative) FRNs. Unfortunately, there were too few trials to examine the amplitude of the FRN across blocks to test this explanation. Second, pramipexole-induced reduced reward learning might have impaired the participants' inability to predict positive feedback, resulting in greater expectancy violations, and consequently increased FRN. As Oliveira et al. [2007] suggested, the FRN may reflect activity of a general performance monitoring system that detects violations in feedback expectancies, whether good or bad. Third, according to Nieuwenhuis et al. [2005], activity from regions associated with positive and negative feedback create a baseline negativity (or ERP deflection). Activity from distinct areas associated with *positive* feedback (e.g., mPFC and rostral ACC regions) push the baseline negativity in a positive direction, yielding a less negative FRN. In this sense, blunted phasic increases in DA induced by pramipexole might have inhibited this positive push, leading to a more negative FRN in the pramipexole group. This latter interpretation is consistent with recent ERP and modeling data showing that positive prediction errors can reduce the FRN (i.e., diminish its negativity) [Holroyd and Coles, 2008]. In the pramipexole group, blunted positive prediction errors to rewards could have contributed to a relatively more negative FRN compared

to placebo. Future studies will be required to evaluate the relative contributions of these accounts to the present findings.

A substantial body of work derived from human functional neuroimaging and single-unit recordings in animals has emphasized the role of various prefrontal regions in reinforcement-guided decision-making. Recent evidence indicates, however, that the dACC and other mPFC regions may make distinct contributions to reinforcement-guided decision-making. Whereas the dACC has been implicated in integrating action-outcome patterns over time and in mediating the link between previous action-reinforcement histories and the upcoming behavioral choices, mPFC regions (including the OFC) have been shown to be critically involved in the representation of reward values [Rushworth et al., 2007]. Interestingly, in the present study, a low dose of a D2/3 agonist was associated with relatively reduced activation in the dACC but relatively increased activation in more rostral mPFC regions (BA 10,11,32). Of note, in a recent study in a larger sample of healthy adults tested with the same probabilistic reward task, we found that participants failing to develop a response bias had significantly lower dACC activation to reward feedback compared to those developing a bias toward the more frequently rewarded stimulus [Santesso et al., unpublished data]. Moreover, a positive correlation between dACC activation to reward feedback and the ability to develop a response bias emerged. In the present study, exploratory analyses confirmed a positive correlation between dACC activity and reward learning for the pramipexole ($r = .71$, $P = .04$, one-tailed) but not placebo ($r = .31$, $P = .15$, one-tailed) group. Collectively, findings from these two independent studies are consistent with the hypotheses that (1) the dACC plays an important role in representing reinforcement histories to guide adaptive behavior [Amiez et al., 2006; Holroyd and Coles, 2008; Kennerley et al., 2006], and (2) phasic DA bursts act as teaching signals that reinforce rewarding behaviors

[Bayer and Glimacher, 2005; Garris et al., 1999]. Furthermore, the current findings extend recent evidence suggesting that acute DA precursor depletion impaired the ability to preferentially respond to stimuli predicting reward in healthy subjects, a finding that was reversed by L-DOPA administration [Leyton et al., 2007]. Additional studies with larger sample sizes using a DA manipulation are needed to confirm a key role of the dACC in inferring which stimulus is more advantageous based on the reinforcement history [Rushworth et al., 2007].

Several limitations of the present study should be acknowledged. First, a negative feedback condition was not included in the present task. The FRN deflection is larger following negative versus positive feedback, and might be generated by distinct areas in the mPFC/ACC [Nieuwenhuis et al., 2005]. Unfortunately, the design of the present signal detection task precludes the examination of ERP difference waves and/or source localization during positive versus negative feedback processing. Additionally, whereas the present computational modeling indicated that blunted reward learning was reproduced by reduced DA burst in the Go learning pathway disrupting the ability to learn from positive feedback, empirical and modeling data have also emphasized the role of the NoGo pathway in reinforcement learning [Frank et al., 2004, 2006; Sumners et al., 1981]. Along similar lines, the computational model postulated that low doses of pramipexole suppressed DA cell firing rates through D2 autoreceptor stimulation. Although this mechanism successfully modeled behavioral performance and was consistent with prior finding of reduced reward learning after administration of cabergoline, which is more selective for D2 receptors than pramipexole, it is important to emphasize that pramipexole has both D2 and D3 effects; accordingly, it is currently unclear whether D3 receptors may play a role in the effects reported in this study.

Second, although the methodology used in the present study allowed us to investigate the spatio-temporal dynamics of brain mechanisms underlying reinforcement learning with millisecond time resolution, it prevented us from examining brain activation in subcortical regions (e.g., midbrain) as well as interactions between midbrain and cingulate regions. An integration of electrophysiological and hemodynamic neuroimaging techniques will be required for a definite test of temporal unfolding of brain mechanisms underlying reinforcement learning in humans, particularly since animal studies have shown that cingulate neurons can modulate activity in the striatum and midbrain and vice versa [Eblen and Graybiel, 1995; Onn and Wang, 2005]. Interestingly, in humans, DA synthesis capacity in the ventral striatum has been found to positively correlate with BOLD signal in the dACC in response to positive, but not negative, pictures [Sissmeier et al., 2006].

Third, although the LORETA algorithm has received important cross-modal validation, the coarse spatial resolution of this source localization technique (1-2 cm) as well as the use of a spherical head model (as opposed to realistic head models derived from individual subjects' MRI) represent further limitations of this study. We note, however, that the present findings of relatively decreased dACC activation and relatively increased mPFC activation in the pramipexole group are consistent with recent hemodynamic findings showing that (1) administration of a DA antagonist decreased BOLD signal in the dACC during the anticipation of a potential reward [Ablner et al., 2007] and increased mPFC activation compared to both amphetamine and placebo administration [Menon et al., 2007]; and (2) administration of DA-enhancing drugs (amphetamine, cocaine) increased cerebral blood flow [Jenkins et al., 2004], glucose metabolism [Vollenweider 1998] and BOLD signal [Breiter et al., 1997] in the dACC. Finally, the sample size of the present study was small due, in part, to participant withdrawal

from the side effects of pramipexole. Consequently, these results should be interpreted with caution and replicated with a larger sample size. Finally, it will be important to replicate these findings using a crossover design (see e.g., [de Bruijn et al., 2004]) to control for potential group differences on demographic or mood variables not considered here.

In sum, the present study provides converging behavioral, electrophysiological, and computational modeling evidence highlighting the critical role of phasic DA signaling and dACC regions in reinforcement learning in humans. These preliminary results suggest that learned response-outcome associations relies on the dACC, which due to its contribution to control of motor behavior and use of DA-reinforcement signals might guide adaptive behavior by integrating reinforcement history and selecting the optimal stimulus. These findings do not only provide initial information about the spatio-temporal dynamics of brain mechanisms underlying reinforcement learning in humans but offer an useful framework for testing the role of dysfunctional DA pathways in various forms of psychopathology, including substance abuse, schizophrenia, and depression.

Acknowledgments

This work was supported by grants from NIMH (R01 MH68376; DAP) and Harvard College Research Program (ECS). Dr. Frank was supported by National Institute on Drug Abuse grant DA022630. The authors would like to thank Dr. Catherine Fullerton, Melissa Culhane, and Avram Holmes for their assistance with subject recruitment and clinical interviews, as well as Petra Pajtas and Kyle Ratner for their skilled assistance with the project.

Disclosure/Conflict of Interest

Dr. Pizzagalli has received research support from GlaxoSmithKline and Merck & Co., Inc. for projects unrelated to the present study. Dr. Evins has received research grant support from Janssen Pharmaceutica, Sanofi-Aventis, Astra Zeneca; research materials from GSK and Pfizer, and honoraria from Primedia, Inc. Moreover, Dr. Evins is an investigator in a NIDA-funded collaborative study with GSK. Dr. Santesso, Dr. Frank, Ms. Cowman Schetter, and Mr. Bogdan report no competing interests.

References

- Abler B, Erk S, Walter H (2007): Human reward system activation is modulated by a single dose of olanzapine in healthy subjects in an event-related, double-blind, placebo-controlled fMRI study. *Psychopharmacology* 191: 823-833.
- Akitsuki Y, Sugiura M, Watanabe J, Yamashita K, Sassa Y, Awata S, Matsuoka H, Maeda Y, Matsue Y, Fukuda H, Kawashima R (2003): Context-dependent cortical activation in response to financial reward and penalty: an event-related fMRI study. *Neuroimage* 19: 1674-1685.
- Amiez C, Joseph JP, Procyk E (2006): Reward encoding in the monkey anterior cingulate cortex. *Cereb Cortex* 16: 1040-1055.
- Bayer HM, Glimcher PW (2005): Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* 47: 129-141.
- Beck AT, Steer RA, Brown GK (1996): *Beck Depression Inventory Manual*, 2nd ed. San Antonio, TX: The Psychological Corporation.
- Berretta S, Parthasarathy HB, Graybiel AM (1997): Local release of GABAergic inhibition in the motor cortex induces immediate-early gene expression in indirect pathway neurons of the striatum. *J Neurosci* 17: 4752-4763.
- Berretta S, Sachs Z, Graybiel AM (1999): Cortically driven Fos induction in the striatum is amplified by local dopamine D2-class receptor blockade. *Eur J Neurosci* 11: 4309-4319.
- Breiter HC, Gollub RL, Wiesskoff RM, Kennedy DN, Makris N, Berke JD, Goodman JM, Kantor HL, Gastfriend DR, Riorden JP, Mathew RT, Rosen BR, Hyman SE (1997): Acute effects of cocaine on human brain activity and emotion. *Neuron* 19: 591-611.
- Chapman LJ, Chapman JP (1987): The measurement of handedness. *Brain Cognition* 6:175-83.

- Cheng J, Feenstra MG (2006): Individual differences in dopamine efflux in nucleus accumbens shell and core during instrumental learning. *Learn Mem* 13: 168-177.
- Cooper JR, Bloom FE, Roth RH (2003): *The biochemical basis of neuropharmacology*, 8th Ed. Oxford: Oxford University Press.
- Davison MC, Tustin RD (1978): The relation between the generalized matching law and signal-detection theory. *J Exp Anal Behav* 29: 331-336.
- Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ (2006): Cortical substrates for exploratory decisions in humans. *Nature* 441: 876-879.
- de Bruijn ER, Hulstijn W, Verkes RJ, Ruigt GS, Sabbe BG (2004): Drug-induced stimulation and suppression of action monitoring in healthy volunteers. *Psychopharmacology* 177: 151-60.
- Donkers FCL, Nieuwenhuis S, van Boxtel GJM (2005): Medial frontal negativities in the absence of responding. *Cog Brain Res* 25: 777-787.
- Eblen F, Graybiel AM (1995): Highly restricted origin of prefrontal cortical inputs to striosomes in the macaque monkey. *Neurosci* 15: 5999-6013.
- Ernst M, Nelson EE, McClure EB, Monk CS, Munson S, Eshel N, Zarah E, Leibenluft E, Zametkin A, Towbin K, Blair J, Charney D, Pine DS (2004): Choice selection and reward anticipation: an fMRI study. *Neuropsychologia* 42: 1585-1597.
- Fiorillo CD, Tobler PN, Schultz W (2003): Discrete coding of reward probability and uncertainty by dopamine neurons. *Science* 299: 1898-1902.
- First MB, Spitzer RL, Gibbon M, Williams JBW (2002): *Structured Clinical Interview for DSM-IV Axis I Disorders, Research Version, Patient Edition (SCID-I/P)*. New York: Biometrics Research, New York State Psychiatric Institute.

- Frank MJ (2005): Dynamic dopamine modulation in the basal ganglia: A neurocomputational account of cognitive deficits in medicated and non-medicated Parkinsonism. *J Cogn Neurosci* 17: 51-72
- Frank M.J (2006): Hold your horses: a dynamic computational role for the subthalamic nucleus in decision making. *Neural Netw* 19: 1120-1136.
- Frank MJ, O'Reilly RC (2006): A mechanistic account of striatal dopamine function in human cognition: Psychopharmacological studies with cabergoline and haloperidol. *Behav Neurosci* 120: 497-517.
- Frank MJ, Seeberger LC, O'Reilly RC (2004): By carrot or by stick: Cognitive reinforcement learning in Parkinsonism. *Science* 306: 1940-1943.
- Fuller RW, Clemens JA, Hynes MD^{3rd} (1982): Degree of selectivity of pergolide as an agonist at presynaptic versus postsynaptic dopamine receptors: implications for prevention or treatment of tardive dyskinesia. *J Clin Psychopharmacol* 2: 371-375.
- Gamma A, Lehmann D, Frei E, Iwata K, Pascual-Marqui R D, Vollenweider FX (2004): Comparison of simultaneously recorded [H215O]-PET and LORETA during cognitive and pharmacological activation. *Hum Brain Mapp* 22: 83-96.
- Garris PA, Kilpatrick M, Bunin MA, Michael D, Walker QD, Wightman RM (1999): Dissociation of dopamine release in the nucleus accumbens from intracranial self-stimulation. *Nature* 398: 67-69.
- Gehring WJ, Willoughby AR (2002): The medial frontal cortex and the rapid processing of monetary gains and losses. *Science* 295: 2279-2282.

- Grace AA (1995): The tonic/phasic model of dopamine system regulation: its relevance for understanding how stimulant abuse can alter basal ganglia function. *Drug Alcohol Depend* 37: 111-129.
- Hadland KA, Rushworth MF, Gaffan D, Passingham RE (2003): The anterior cingulate and reward-guided selection of actions. *J Neurophysiol* 89: 1161-1164.
- Hajcak G, Holroyd CB, Moser JS, Simons RF (2005): Brain potentials associated with expected and unexpected good and bad outcomes. *Psychophysiology* 42: 161-170.
- Hajcak G, Moser JS, Holroyd CB, Simons RF (2007): It's worse than you thought: the feedback negativity and violations of reward prediction in gambling tasks. *Psychophysiology* 44: 905-912.
- Holroyd CB, Coles MGH (2002): The neural basis of human error processing: Reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev* 109: 679-709.
- Holroyd CB, Coles MGH (2008). Dorsal anterior cingulate cortex integrates reinforcement history to guide voluntary behavior. *Cortex* 44: 548-559.
- Holroyd CB, Krigolson OE (2007): Reward prediction error signals associated with a modified time estimation task. *Psychophysiology*, doi:10.1111/j.1469-8986.2007.00561.x.
- Kennerley SW, Walton ME, Behrens TEJ, Buckley MJ, Rushworth, MFS (2006): Optimal decision making and the anterior cingulate cortex. *Nat Neurosci* 9: 940-947.
- Knutson B, Fong GW, Bennett SM, Adams CM, Hommer D (2003): A region of mesial prefrontal cortex tracks monetarily rewarding outcomes: characterization with rapid event-related fMRI. *Neuroimage* 18: 263-272.
- Kreitzer AC., Malenka RC (2007): Endocannabinoid-mediated rescue of striatal LTD and motor deficits in Parkinson's disease models. *Nature* 445: 643-647.

- Lancaster JL, Laird AR, Fox PM, Glahn DE, Fox PT (1997): Automated labeling of the human brain- a preliminary report on the development and evaluation of a forward-transformed method. *Hum Brain Map* 5: 238-242.
- Lei W, Jiao Y, Del Mar N, Reiner A (2004): Evidence for differential cortical input to direct pathway versus indirect pathway striatal projection neurons in rats. *J Neurosci* 24: 8289-8299.
- Leyton M, van der Rot M, Booij L, Baker GB, Young SN, Benkelfat C (2007): Mood-elevating effects of d-amphetamine and incentive salience: the effect of acute dopamine precursor depletion. *J Psychiatry Neurosci* 32: 129-136.
- Luu P, Collins P, Tucker DM (2000): Mood, personality, and self-monitoring: negative affect and emotionality in relation to frontal lobe mechanisms of error monitoring. *J Exp Psychol Gen* 129: 43-60.
- Marsh AA, Blair KS, Vythilingam M, Busis S, Blair RJR (2007): Response options and expectations of reward in decision-making: The differential roles of dorsal and rostral anterior cingulate cortex. *Neuroimage* 35: 979-988.
- Martres MP, Costentin J, Baudry M, Marçais H, Protais P, Schwartz JC (1977): Long-term changes in the sensitivity of pre-and postsynaptic dopamine receptors in mouse striatum evidenced by behavioural and biochemical studies. *Brain Res* 136: 319-337.
- McClure SM, Laibson DI, Loewenstein G, Cohen JD (2004): Separate neural systems value immediate and delayed monetary rewards. *Science* 306: 503-507.
- Menon M, Jensen J, Vitcu I, Graff-Guerrero A, Crawley A, Smith MA, Kapur S (2007): Temporal difference modeling of the Blood-Oxygen Level Dependent response during

aversive conditioning in humans: Effects of dopaminergic modulation. *Biol Psychiatry* 62: 765-772.

Miltner WHR, Braun CH, Coles MGH (1997): Event-related brain potentials following incorrect feedback in a time-estimation task: Evidence for a “generic” neural system for error detection. *J Cogn Neurosci* 9: 788-798.

Montague PR, Dayan P, Sejnowski TJ (1996): A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* 16: 1936-1347

Mulert C, Jager L, Schmitt R, Bussfeld P, Pogarell O, Moller HJ, Juckel G, Hegerl U (2004): Integration of fMRI and simultaneous EEG: towards a comprehensive understanding of localization and time-course of brain activity in target detection. *Neuroimage* 22: 83-94.

Muller SV, Rodriguez-Fornells A, and Munte TF (2005): Brain potentials related to self-generated information used for performance monitoring. *Clin Neurophysiol* 116: 63-74.

Myers RE, Anderson LI, Dluzen DE (2003): Estrogen, but not testosterone, attenuates methamphetamine-evoked dopamine output from superfused striatal tissue of female and male mice. *Neuropharmacology* 44: 624-632.

Nieuwenhuis S, Holroyd CB, Mol N, Coles MG. (2004): Reinforcement-related brain potentials from medial frontal cortex: origins and functional significance. *Neurosci Biobehav Rev* 28: 441-448.

Nieuwenhuis S, Slagter HA, von Geusau A, Heslenfeld DJ, Holroyd CB (2005): Knowing good from bad: Differential activation of human cortical areas by positive and negative outcomes. *Eur J Neurosci* 21: 3161-3168.

- Oliveira FTP, McDonald JJ, Goodman D (2007): Performance monitoring in the anterior cingulate is not all error related: expectancy deviation and the representation of action–outcome associations. *J Cogn Neurosci* 19: 1-11.
- Onn SP, Wang XB (2005): Differential modulation of anterior cingulate cortical activity by afferents from ventral tegmental area and mediodorsal thalamus. *Eur J Neurosci* 21: 2975-2992.
- Pascual-Marqui RD, Lehmann D, Koenig T, Kochi K, Merlo MC, Hell D, Koukkou M (1999): Low resolution brain electromagnetic tomography LORETA functional imaging in acute, neuroleptic-naive, first-episode, productive schizophrenia. *Psychiatry Res: Neuroimaging* 90: 169-179.
- Piercey MF, Hoffmann, WE, Smith MW, Hyslop DK (1996): Inhibition of dopamine neuron firing by pramipexole, a dopamine D3 receptor-preferring agonist: comparison to other dopamine receptor agonists. *Eur J Pharmacol* 312: 35-44.
- Pizzagalli DA, Evins AE, Schetter E, Frank MJ, Pajtas PE, Santesso DL, Culhane M (2008): Single dose of a dopamine agonist impairs reinforcement learning in humans: Behavioral evidence from a laboratory-based measure of reward responsiveness. *Psychopharmacology (Berl)* 196: 221-232.
- Pizzagalli DA, Jahn AL, O’Shea JP (2005): Toward an objective characterization of an anhedonic phenotype: a signal-detection approach. *Biol Psychiatry* 57: 319-327.
- Pizzagalli DA, Oakes TR, Fox AS, Chung MK, Larson CL, Abercrombie HC, Schaefer SM, Benca RM, Davidson RJ (2004): Functional but not structural subgenual prefrontal cortex abnormalities in melancholia. *Mol Psychiatry* 9: 393-405.

- Pizzagalli D, Pascual-Marqui RD, Nitschke JB, Oakes TR, Larson CL, Abercrombie HC, Schaefer SM., Koger JV, Benca RM, Davidson RJ (2001): Anterior cingulate activity as a predictor of degree of treatment response in major depression: evidence from brain electrical tomography analysis. *Am J Psychiatry* 158: 405-415.
- Reynolds JN, Hyland BI, Wickens JR (2001): A cellular mechanism of reward-related learning. *Nature* 413: 67-70.
- Robinson S, Rainwater AJ, Hnasko TS, Palmiter RD (2007): Viral restoration of dopamine signaling to the dorsal striatum restores instrumental conditioning to dopamine-deficient mice. *Psychopharmacology (Berl)* 191: 567-578.
- Rogers RD, Ramnani N, Mackay C, Wilson JL, Jezzard P, Carter CS, Smith SM (2004): Distinct portions of anterior cingulate cortex and medial prefrontal cortex are activated by reward processing in separable phases of decision-making cognition. *Biol Psychiatry* 55: 594-602.
- Rushworth MF, Buckley MJ, Behrens TE, Walton, ME, Bannerman DM (2007): Functional organization of the medial frontal cortex. *Curr Opin Neurobiol* 17: 220-227.
- Schultz W (2002): Getting formal with dopamine and reward. *Neuron*. 36: 241-263.
- Schultz W (2007): Behavioral dopamine signals. *Trends Neurosci* 30: 203-210.
- Schultz W, Dayan P, Montague PR (1997): A neural substrate of prediction and reward. *Science* 275: 1593-1599.
- Schwabe K, Koch M (2007): Effects of aripiprazole on operant responding for a natural reward after psychostimulant withdrawal in rats. *Psychopharmacology* 191: 759-765.
- Siessmeier T, Kienast T, Wrase J, Larsen JL, Braus DF, Smolka MN, Buchholz HG, Schreckenberger M, Rosch F, Cumming P, Mann K, Bartenstein P, Heinz A (2006): Net

- influx of plasma 6-[18F]fluoro-L-DOPA FDOPA to the ventral striatum correlates with prefrontal processing of affective stimuli. *Eur J Neurosci* 24: 305-313.
- Spielberger CD, Gorsuch RL, Lushere RE (1970): *Manual of the State-Trait Anxiety Inventory*. Palo Alto, CA: Consulting Psychologists Press.
- Sumners C, de Vries JB, Horn AS (1981): Behavioural and neurochemical studies on apomorphine-induced hypomotility in mice. *Neuropharmacol* 20: 1203-1208.
- Tissari AH, Rossetti ZL, Meloni M, Frau MI, Gessa GL (1983): Autoreceptors mediate the inhibition of dopamine synthesis by bromocriptine and lisuride in rats. *Eur J Pharmacol* 91: 463-468.
- Towle VL, Bolaños J, Suarez D, Tan K, Grzeszczuk R, Levin DN, Cakmur R, Frank SA, Spire JP (1993): The spatial location of EEG electrodes: locating the best-fitting sphere relative to cortical anatomy. *Electroencephalogr Clin Neurophysiol* 86: 1-6.
- Van Veen V, Holroyd CB, Cohen JD, Stenger VA, Carter CS (2004): Errors without conflict: Implications for performance monitoring theories of anterior cingulate cortex. *Brain Cogn* 56: 267-276.
- Vitacco D, Brandeis D, Pascual-Marqui R, Martin E (2002): Correspondence of event-related potential tomography and functional magnetic resonance imaging during language processing. *Hum Brain Mapp* 17: 4-12.
- Vollenweider FX, Maguire RP, Leenders KL, Mathys K, Angst J (1998): Effects of high amphetamine dose on mood and cerebral glucose metabolism in normal volunteers using positron emission tomography PET. *Psychiatry Res* 83: 149-162.
- Wright CE, Sisson TL, Ichhpurani AK, Peters GR (1997): Steady-state pharmacokinetic properties of pramipexole in healthy volunteers. *J Clin Pharmacol* 37: 520-525.

Yeung N, Cohen JD, Botvinick MM (2004): The neural basis of error detection: conflict monitoring and the error-related negativity. *Psychol Rev* 111: 931-959.

Yeung N, Sanfey AG (2004): Independent coding of reward magnitude and valence in the human brain. *J Neurosci* 24: 6258-6264.

Zirnheld PJ, Carroll CA, Kieffaber PD, O'Donnell BF, Shekhar A, Hetrick WP (2004): Haloperidol impairs learning and error-related negativity in humans. *J Cog Neurosci* 16: 1098-112.

Zumsteg D, Friedman A, Wennberg RA, Wieser HG (2005): Source localization of mesial temporal interictal epileptiform discharges: correlation with intracranial foramen ovale electrode recordings. *Clin Neurophysiol* 116: 2810-2818.

Table 1

Means (SD) for FRN time-locked to feedback presentation and the N1 and P3 time-locked to stimulus presentation for the placebo ($n = 13$) and pramipexole ($n = 7$) group.

| | Fz | FCz | Cz | Pz |
|-------------|--------------|--------------|--------------|--------------|
| Placebo | | | | |
| FRN | 2.51 (2.07) | 2.92 (2.66) | 1.19 (3.42) | |
| N1 | -1.16 (1.12) | -1.84 (1.05) | -1.86 (1.51) | -1.59 (1.29) |
| P3 | 1.08 (2.27) | 2.64 (2.23) | 3.38 (2.17) | 3.27 (1.81) |
| Pramipexole | | | | |
| FRN | -0.84 (2.22) | 0.15 (2.78) | -0.37 (2.50) | |
| N1 | -0.92 (0.49) | -1.43 (0.53) | -1.81 (0.90) | -2.23 (1.64) |
| P3 | 0.74 (0.75) | 1.79 (0.86) | 3.44 (1.77) | 4.13 (2.13) |

Table 2

Summary of significant results emerging from whole-brain LORETA analyses contrasting the placebo ($n = 13$) and pramipexole ($n = 7$) group.

| Region | MNI Coordinates | | | Brodmann's Area | Voxels | t -value | p-value |
|----------------------------------|-----------------|----|----|-----------------|--------|------------|---------|
| | x | y | z | | | | |
| Dorsal anterior cingulate cortex | 11 | -4 | 43 | 24 | 16 | 3.57 | 0.0022 |
| Medial prefrontal cortex | -3 | 59 | 1 | 10 | 15 | -4.18 | 0.0006 |

The anatomical regions, MNI coordinates, and BAs of extreme t -values are listed. Positive t -values are indicative of stronger current density for the placebo than pramipexole group, and vice versa for negative t -values. The numbers of voxels exceeding the statistical threshold are also reported ($p < 0.005$). Coordinates in mm (MNI space), origin at anterior commissure; (X) = left (-) to right (+); (Y) = posterior (-) to anterior (+); (Z) = inferior (-) to superior (+).

Figure Captions

Figure 1. Neural network model of cortico-striatal circuitry (squares represent units, with height and color reflecting neural activity; yellow = most active, red = less active, grey = not active). The model includes the direct (Go) and indirect (NoGo) pathways of the basal ganglia (Frank, 2005, 2006). The Go cells disinhibit the thalamus via the internal segment of globus pallidus (GPi) and thereby facilitate the execution of an action represented in cortex. The NoGo cells have an opposing effect by increasing inhibition of the thalamus, which suppresses actions and thereby keeps them from being executed. Dopamine from the substantia nigra pars compacta (SNc) projects to the dorsal striatum. A tonic level of dopamine is shown in SNc; a burst or dip ensues in a subsequent error feedback phase, causing corresponding changes in Go/NoGo unit activations, which drive learning, via simulated D1 and D2 receptors. Pramipexole was simulated by reducing the size of DA bursts during rewards to simulate presynaptic autoreceptor effects induced by the low dose.

Figure 2. Left panel: Summary of (a) response bias; (b) discriminability; (c) accuracy for the more frequently rewarded (rich) stimulus; and (d) accuracy for the less frequently rewarded (lean) stimulus. Figures modified from Pizzagalli et al. [2008] with permission. Right panel: Corresponding variables for the intact neural network of cortico-striatal circuitry (“placebo groups”) and the neural network simulating reduced pre-synaptic DA bursts in response to rewards (“pramipexole group”). Error bars refer to standard errors

Figure 3. (a) Averaged ERP waveforms from 200 ms before to 600 ms after the presentation of correct feedback during the probabilistic reward task for the pramipexole (heavy line) and

placebo (light line) group averaged across Fz, FCz and Cz; **(b)** Topographic map of the FRN difference wave between the pramipexole and placebo group (pramipexole minus placebo); and **(c)** Results of voxel-by-voxel independent t-tests contrasting current density for the placebo and pramipexole group in response to reward feedback. Red: relatively higher activity for placebo subjects. Blue: relatively higher activity for pramipexole subjects. Statistical map is thresholded at $P < 0.005$ and displayed on the MNI template.

Figure 1

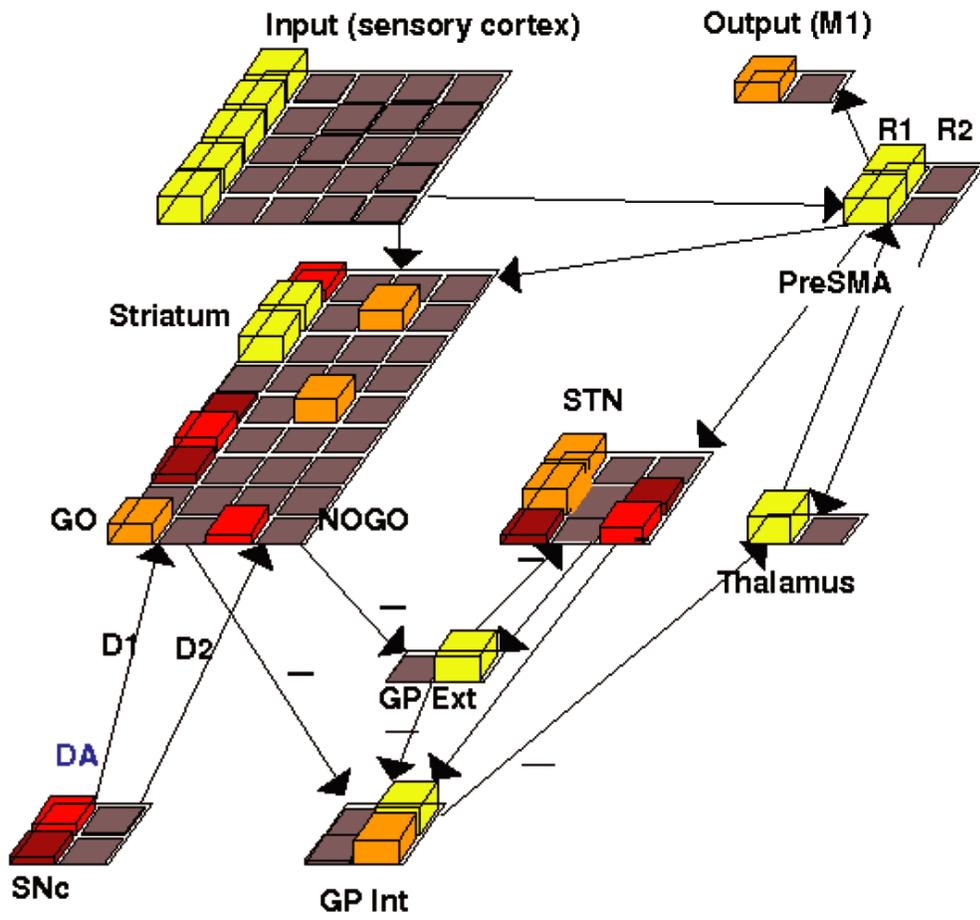


Figure 2

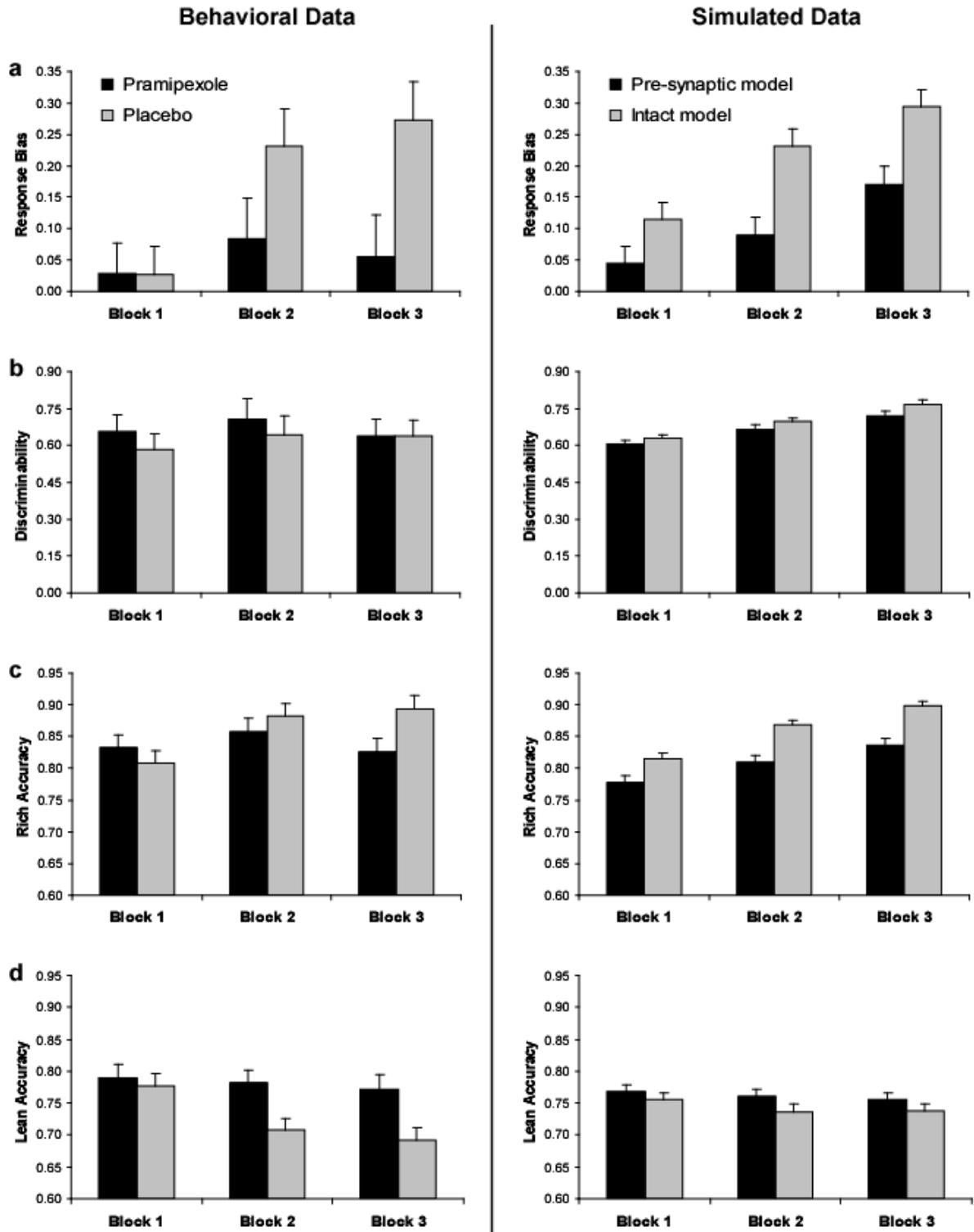


Figure 3

