Neuropsychopharmacology (2014) 39, 2366–2375

www.neuropsychopharmacology.org

Role of Dopamine D2 Receptors in Human Reinforcement Learning

Christoph Eisenegger*[,1,2], Michael Naef[3], Anke Linssen[4], Luke Clark[1,5], Praveen K Gandamaneni[6], Ulrich Müller[1,7] and Trevor W Robbins[1,5]

[1]Behavioural and Clinical Neuroscience Institute, Department of Psychology, University of Cambridge, Cambridge, UK; [2]Social, Cognitive and Affective Neuroscience Unit, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, University of Vienna, Vienna, Austria; [3]Department of Economics, Royal Holloway, University of London, Egham, UK; [4]Department of Neuropsychology and Psychopharmacology, Maastricht University, Maastricht, The Netherlands; [5]Department of Psychology, University of Cambridge, Cambridge, UK; [6]Department of Psychiatry, University of Cambridge, Cambridge, UK; [7]Cambridgeshire & Peterborough NHS Foundation Trust, Adult ADHD Service, Cambridge, UK

Influential neurocomputational models emphasize dopamine (DA) as an electrophysiological and neurochemical correlate of reinforcement learning. However, evidence of a specific causal role of DA receptors in learning has been less forthcoming, especially in humans. Here we combine, in a between-subjects design, administration of a high dose of the selective DA D2/3-receptor antagonist sulpiride with genetic analysis of the DA D2 receptor in a behavioral study of reinforcement learning in a sample of 78 healthy male volunteers. In contrast to predictions of prevailing models emphasizing DA's pivotal role in learning via prediction errors, we found that sulpiride did not disrupt learning, but rather induced profound impairments in choice performance. The disruption was selective for stimuli indicating reward, whereas loss avoidance performance was unaffected. Effects were driven by volunteers with higher serum levels of the drug, and in those with genetically determined lower density of striatal DA D2 receptors. This is the clearest demonstration to date for a causal modulatory role of the DA D2 receptor in choice performance that might be distinct from learning. Our findings challenge current reward prediction error models of reinforcement learning, and suggest that classical animal models emphasizing a role of postsynaptic DA D2 receptors in motivational aspects of reinforcement learning may apply to humans as well.
Neuropsychopharmacology (2014) **39,** 2366–2375; doi:10.1038/npp.2014.84; published online 7 May 2014

INTRODUCTION

Animals and humans flexibly choose actions in pursuit of rewards on a trial-and-error basis by forming stimulus–outcome associations that optimize the likelihood of obtaining future rewards, a process known as reinforcement learning. The neurotransmitter dopamine (DA) is thought to be central to this process, as evidenced by research in both non-human species (Bayer and Glimcher, 2005; Glimcher, 2011; Schultz, 1998; Schultz *et al*, 1997; Wise, 2004; Wise and Rompre, 1989) and humans (Chowdhury *et al*, 2013; Glimcher, 2011; Pessiglione *et al*, 2006). Mechanistic accounts hold that such learning is driven by a so-called 'prediction error' signaling the difference between expected and obtained events, which is then used to update predictions for events in the environment (Sutton and Barto, 1998). A putative neurobiological substrate of the prediction error signal for reward is the phasic firing of DA neurons in the midbrain (Montague *et al*, 1996) projecting to the striatum, one of the major input structures of the basal ganglia. Midbrain dopaminergic reward prediction error signals are assumed to regulate the plasticity of cortico-striatal synaptic transmission by enhancing NMDA-receptor-mediated postsynaptic currents (Calabresi *et al*, 2000; Seamans *et al*, 2001; Wang and O'Donnell, 2001). In doing so, they are thought to contribute to the strengthening of associations leading to rewarding, but not aversive, outcomes (Schultz *et al*, 1997).

So far, human evidence for this account has derived mainly from clinical studies on the effects of dopaminergic medications in Parkinson's disease (eg, Frank *et al*, 2004; Palminteri *et al*, 2009; Voon *et al*, 2010) and Tourette's syndrome (Palminteri *et al*, 2009). Studies of DA agonists or antagonists in healthy volunteers have produced inconclusive results, likely due to a reliance on low doses of available agents (Cohen *et al*, 2007; Jocham *et al*, 2011; Kirsch *et al*, 2005; McCabe *et al*, 2011; Mehta *et al*, 2005; Pessiglione *et al*, 2006; Pizzagalli *et al*, 2008; Riba *et al*, 2008; van der Schaaf *et al*, 2012), which do not cause a necessary level of postsynaptic DA D2 receptor occupancy. Low doses of amisulpride (similar to sulpiride, both are selective for

*Correspondence: Dr C Eisenegger, Department of Basic Psychological Research and Research Methods, Faculty of Psychology, Liebiggasse 5, University of Vienna, Vienna 1010, Austria, Tel: +43 1 660 458 8188, Fax: +43 1 4277 47193, E-mail: christoph.eisenegger@univie.ac.at
Received 13 December 2013; revised 10 March 2014; accepted 27 March 2014; accepted article preview online 9 April 2014

DA D2/3 receptors) may also exert greater functional blockade of cortical and limbic DA receptors, rather than striatal receptors (Bressan *et al*, 2003; Xiberas *et al*, 2001). Thus, sufficiently high doses of sulpiride have to be administered in order to achieve an effective level of postsynaptic DA D2 receptor occupancy in the striatum. Previous studies have shown that a single dose of 400 mg of sulpiride occupies ~30% of striatal DA D2 receptors (Mehta *et al*, 2008), whereas 800 mg results in ~60% occupancy levels, without causing side effects in healthy volunteers (Takano *et al*, 2006). Thus, a dose of 800 mg should allow a direct test of striatal postsynaptic DA D2 receptor involvement in reinforcement learning.

Striatal DA D2 receptor density is assumed to be influenced by genetic factors, with the DA D2 receptor Taq1A polymorphism being the most widely investigated variation, as the minor A1 allele has been associated with a reduction in striatal DA D2 receptor density of up to 30% (Jonsson *et al*, 1999; Pohjalainen *et al*, 1998; Ritchie and Noble, 1996, 2003; Thompson *et al*, 1997). One might expect these A1+ carriers to be disproportionately sensitive to DA D2 receptor antagonism in terms of behavioral impairments during reinforcement learning. Such a pharmacogenetic approach (Eisenegger *et al*, 2010, 2013; Frank and Fossella, 2008), targeted to the D2 receptor, augments previous behavioral genetic studies that have been conducted without drug administration (Jocham *et al*, 2009; Klein *et al*, 2007). We investigate how behavioral impairments during reinforcement learning following DA D2/3 receptor antagonist administration are modulated by genetically determined differences in striatal DA D2 receptor occupancy. This approach provides a means of addressing specificity for the DA D2 receptor.

Although the role of the dopaminergic system in reward prediction has been widely investigated in animals and imaging studies in humans, the functions of this system extend well beyond reinforcement learning. One of the classical functions associated with DA is its control of the motivational aspects of behavior (Ahlenius *et al*, 1977; Bardgett *et al*, 2009; Beninger and Phillips, 1981; Berridge and Robinson, 1998; Lex and Hauber, 2010; Niv, 2007). For instance, studies in rodents have shown that administration of DA antagonists disrupts the ability to associate a reward with the actions necessary to obtain it, but leaves consummatory behavior unaffected (Ikemoto and Panksepp, 1999; Wise, 2004). Therefore, it is important to bear in mind the role of DA not only in reinforcement learning but also in the modulation of the expression of such learning in performance (Salamone, 1994; Shiner *et al*, 2012; Smittenaar *et al*, 2012).

In sum, whereas neurocomputational models emphasize DA as a neurochemical correlate of reinforcement learning via prediction errors, and impaired learning following DA antagonism as a consequence, animal models suggest impairments in expression of learned associations.

Therefore, it is important to clarify whether a high dose of a DA D2/3-receptor antagonist impairs behavior during the acquisition phase of reinforcement learning or during the expression and maintenance of accurately learned associations. Furthermore, by investigating the influence of genetic differences in striatal DA D2 receptor density, we are able to determine the specificity of any behavioral effects of the high-dose DA D2/3 receptor antagonist during reinforcement learning.

We administered 800 mg of sulpiride or placebo to 78 volunteers, genotyped for the DA D2 receptor Taq1A polymorphism, in a behavioral genetic study of reinforcement learning. We studied learning using an established instrumental conditioning paradigm (Pessiglione *et al*, 2006), during which volunteers are required to choose between two visual stimuli that are probabilistically associated with monetary gains and losses. In our version of the paradigm, there were two pairs of stimuli, one pair was associated with monetary gains (winning £1 with a probability of 75% or winning nothing with a probability of 25%), and a second pair was associated with a monetary loss (losing £1 with a probability of 75% or losing nothing with a probability of 25%). For the first pair, volunteers should seek out the symbol associated with a higher likelihood of winning £1 (ie to choose the 'correct' symbol), whereas for the other pair volunteers should avoid the symbol associated with a higher likelihood of losing £1 (ie to avoid choosing the 'incorrect' symbol).

We hypothesized that sulpiride would produce behavioral impairments during reinforcement learning, and that these effects should vary as a function of individual differences in drug absorption. Furthermore, we expected volunteers with genetically determined reductions in striatal DA D2 receptor density would show the most pronounced behavioral impairments following sulpiride administration.

MATERIALS AND METHODS

Volunteers

Seventy-eight healthy male participants with an age range of 19–44 years (mean = 32.1) participated in the study. All participants were recruited from the Cambridge BioResource, a large community-based panel of volunteers that agreed to take part in research linking genotype with phenotype (http://www.cambridgebioresource.org.uk). All volunteers were right-handed European or North American caucasians. Participants were stratified based on their DA D2 receptor Taq1A genotype, with one group consisting of individuals carrying one or two copies of the A1 allele and the other group consisting of A2 allele homozygotes.

All participants' mental and physical health was screened prior to genotyping using a detailed medical history questionnaire used by Cambridge BioResource. This revealed no history of neurological disease or psychiatric disorders. In addition, the psychiatrist on site performed another structured interview, confirming that volunteers had no significant general psychiatric, medical, or neurological disorder and were not currently taking any prescription medicine, nor drugs of abuse. All volunteers were required to perform an alcohol test upon arrival to the laboratory using a commercially available breath alcohol analyzer. This confirmed that no volunteer had consumed alcohol on the study day.

The study was performed in accordance with the Declaration of Helsinki and approved by the National Research Ethics Committee of Hertfordshire (11/EE/0111). All participants were included in the study after having provided written informed consent. Data collection for two volunteers

was unsuccessful, because one did not understand the instructions to the task (placebo group, A1 −), and the other because he felt uncomfortable in the testing room (sulpiride group, A1 −).

Experimental Design

Volunteers were assigned to receive a single oral dose of either 800 mg of sulpiride or a respective placebo in a randomized and double-blind manner. The resulting four groups of participants (A1 + with sulpiride, $n = 21$; A1 + with placebo, $n = 17$; A1 − with sulpiride, $n = 20$; A1 − with placebo, $n = 18$) were all matched for age and BMI; *post hoc* tests revealed that there was no difference in general intelligence across groups (Mann–Whitney tests, $p > 0.459$).

Procedure

Upon arrival (between 0830 and 1000), volunteers completed the National Adult Reading Test assessing general intelligence and visual analog scales (VAS) assessing alertness (Bond and Lader, 1974). They gave a first blood sample (10 ml), underwent assessments of heart rate and blood pressure and were then required to ingest either the placebo or the sulpiride pill. Volunteers then entered a waiting period during which they were required to stay in the premises in separate and quiet rooms, and were allowed to read newspapers. To increase absorption of sulpiride, volunteers were required to ingest a small snack.

Three hours after drug loading (Mehta *et al*, 2003), when sulpiride plasma levels reached their peak, volunteers had to fill out a comprehensive side-effects questionnaire (Rush *et al*, 2003) and again a VAS assessing alertness. They then provided a second blood sample, and blood pressure and heart rate were measured again. At the end of the study, we asked volunteers to guess whether they received sulpiride or the placebo (Supplementary Information).

The reinforcement learning task was implemented in Visual Basic software and presented on computer screens. Instructions were first presented on-screen, then they were summarized orally, and following this volunteers performed the practice block, to get used to the task.

All volunteers received a flat fee of £50 for participation in the study and an additional payment of 5% of their earnings in the reinforcement learning task. Each volunteer received payment in cash in private at the end of the study.

Sulpiride and Prolactin Serum Concentration Measurements

Serum sulpiride was measured by high-performance liquid chromatography utilizing fluorescence endpoint detection with prior solvent extraction. The excitation and emission wavelengths were 300 and 360 nm, respectively. Both intra- and inter-assay coefficients of variation (CVs) were $<10\%$ and the limit of detection was 5–10 ng/ml.

Serum prolactin was measured by a commercial immunoradiometric assay (MP Biomedicals, UK), which utilized ^{125}I as the ligand. The intra- and inter-assay CVs were 4.2% and 8.2%, respectively, and the limit of detection was 0.5 ng/ml.

Reinforcement Learning Task

The task contained a total of 104 trials, and employed different pairs of geometrical shapes as visual stimuli. The first 24 trials were practice trials, after which new pairs of stimuli appeared. Following this, the main task started, which consisted of 40 trials of a 'gain' domain, randomly interspersed with 40 trials of a 'loss' domain amounting to a total of 80 trials.

Each of the two pairs of stimuli was associated with pairs of outcomes, ie a win of £1 or nil in the 'gain' domain and a loss of £1 or nil in the 'loss' domain. In the 'gain' domain, one stimulus was associated with a probability of winning £1 of 75% and a probability of winning nil of 25%, whereas the other stimulus was associated with a probability of winning £1 of 25% and winning nil of 75%. In the 'loss' pair, one stimulus was associated with a probability of losing £1 of 75% and a probability of losing nil of 25%, whereas the other stimulus was associated with a probability of losing £1 of 25% and losing nil of 75%. Volunteers were unaware of these percentages.

On each trial, one pair was randomly presented and the two stimuli were displayed on the screen, right and left of a central fixation cross, their relative position being counterbalanced across trials. The volunteer was required to choose one of the two stimuli by pressing a corresponding keyboard button. Immediately after the decision, the choice was framed in bold and afterwards £1 coin was displayed in case of a gain, a crossed-out coin was displayed in case of a loss and an empty white circle was displayed in case of the outcome nil. Volunteers were required to select between the two stimuli within a restricted time frame of 1700 ms. If volunteers did not respond within 1700 ms, they were penalized with a loss of £1, in both domains, along with text on-screen showing 'too late'. Thus, in order to accumulate money, volunteers had to learn, by trial and error, the stimulus–outcome associations.

Statistical Methods

We used non-parametric Mann–Whitney tests to test for group differences in the distribution of behavioral choices in the learning task. As a robustness check of our results, we fitted a log growth curve model to the data (Supplementary Table S1).

We used parametric Student's *t*-tests to compare differences in the associated average response latencies. These were log transformed to meet statistical distributional assumptions (Judd and McClelland, 1989). We used raw response latencies for graphical representation and for reporting averages.

To investigate whether a high dose of a DA D2/3-receptor antagonist impairs behavior during the acquisition phase of reinforcement learning and/or during the expression of accurately learned associations, we tested whether the learning rate changes over the 40 trials. We used a method developed by Bai and Perron (1998), which endogenously identifies the number and location of structural breaks in the learning rate over the 40 trials. Using E-views 8.0, the Bai–Perron multiple structural break test identified three basic phases of learning in our data. In the gain domain, we identified pronounced and significant learning in early

trials (trials 1–8), less pronounced but significant learning in the middle (9–24) and absent learning in later trials, ie, we found that learning reached asymptote in trials $\geqslant 25$. The according three phases in the loss domain were trials 1–12 (early), 13–30 (middle) and reaching asymptote in trials $\geqslant 31$ (see Supplementary Figure S1).

Q-Learning Model

To reveal the nature of the observed behavioral impairments, we then applied to the data a Q-learning model that also takes feedback processing into account and thus allows a more sophisticated interpretation of the data. The Q-learning model assumes that each volunteer forms a subjective value for each stimulus and updates this value based on the feedback received in each trial (Sutton and Barto, 1998). For each trial, we calculated the subjective value volunteers assign to each pair of stimuli A and B, indicated as Q_t^A and Q_t^B. This can be interpreted as the expected reward for choosing a certain stimulus A or B. Q_t^A and Q_t^B are updated with the feedback volunteers receive (R_{t-1}) in each trial. Note that (R_{t-1}) is not indexed with A and B, because if a volunteer for instance chooses A and feedback indicates it was the 'correct' choice, this feedback implies that B is the 'incorrect' choice, simply because there are only two symbols to choose from. The following updating rule is used: $Q_t^A = (1-\alpha)Q_{t-1}^A + \alpha R_{t-1}$. The extent to which feedback (ie, $+£1, 0, -£1$) influences the subjective values of the chosen stimulus is referred to as the learning rate and captured by model parameter α (Sutton and Barto, 1998). The higher the learning rate α, the higher the influence of recent feedback on Q_t^A and Q_t^B. Thus, a low α estimate reflects a relatively small impact of prior feedback on the current decision, whereas a higher α estimate indicates a larger impact of feedback. In other words, the current subjective value of each stimulus is updated with the difference between feedback in the previous trial and its previous subjective value, ie with the prediction error term $\left(R_{t-1} - Q_{t-1}^A\right)$.

We then used the softmax function $P_t^A = \exp\left(Q_t^A/\beta\right) / \left[\exp\left(Q_t^A/\beta\right) + \exp\left(Q_t^B/\beta\right)\right]$ to estimate the probability (P_t) with which each stimulus is chosen, and maximum likelihood methods to estimate the learning rate (α) and temperature (β). The starting values are $Q_1^A = Q_1^B = 0$, which implies equiprobable choices at the beginning $\left(P_1^A = P_1^B = 0.5\right)$.

The temperature parameter (β) specifies noise that reflects on the accuracy of response choice (Sutton and Barto, 1998). For example, for a volunteer who chooses randomly between the two stimuli, and thus whose choices do not correlate with the subjective value of the two stimuli, the β estimate is high. Vice versa, if a volunteer always chooses the stimulus with the higher subjective value, then parameter β will be close to zero. Note that this implies that volunteers who are more likely to switch their choice when receiving an unexpected feedback show a higher β.

Earlier research suggests that DA neurons might be differentially involved in learning from positive and negative feedback (Daw *et al*, 2002; Frank *et al*, 2004). Hence, we estimate separate parameters (α and β) for the gain and loss domains. To test whether a model with separate parameters is indeed a better specification than models with combined or partially combined parameters (eg separate α and combined β), we compare these different models using the Bayesian Information Criteria. This model comparison confirms that separate parameters give a better fit than combined or partially combined parameters (Supplementary Table S3).

All reported p values are based on two-sided z-tests. The detailed results for the Q-learning model can be found in the (Supplementary Table S2).

RESULTS

Sulpiride Decreases Frequency of Correct Choices

We found that volunteers successfully learned over 40 trials to choose the high-probability gain symbol and avoid choosing the high-probability loss symbol (Figure 1a). Although the probability of the volunteers choosing the 'correct' symbol increased over the task, sulpiride impaired this selectively for gains: ie for choosing a stimulus indicating a high probability of monetary gain, but not for avoiding the stimulus associated with monetary loss (Figure 1a).

Significant drug effects were not apparent until learning had reached a stable level (trials $\geqslant 25$). In this asymptotic phase, identified via a multiple break-point analysis (Bai and Perron, 1998), volunteers receiving sulpiride chose the high-probability gain symbol (85% 'correct' choices) less often compared with those receiving placebo (94% 'correct' choices, Mann–Whitney test, $p = 0.016$, $n = 76$). No such effects were observed in the earlier phases of learning, and in the loss domain (all $p > 0.15$). Together this suggests that sulpiride affected behavior in the asymptotic phase of learning, selectively in the gain, but not in the loss domain.

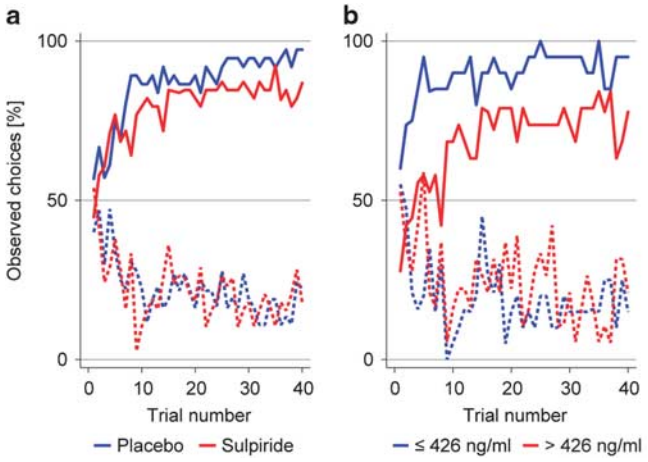

Figure 1 Sulpiride effects on reinforcement learning. The learning curves depict the ratio of volunteers that chose the 'correct' stimulus in the gain domain (upper graph, solid lines), and the 'incorrect' stimulus in the loss domain (lower graph, dashed lines). In the gain domain, the 'correct' stimulus is associated with a probability of 0.75 of winning £1, whereas in the loss domain the 'incorrect' stimulus is associated with a probability of 0.75 of losing £1. (a) Sulpiride (red) compared with placebo (blue) induces behavioral impairments in the gain domain after learning has plateaued (Mann–Whitney test, $p = 0.016$, $n = 76$). (b) Sulpiride group divided by serum values through median split. Higher serum levels (red) relate to more prominent impairments in the gain domain when compared with volunteers with lower serum levels (blue, Mann–Whitney test, $p = 0.031$, $n = 39$).

To reveal the nature of the observed behavioral impairments, we applied to the data a Q-learning model that distinguishes two major components of reinforcement learning, ie the learning rate and choice performance.

Sulpiride Selectively Affects Choice Performance

We did not observe any significant differences between the placebo and the sulpiride group on the learning rate α (Figure 2a, Supplementary Table S2), neither in the gain (placebo $\alpha = 0.06$, sulpiride $\alpha = 0.06$, $p = 0.819$) nor in the loss domain (placebo $\alpha = 0.15$, sulpiride $\alpha = 0.13$, $p = 0.389$). In contrast, β was 57% higher in the sulpiride group compared with the placebo group (placebo $\beta = 0.14$, sulpiride $\beta = 0.22$, $p = 0.005$). This effect was selective for the gain domain (treatment × domain interaction term $= -0.10$, $p = 0.008$), with sulpiride having no effect on β in the loss domain (placebo $\beta = 0.32$, sulpiride $\beta = 0.30$, $p = 0.368$). These results suggest that sulpiride impairs learned choice performance of symbols indicating a high probability of monetary gain.

Analysis of response latencies revealed a similar pattern of results (Figure 3 and Supplementary Table S5). Sulpiride induced a significant prolongation of response latencies in the gain domain, when learning had converged (566 *vs* 611 ms, *t*-test, $p = 0.044$). The increase in response times was not significant in the loss domain (767 *vs* 792 ms, *t*-test, $p = 0.621$).

We also found no apparent side effects on heart rate, blood pressure and self-reported neurovegetative symptoms, as well as alertness, calmness and contentedness (Supplementary Table S4). Moreover, volunteers were at chance levels in guessing whether they had received sulpiride or placebo (Supplementary Information). Furthermore, as an indication of postsynaptic DA D2 effects, we found that sulpiride administration induced a significant rise in serum prolactin values (Supplementary Information).

Serum Sulpiride Values Predict Impairments in Choice Performance

Substantial individual differences were observed in serum-sulpiride concentrations (range: 164–1782 ng/ml, median = 426 ng/ml), providing a means of estimating possible dose-response effects (Dodds *et al*, 2009). To reduce noise, we aggregated these values in two groups by median split. Note that Taq1A genotype was not significantly associated with the two groups (below median: A1+ ($n = 9$) and A1− ($n = 11$); above median: A1+ ($n = 12$) and A1− ($n = 7$) carriers, *t*-test, $p = 0.27$).

If the sulpiride effects observed above are in fact primarily a result of postsynaptic DA D2/3 blockade, then these effects should be most pronounced in those who have high serum levels.

Indeed, we found that the above reported sulpiride effect was driven by volunteers with higher serum levels (Figure 1b). In the asymptotic phase of learning, volunteers with higher serum levels choose the high-probability gain symbol less frequently (75% on average) than volunteers with low serum values (94% on average, Mann–Whitney test, $p = 0.031$, $n = 39$).

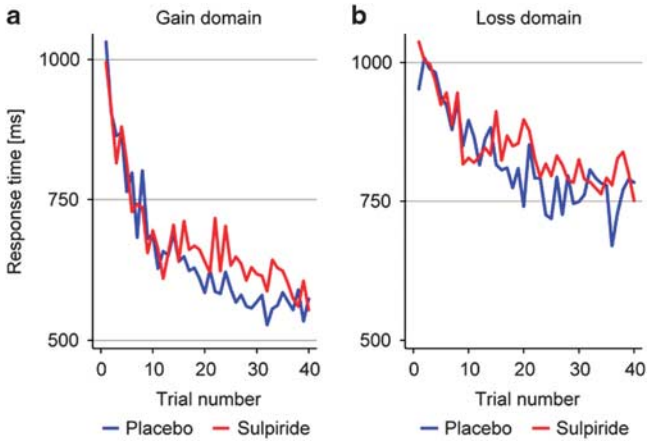

Figure 3 Response latencies over trials. (a) Whereas response latencies decrease over trials in both groups, volunteers in the sulpiride (red) compared with the placebo (blue) group show higher response latencies in the gain domain in the plateau phase of learning (*t*-test, $p = 0.044$, $n = 76$). (b) Response latencies in the loss domain are higher than in the gain domain, and there is no significant difference between the sulpiride group (red) and the placebo group in the loss domain (blue).

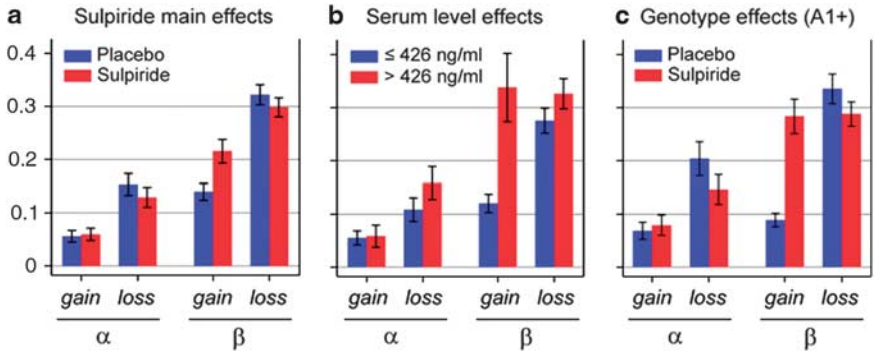

Figure 2 Parameter estimates of the Q-learning model across drug, serum value and genotype groups, separately for the gain and loss domain. (a) Temperature parameter β_{gain} is significantly higher in the sulpiride group (57% increase compared with the placebo, $p = 0.005$), but the learning rate α_{gain} is not affected, and there are no effects in the loss domain (α_{loss}, β_{loss}). (b) Higher sulpiride serum values selectively affect the temperature parameter β_{gain} (183% increase in high compared with low serum values, $p = 0.001$), with no effects on either α_{gain}, α_{loss} or β_{loss}. (c) Pronounced sulpiride effects on β_{gain} are observed in A1+ genotype carriers (211% increase following sulpiride compared with placebo administration, $p < 0.001$), but not in A1− genotype carriers.

Analysis of model parameters showed that effects were linked to a higher parameter β (Figure 2b, Supplementary Table S2), with higher serum levels being associated with a substantial 183% increase in β (lower sulpiride $\beta = 0.12$, higher sulpiride $\beta = 0.34$, $p = 0.001$), selectively in the gain domain (serum level group × domain interaction term $= -0.17$, $p = 0.029$), with effects on β being absent in the loss domain (lower sulpiride $\beta = 0.28$, higher sulpiride $\beta = 0.33$, $p = 0.168$). There were no significant effects on α (main and interaction effects, all $p > 0.190$). Furthermore, we find that volunteers with lower serum levels do not behave differently than volunteers receiving placebo. In the asymptotic phase, volunteers with lower serum levels (< 426 ng/ml) chose the correct symbol as often as volunteers receiving placebo (94% on average in both groups, Mann–Whitney test, $p = 0.44$, $n = 57$). The estimated parameters of the Q-learning model are also not different between volunteers receiving placebo and those receiving sulpiride and low serum levels (placebo $\beta = 0.14$, lower sulpiride $\beta = 0.12$, $p = 0.403$; placebo $\alpha = 0.06$, lower sulpiride $\alpha = 0.05$, $p = 0.948$). Thus, earlier studies used DA D2 antagonist doses that were probably insufficiently high to occupy a substantial proportion of postsynaptic striatal DA D2 receptors (Jocham *et al*, 2011; Mehta *et al*, 2005, 2008; Pessiglione *et al*, 2006; van der Schaaf *et al*, 2012). Our results indicate that high-dose sulpiride impairs learned choice performance, driven by those who achieve higher serum levels, thus indicating that effects are driven by blockade of postsynaptic DA D2/3 receptors.

Impairments in Choice Performance are Linked to DA D2 Receptor Genotype

Although sulpiride has higher selectivity for the DA D2 type receptors (with no significant binding to DA D1 type, α adrenergic, histaminergic or serotoninergic receptors), it does not distinguish between DA D2 and D3 receptors. Stratification of our sample by a D2 receptor-specific genetic polymorphism, the Taq1A genotype, provided a means of addressing specificity for the DA D2 receptor. We tested whether volunteers who carry a genotype associated with a ~30% reduction in DA D2 receptor density showed a more marked impairment in learned choice performance following sulpiride administration. This pharmacogenetic approach also sheds light on the issue of whether this polymorphism is involved in generating behavioral variation or whether it is only spuriously correlated with it.

We found that volunteers carrying at least one copy of the A1 allele (A1 +) drove the sulpiride effect (Figure 4a and b). These volunteers showed a prominent behavioral impairment during learning about reward-predicting stimuli, with A1 + carriers receiving placebo choosing the high-probability reward symbol on average 99% of the time in the asymptotic phase, whereas A1 + volunteers receiving sulpiride chose this symbol less often (82% on average, Mann–Whitney test, $p = 0.048$, $n = 38$). However, there was no significant effect in volunteers who were not carrying a copy of the minor A1 allele (A1 − , Mann–Whitney test, $p = 0.165$, $n = 38$).

Q-learning model results showed that sulpiride significantly affected A1 + , but not A1 − , volunteers' choice performance (treatment × genotype interaction term $= -0.20$, $p = 0.001$, Figure 2c, Supplementary Table S2). A1 + volunteers receiving sulpiride showed a 211% increase in β (placebo $\beta = 0.09$, sulpiride $\beta = 0.28$, $p < 0.001$), selectively in the gain domain (treatment × domain interaction term $= -0.24$, $p < 0.001$), with effects on β being absent in the loss domain (placebo $\beta = 0.34$, sulpiride $\beta = 0.29$, $p = 0.190$). There were no significant effects on α (main and interaction effects, all $p > 0.172$).

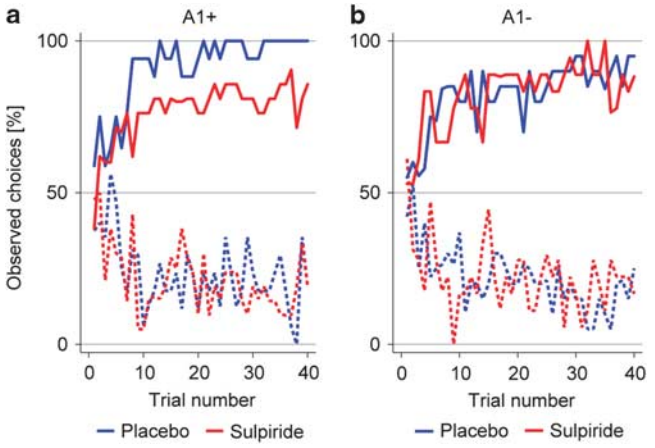

Figure 4 Pharmacogenetic interaction of sulpiride and DA D2 receptor Taq1A polymorphism on reinforcement learning. (a) Volunteers with a genetically determined reduction in DA D2 receptor density (A1 + allele carriers) show behavioral impairments during reinforcement learning following sulpiride (red) compared with placebo (blue) administration in the gain domain (upper graph, solid lines, Mann–Whitney test, $p = 0.048$, $n = 38$), but not in the loss domain (lower graph, dashed lines). (b) Sulpiride (red) compared with placebo (blue) has no significant effect in either the gain (upper graph, solid lines) or the loss (lower graph, dashed lines) domain in volunteers who carry the common variant (A1 −) of the polymorphism.

DISCUSSION

Administration of a high dose (800 mg) of the selective DA D2/3-receptor antagonist sulpiride had no effect on reinforcement learning, but impaired the expression of this learning in choice performance for gains, though not monetary losses. This effect was mirrored in response time slowing and was behaviorally selective with no side effects of the drug on blood pressure, heart rate or self-report measures of sedation. The impairment in choice performance was dose dependent in the sense that volunteers who had higher levels of serum sulpiride (based on a median split) showed the most prominent deficits. We also found that the sulpiride effect was greater in volunteers carrying at least one copy of the minor allele of the DA D2 receptor Taq1A polymorphism, which is known to be associated with a 30% reduction in striatal DA D2 receptors. These results bear on the precise functions of different DA receptors in reinforcement learning and performance, with the DA D2 receptor being implicated primarily in appetitive performance.

Our main finding is that postsynaptic DA D2 receptor blockade affects the asymptote of correct percentage choice

performance. We used a Q-learning algorithm with a learning parameter α and temperature β reflecting choice performance in the task to model behavioral choices. Although these two parameters are mathematically not entirely independent of each other, our results show that sulpiride selectively increased β, suggesting impairments in choice performance.

The lack of effect of DA D2 receptor blockade on the learning rate α in our data is at first sight difficult to reconcile with the postulated role of DA in learning through reward prediction errors (Montague *et al*, 1996; Pessiglione *et al*, 2006; Schultz, 1998). However, the DA D1 receptor could be hypothesized to have a more specific role in learning, because it has a relatively low affinity for DA binding, and is more closely associated with phasic DA release following receipt of unexpected rewards (Dreyer *et al*, 2010). Hence, the DA D1 receptor is implicated in the phasic DA surges that result from unexpected rewards obtained in the initial phase of the reinforcement learning task and might facilitate learning via NMDA-dependent long-term plasticity (Lovinger, 2012; Zweifel *et al*, 2009). In contrast, the DA D2 receptor has high affinity (Rice and Cragg, 2008) for DA binding and is more closely related to tonic dopaminergic activity, which has been linked to response vigor and motivational effects (Robbins and Everitt, 1992). Consequently, blockade of the DA D2 receptor may spare learning from reward predictions, but results in impairments in expression of such learning, in the form of choice performance impairments and response time slowing.

Apparently consistent with the reward prediction error account of DA function, Pessiglione *et al* (2006) reported that administration of a low dose of the non-selective DA D2 antagonist haloperidol decreased the sum of correct choices in a reinforcement learning task compared with administration of L-DOPA (the biochemical precursor of DA), which leads to a non-specific increase in brain DA levels. However, as statistical tests against a placebo group were not significant, their results precluded deriving conclusions regarding the role of the DA D2 receptor in reinforcement learning and performance.

The fact that we did not observe effects in the loss condition is relevant in the light of a basal ganglia model of DA function proposing that avoidance learning is mediated by DA D2 receptors. The model suggests that the DA decreases below baseline ('dips') that occur when an outcome is worse than expected (Schultz, 2002) would release otherwise tonically activated D2 receptors on striatopallidal neurons (Frank, 2005). Accordingly, release of D2 receptors is thought to facilitate learning from losses via the so-called 'indirect' pathway (Frank, 2005). The prediction then is that a pharmacological blockade of postsynaptic D2 receptors (simulating a lack of D2 receptor stimulation during dips) would enhance avoidance learning. Although this model has received support from studies in Parkinson's disease patients (Frank *et al*, 2004) and from behavioral genetic studies of the DA D2 receptor without drug administration (Jocham *et al*, 2009; Klein *et al*, 2007), we found neither a main effect of sulpiride, a serum-sulpiride dependent, nor a pharmacogenetic interaction effect on avoidance learning. This clearly points to the DA D2 receptor as being less critically implicated in aversive instrumental learning than predicted by this model (Frank, 2005) and adds psychopharmacological evidence in healthy humans to the general discussion of the relative involvement of DA in reward over punishment processing (Brischoux *et al*, 2009; Lammel *et al*, 2011; Matsumoto and Hikosaka, 2009; Mirenowicz and Schultz, 1996).

The lack of behavioral effects and absence of response time slowing in the loss condition, together with a general absence of physiological and self-reported side effects, excludes general impairments in sensorimotor coordination and attention in accounting for the sulpiride effects in the gain condition. This is consistent with animal research showing that DA receptor blockade can impair appetitive performance both by diminishing incentive motivation as well by impairing the initiation and selection of action (Berridge and Robinson, 1998; Blackburn *et al*, 1987; Ikemoto and Panksepp, 1999; Robbins and Everitt, 1992; Wise, 2004), and with the findings that tonic DA acting on DA D2 receptors primarily affect late-stage performance, via modulating PFC-NAcc information processing (Goto and Grace, 2005). Sulpiride might also have affected performance through effects on the matching *versus* maximization trade-off (Morris *et al*, 2006). For example, volunteers treated with sulpiride could be less motivated to maximize their rewards in the probabilistic reinforcement task and rather tend to match reinforcement rates across the two options. Overall, the findings raise the interesting issue of how motivational factors interact with reinforcement learning to generate not only the performance of a task but also its initial acquisition. Specifically, dopaminergic medication altered the ability to apply instructions concerning which outcomes were rewarding to existing stimulus–outcome associations.

Our study has limitations in that we restricted our study to investigating male participants, owing to the rationale that menstrual cycle effects in females is an unwanted source of variance. Thus, future studies might investigate whether the same results can be obtained in healthy young females. Finally, the stratified genotype groups were comparatively small. Thus, although our pharmacogenetic approach goes beyond mere genotype behavior correlations, independent replication is warranted.

In sum, our results shed new light on the role of DA D2 receptor blockade in reinforcement learning in healthy volunteers by showing that DA D2 receptor blockade impairs measures of choice performance rather than the learning rate. Our findings might have relevance in the context of pharmacological treatment of psychosis, by pointing to genetic susceptibilities in causing impairments in motivation as a result of DA receptor antagonist administration. The results may have implications for understanding the behavioral mechanisms of action of anti-psychotic drugs and may, for example, be consistent with the view that they reduce 'aberrant motivational salience' of stimuli hypothetically contributing to positive symptoms (Kapur, 2003). Furthermore, our findings may have relevance in the context of anhedonia, which is characterized by a low motivation to provide effort for rewards (Treadway and Zald, 2013), and is a common symptom in schizophrenic patients treated with DA receptor blockers.

FUNDING AND DISCLOSURE

This research work was funded by a Core Award from the Medical Research Council and the Wellcome Trust to the Behavioural and Clinical Neuroscience Institute (MRC Ref G1000183; WT Ref 093875/Z/10/Z). CE was supported by the Swiss National Science Foundation (PA00P1_134135) and the Vienna Science and Technology Fund (WWTF VRG13-007). AL is an employee of Medpace Medical Device B.V. LC is a Principal Consultant for Cambridge Cognition. UM discloses consultancy for Janssen-Cilag, Lilly, Heptares and Shire, and educational funding from AstraZeneca, Bristol-Myers Squibb, Janssen-Cilag, Lilly, Lundbeck and Pharmacia-Upjohn. TWR discloses consultancy with Lilly, Lundbeck, Teva, Shire Pharmaceuticals, ChemPartners and Cambridge Cognition Ltd and research grants with Lilly, Lundbeck and GlaxoSmithKline. The remaining authors declare no conflict of interest.

ACKNOWLEDGEMENTS

We gratefully acknowledge the participation of all NIHR Cambridge BioResource (CBR) volunteers. We thank the Cambridge BioResource staff for their help with volunteer recruitment. We also thank members of the Cambridge BioResource SAB and Management Committee for their support given to our study and the National Institute for Health Research Cambridge Biomedical Research Centre for funding. Access to CBR volunteers and their data and samples is governed by the CBR SAB. Documents describing access arrangements and contact details are available at http://www.cambridgebioresource.org.uk. We thank Violetta Dalla and Rudolf Cardinal for their support in the statistical analysis.

Author Contributions

CE, MN, UM, LC and TWR designed the research; CE, MN, AL, and UM performed the research; UM and PKG provided medical cover; MN analyzed the data; CE, MN and TWR wrote the paper.

REFERENCES

Ahlenius S, Engel J, Zoller M (1977). Effects of apomorphine and haloperidol on exploratory-behavior and latent learning in mice. *Phys Psychol* **5**: 290–294.

Bai JS, Perron P (1998). Estimating and testing linear models with multiple structural changes. *Econometrica* **66**: 47–78.

Bardgett ME, Depenbrock M, Downs N, Points M, Green L (2009). Dopamine modulates effort-based decision making in rats. *Behav Neurosci* **123**: 242–251.

Bayer HM, Glimcher PW (2005). Midbrain dopamine neurons encode a quantitative reward prediction error signal. *Neuron* **47**: 129–141.

Beninger RJ, Phillips AG (1981). The effects of pimozide during pairing on the transfer of classical conditioning to an operant discrimination. *Pharmacol Biochem Behav* **14**: 101–105.

Berridge KC, Robinson TE (1998). What is the role of dopamine in reward: hedonic impact, reward learning, or incentive salience? *Brain Res Brain Res Rev* **28**: 309–369.

Blackburn JR, Phillips AG, Fibiger HC (1987). Dopamine and preparatory behavior: I. effects of pimozide. *Behav Neurosci* **101**: 352–360.

Bond A, Lader M (1974). Use of analog scales in rating subjective feelings. *Br J Med Psychol* **47**: 211–218.

Bressan RA, Erlandsson K, Jones HM, Mulligan R, Flanagan RJ, Ell PJ *et al* (2003). Is regionally selective D2/D3 dopamine occupancy sufficient for atypical antipsychotic effect? An *in vivo* quantitative [123I]epidepride SPET study of amisulpride-treated patients. *Am J Psychiatry* **160**: 1413–1420.

Brischoux F, Chakraborty S, Brierley DI, Ungless MA (2009). Phasic excitation of dopamine neurons in ventral VTA by noxious stimuli. *Proc Natl Acad Sci USA* **106**: 4894–4899.

Calabresi P, Centonze D, Gubellini P, Marfia GA, Pisani A, Sancesario G *et al* (2000). Synaptic transmission in the striatum: from plasticity to neurodegeneration. *Prog Neurobiol* **61**: 231–265.

Chowdhury R, Guitart-Masip M, Lambert C, Dayan P, Huys Q, Duzel E *et al* (2013). Dopamine restores reward prediction errors in old age. *Nat Neurosci* **16**: 648–653.

Cohen MX, Krohn-Grimberghe A, Elger CE, Weber B (2007). Dopamine gene predicts the brain's response to dopaminergic drug. *Eur J Neurosci* **26**: 3652–3660.

Daw ND, Kakade S, Dayan P (2002). Opponent interactions between serotonin and dopamine. *Neural Netw* **15**: 603–616.

Dodds CM, Clark L, Dove A, Regenthal R, Baumann F, Bullmore E *et al* (2009). The dopamine D2 receptor antagonist sulpiride modulates striatal BOLD signal during the manipulation of information in working memory. *Psychopharmacology (Berl)* **207**: 35–45.

Dreyer JK, Herrik KF, Berg RW, Hounsgaard JD (2010). Influence of phasic and tonic dopamine release on receptor activation. *J Neurosci* **30**: 14273–14283.

Eisenegger C, Knoch D, Ebstein RP, Gianotti LRR, Sandor PS, Fehr E (2010). Dopamine receptor D4 polymorphism predicts the effect of L-DOPA on gambling behavior. *Biol Psychiatry* **67**: 702–706.

Eisenegger C, Pedroni A, Rieskamp J, Zehnder C, Ebstein R, Fehr E *et al* (2013). DAT1 polymorphism determines L-DOPA effects on learning about others' prosociality. *PLoS One* **8**: e67820.

Frank MJ (2005). Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. *J Cogn Neurosci* **17**: 51–72.

Frank MJ, Fossella JA (2008). Neurogenetics and pharmacology of learning, motivation, and cognition. *Neuropsychopharmacology* **36**: 133–152.

Frank MJ, Seeberger LC, O'Reilly RC (2004). By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science* **306**: 1940–1943.

Glimcher PW (2011). Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proc Natl Acad Sci USA* **108**(Suppl 3): 15647–15654.

Goto Y, Grace AA (2005). Dopaminergic modulation of limbic and cortical drive of nucleus accumbens in goal-directed behavior. *Nat Neurosci* **8**: 805–812.

Ikemoto S, Panksepp J (1999). The role of nucleus accumbens dopamine in motivated behavior: a unifying interpretation with special reference to reward-seeking. *Brain Res Brain Res Rev* **31**: 6–41.

Jocham G, Klein TA, Neumann J, von Cramon DY, Reuter M, Ullsperger M (2009). Dopamine DRD2 polymorphism alters reversal learning and associated neural activity. *J Neurosci* **29**: 3695–3704.

Jocham G, Klein TA, Ullsperger M (2011). Dopamine-mediated reinforcement learning signals in the striatum and ventromedial prefrontal cortex underlie value-based choices. *J Neurosci* **31**: 1606–1613.

Jonsson EG, Nothen MM, Grunhage F, Farde L, Nakashima Y, Propping P *et al* (1999). Polymorphisms in the dopamine D2 receptor gene and their relationships to striatal dopamine receptor density of healthy volunteers. *Mol Psychiatry* **4**: 290–296.

Judd CM, McClelland GH (1989). *Data Analysis, A Model-Comparison Approach* Orlando, FL, USA.

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

Kirsch P, Reuter M, Mier D, Vaitl D, Hennig J (2005). Imaging gene-substance interactions: the effect of the DRD2 Taq 1A polymorphism and the dopamine agonist bromocriptine on the brain activation during the anticipation of reward. *J Psychophysiol* **19**: 126–126.

Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M (2007). Genetically determined differences in learning from errors. *Science* **318**: 1642–1645.

Lammel S, Ion DI, Roeper J, Malenka RC (2011). Projection-specific modulation of dopamine neuron synapses by aversive and rewarding stimuli. *Neuron* **70**: 855–862.

Lex B, Hauber W (2010). The role of nucleus accumbens dopamine in outcome encoding in instrumental and Pavlovian conditioning. *Neurobiol Learn Mem* **93**: 283–290.

Lovinger DM (2012). Neurotransmitter roles in synaptic modulation, plasticity and learning in the dorsal striatum. *Neuropharmacology* **58**: 951–961.

Matsumoto M, Hikosaka O (2009). Two types of dopamine neuron distinctly convey positive and negative motivational signals. *Nature* **459**: 837–U834.

McCabe C, Huber A, Harmer CJ, Cowen PJ (2011). The D2 antagonist sulpiride modulates the neural processing of both rewarding and aversive stimuli in healthy volunteers. *Psychopharmacology* **217**: 271–278.

Mehta MA, Hinton EC, Montgomery AJ, Bantick RA, Grasby PM (2005). Sulpiride and mnemonic function: effects of a dopamine D2 receptor antagonist on working memory, emotional memory and long-term memory in healthy volunteers. *J psychopharmacol* **19**: 29–38.

Mehta MA, McGowan SW, Lawrence AD, Aitken MR, Montgomery AJ, Grasby PM (2003). Systemic sulpiride modulates striatal blood flow: relationships to spatial working memory and planning. *Neuroimage* **20**: 1982–1994.

Mehta MA, Montgomery AJ, Kitamura Y, Grasby PM (2008). Dopamine D2 receptor occupancy levels of acute sulpiride challenges that produce working memory and learning impairments in healthy volunteers. *Psychopharmacology* **196**: 157–165.

Mirenowicz J, Schultz W (1996). Preferential activation of midbrain dopamine neurons by appetitive rather than aversive stimuli. *Nature* **379**: 449–451.

Montague PR, Dayan P, Sejnowski TJ (1996). A framework for mesencephalic dopamine systems based on predictive Hebbian learning. *J Neurosci* **16**: 1936–1947.

Morris G, Nevet A, Arkadir D, Vaadia E, Bergman H (2006). Midbrain dopamine neurons encode decisions for future action. *Nat Neurosci* **9**: 1057–1063.

Niv Y (2007). Cost, benefit, tonic, phasic: what do response rates tell us about dopamine and motivation? *Ann N Y Acad Sci* **1104**: 357–376.

Palminteri S, Lebreton Ml, Worbe Y, Grabli D, Hartmann A, Pessiglione M (2009). Pharmacological modulation of subliminal learning in Parkinson's and Tourette's syndromes. *Proc Natl Acad Sci USA* **106**: 19179–19184.

Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD (2006). Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature* **442**: 1042–1045.

Pizzagalli DA, Evins AE, Schetter EC, Frank MJ, Pajtas PE, Santesso DL *et al* (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.

Pohjalainen T, Rinne JO, Nagren K, Lehikoinen P, Anttila K, Syvalahti EK *et al* (1998). The A1 allele of the human D2 dopamine receptor gene predicts low D2 receptor availability in healthy volunteers. *Mol Psychiatry* **3**: 256–260.

Riba J, Kramer UM, Heldmann M, Richter S, Munte TF (2008). Dopamine agonist increases risk taking but blunts reward-related brain activity. *PLoS One* **3**: e2479.

Rice ME, Cragg SJ (2008). Dopamine spillover after quantal release: rethinking dopamine transmission in the nigrostriatal pathway. *Brain Res Rev* **58**: 303–313.

Ritchie T, Noble EP (1996). [3H]naloxone binding in the human brain: alcoholism and the TaqI A D2 dopamine receptor polymorphism. *Brain Res* **718**: 193–197.

Ritchie T, Noble EP (2003). Association of seven polymorphisms of the D2 dopamine receptor gene with brain receptor-binding characteristics. *Neurochem Res* **28**: 73–82.

Robbins TW, Everitt BJ (1992). Functions of dopamine in the dorsal and ventral striatum. *Semin Neurosci* **4**: 119–127.

Rush CR, Stoops WW, Hays LR, Glaser PEA, Hays LS (2003). Risperidone attenuates the discriminative-stimulus effects of d-amphetamine in humans. *J Pharmacol Exp Ther* **306**: 195–204.

Salamone JD (1994). The involvement of nucleus accumbens dopamine in appetitive and aversive motivation. *Behav Brain Res* **61**: 117–133.

Schultz W (1998). Predictive reward signal of dopamine neurons. *J Neurophysiol* **80**: 1–27.

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

Schultz W, Dayan P, Montague PR (1997). A neural substrate of prediction and reward. *Science* **275**: 1593–1599.

Seamans JK, Durstewitz D, Christie BR, Stevens CF, Sejnowski TJ (2001). Dopamine D1/D5 receptor modulation of excitatory synaptic inputs to layer V prefrontal cortex neurons. *Proc Natl Acad Sci USA* **98**: 301–306.

Shiner T, Seymour B, Wunderlich K, Hill C, Bhatia KP, Dayan P *et al* (2012). Dopamine and performance in a reinforcement learning task: evidence from Parkinson's disease. *Brain* **135**: 1871–1883.

Smittenaar P, Chase HW, Aarts E, Nusselein B, Bloem BR, Cools R (2012). Decomposing effects of dopaminergic medication in Parkinson's disease on probabilistic action selection–learning or performance? *Eur J Neurosci* **35**: 1144–1151.

Sutton RS, Barto AG (1998). Reinforcement learning: an introduction. *IEEE Trans Neural Netw* **9**: 1054–1054.

Takano A, Suhara T, Yasuno F, Suzuki K, Takahashi H, Morimoto T *et al* (2006). The antipsychotic sultopride is overdosed—a PET study of drug-induced receptor occupancy in comparison with sulpiride. *Int J Neuropsychopharmacol* **9**: 539–545.

Thompson J, Thomas N, Singleton A, Piggott M, Lloyd S, Perry EK *et al* (1997). D2 dopamine receptor gene (DRD2) Taq1 A polymorphism: reduced dopamine D2 receptor binding in the human striatum associated with the A1 allele. *Pharmacogenetics* **7**: 479–484.

Treadway MT, Zald DH (2013). Parsing anhedonia: translational models of reward-processing deficits in psychopathology. *Curr Dir Psychol Sci* **22**: 244–249.

van der Schaaf ME, van Schouwenburg MR, Geurts DE, Schellekens AF, Buitelaar JK, Verkes RJ *et al* (2012). Establishing the dopamine dependency of human striatal signals during reward and punishment reversal learning. *Cereb Cortex* **24**: 633–642.

Voon V, Pessiglione M, Brezing C, Gallea C, Fernandez HH, Dolan RJ *et al* (2010). Mechanisms underlying dopamine-mediated reward bias in compulsive behaviors. *Neuron* **65**: 135–142.

Wang J, O'Donnell P (2001). D(1) dopamine receptors potentiate nmda-mediated excitability increase in layer V prefrontal cortical pyramidal neurons. *Cereb Cortex* **11**: 452–462.

Wise RA (2004). Dopamine, learning and motivation. *Nat Rev Neurosci* **5**: 483–494.

Wise RA, Rompre PP (1989). Brain dopamine and reward. *Annu Rev Psychol* **40**: 191–225.

Xiberas X, Martinot JL, Mallet L, Artiges E, Canal M, Loc'h C *et al* (2001). *In vivo* extrastriatal and striatal D2 dopamine receptor blockade by amisulpride in schizophrenia. *J Clin Psychopharmacol* **21**: 207–214.

Zweifel LS, Parker JG, Lobb CJ, Rainwater A, Wall VZ, Fadok JP *et al* (2009). Disruption of NMDAR-dependent burst firing by dopamine neurons provides selective assessment of phasic dopamine-dependent behavior. *Proc Natl Acad Sci USA* **106**: 7281–7288.

Supplementary Information accompanies the paper on the Neuropsychopharmacology website (http://www.nature.com/npp)