Neuropsychopharmacology www.nature.com/npp

ARTICLE

Differential contributions of striatal dopamine D1 and D2 receptors to component processes of value-based decision making

Jeroen P. H. Verharen[1,2], Roger A. H. Adan[1] and Louk J. M. J. Vanderschuren[2]

Dopamine has been implicated in value-based learning and decision making by signaling reward prediction errors and facilitating cognitive flexibility, incentive motivation, and voluntary movement. Dopamine receptors can roughly be divided into the D1 and D2 subtypes, and it has been hypothesized that these two types of receptors have an opposite function in facilitating reward-related and aversion-related behaviors, respectively. Here, we tested the contribution of striatal dopamine D1 and D2 receptors to processes underlying value-based learning and decision making in rats, employing a probabilistic reversal learning paradigm. Using computational trial-by-trial analysis of task behavior after systemic or intracranial treatment with dopamine D1 and D2 receptor agonists and antagonists, we show that negative feedback learning can be modulated through D2 receptor signaling and positive feedback learning through D1 receptor signaling in the ventral striatum. Furthermore, stimulation of D2 receptors in the ventral or dorsolateral (but not dorsomedial) striatum promoted explorative choice behavior, suggesting an additional function of dopamine in these areas in value-based decision making. Finally, treatment with most dopaminergic drugs affected response latencies and number of trials completed, which was also seen after infusion of D2, but not D1 receptor-acting drugs into the striatum. Together, our data support the idea that dopamine D1 and D2 receptors have complementary functions in learning on the basis of emotionally valenced feedback, and provide evidence that dopamine facilitates value-based and motivated behaviors through distinct striatal regions.

Neuropsychopharmacology (2019) 44:2195–2204; https://doi.org/10.1038/s41386-019-0454-0

INTRODUCTION

Many decisions we make in everyday life are the result of a process in which the expected gains and losses associated with different courses of action are weighed and compared, and these expectations are often based on the value of the outcomes of similar actions taken in the past. The process by which these action-outcome associations are acquired, stored, and updated to guide behavior, thereby linking positive and negative experiences to actions, is called reinforcement learning [1–3]. Deficits in this process have been implicated in a wide variety of mental disorders, including depression, mania, attention-deficit/hyperactivity disorder, and addiction [4–11].

Dopamine (DA) is an important modulator of value-based learning and decision making, and it does so by attributing salience to relevant cues [12], facilitating the allocation of effort [13, 14], guiding voluntary movement [15], and by signaling reward prediction errors [16–18]. Especially this latter function of DA is thought to be fundamental for value-based learning. Reward prediction error theory posits that midbrain DA neurons signal a discrepancy between anticipated and received reward or punishment. Downstream dopaminoceptive brain areas can use these signals to update future expectations of actions, in order to optimally adapt to environmental changes. As such, DA has also been implicated in cognitive flexibility, since manipulations of the DA system disrupt performance in tasks such as reversal learning and set shifting [19–22].

It has been proposed that the D1 and D2 subclass of DA receptors mediate behaviors of opposing valence, so that activation of DA D1 receptors stimulates the expression of reward-related behaviors, and activation of DA D2 receptors stimulates the expression of aversion-related behaviors [23–25]. Given that endogenous DA primarily stimulates D1-receptor expressing neurons and inhibits D2-receptor expressing neurons [24], it has been proposed that DA facilitates value-based behaviors through opposing roles of DA D1 and D2-receptor expressing neurons in adapting to positively and negatively valenced feedback [23, 26–29]. However, the striatum is heterogeneous in function, morphology and connectivity, and distinct subregions of the striatum and their DAergic innervation have been implicated in distinct processes [30–33]. Accordingly, striatal subregion-specific neuronal manipulations have differential effects in behavioral tasks of value-based decision making, including reversal learning [34].

Importantly, aberrant value-based behavior may arise from impairments in value-based learning, so that animals fail to flexibly

[1]Department of Translational Neuroscience, Brain Center Rudolf Magnus, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, the Netherlands and [2]Department of Animals in Science and Society, Division of Behavioural Neuroscience, Faculty of Veterinary Medicine, Utrecht University, Yalelaan 2, 3584 CM Utrecht, the Netherlands
Correspondence: Louk J. M. J. Vanderschuren (l.j.m.j.vanderschuren@uu.nl)
These authors contributed equally: Roger A.H. Adan, Louk J.M.J. Vanderschuren

Received: 18 February 2019 Revised: 17 June 2019 Accepted: 21 June 2019
Published online: 29 June 2019

© The Author(s), under exclusive licence to American College of Neuropsychopharmacology 2019

adapt to positive or negative outcomes, or by processes directly related to value-based decision making, such as choice perseveration. However, changes in overt behavior, as apparent from the analysis of conventional task parameters, do not necessarily inform about which of these component processes is altered. One way to address this issue is by assessing trial-by-trial behavior using computational reinforcement learning models. Here, we therefore studied the role of DAergic neurotransmission in the striatum in value-based learning and decision making in rats using a probabilistic reversal learning paradigm [35–37]. By applying a Q-learning model [2, 3, 37, 38] to the data, we tried to unravel how DA D1 and D2 receptors in the ventral striatum (VS), dorsolateral striatum (DLS), and dorsomedial striatum (DMS) contribute to four core components of task performance: reward learning (adapting behavior to reward delivery), punishment learning (adapting behavior to reward omission), stickiness (a preference for the previously chosen option, independent of trial outcome), and choice stochasticity (the balance between exploration versus exploitation)—alongside conventional task parameters. We predicted an important role of DA receptors in the VS in reward and punishment learning, given its function in processing reward prediction errors and facilitating motivation [14, 21, 30, 39], and of DA receptors in the dorsal striatum in aspects of value-based decision making, given its function in perseverative behavior [40, 41] and balancing goal-directed versus habitual behaviors [32, 33, 42].

MATERIALS AND METHODS

Animals

A total of 68 adult (>300 g) male Long-Evans rats (Janvier labs, France) were used for the experiments. Rats were housed in pairs (for systemic drug treatment) or singly (for intracranial infusions) in a humidity- and temperature-controlled room and kept on a 12/12 h reversed day/night cycle (lights off at 8 a.m.). All experiments took place during the dark phase of the animals' day/night cycle. Animals were kept on food restriction (~4.5 g standard lab chow per 100 g body weight per day) during the experiments. All experiments were conducted in accordance with European (Directive 2010/63/EU of the European Parliament and of the Council of 22 September 2010 on the protection of animals used for scientific purposes) and Dutch (Animal Testing Act (Act of 26 November 2014 amending the Animal Testing Act (1977) in connection with implementation of Directive 2010/63 / EU) legislation, and approved by the Dutch Central Animal Testing Committee, and the Animal Ethics Committee and Animal Welfare Body of Utrecht University.

Experimental procedures

Experimental procedures are described in the Supplementary Materials and Methods.

Behavioral task

We used a probabilistic reversal learning task, previously described by Verharen et al. [37]. In brief, animals could earn sucrose pellets by responding on two levers that each differed in the probability of being reinforced (Fig. 1a). At task initiation, one lever was randomly assigned as the high-probability lever and pressing this lever had a 80% chance of being reinforced (delivery of a sucrose pellet) and 20% chance of not being reinforced (a 10 s time-out). The other lever was assigned as the low-probability lever, which gave 20% chance of being reinforced and 80% of not being reinforced. Initial assignment of the left and right lever as high- or low-probability was counterbalanced between animals. When the animal made eight consecutive responses on the high-probability lever, a reversal in reward contingencies occurred, so that the previous high-probability lever became the low-probability lever and vice versa. The task terminated after 90 min.

Computational model

We used computational modeling [2, 3, 37, 38] to extract different subcomponents of reward-based decision-making from the raw behavioral data. Consistent with our previous findings [37], we found, using Bayesian model selection [43], that an extension of the classic Rescorla–Wagner model best described the behavior of the rats (Supplementary Figure 1). The model assumes that on every trial, the rat makes a choice based on a representation of the value of each of these levers. In most cases, the animal chooses the lever with the highest value Q on each trial t (Supplementary Figure 2). The relationship between lever values Q_{left} and Q_{right}, and the probability that the rat chooses left or right ($p_{\text{left},t}$, respectively $p_{\text{right},t}$) lever in every trial is described by a softmax function:

$$p_{right,t} = \frac{\exp\left(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t}\right)}{\exp\left(\beta \cdot Q_{left,t} + \pi \cdot \phi_{left,t}\right) + \exp\left(\beta \cdot Q_{right,t} + \pi \cdot \phi_{right,t}\right)} \quad (1)$$

$$\text{and } p_{left,t} = 1 - p_{right,t} \quad (2)$$

In this function, β is the Softmax' inverse temperature, which indicates how value-driven the animal's choices are. If β becomes very large, then the value function $\beta \cdot Q_{s,t}$ of the highest valued side becomes dominant, and the probability that the rat chooses that side approaches one. Is β zero, then $p_{\text{left},t} = p_{\text{right},t} = e^0/(e^0 + e^0) = 0.5$ (π not taken into account), so that choice behavior becomes random. β is sometimes referred to as the explore/exploit parameter, where a low β favors exploration (i.e., sampling of all options) and a high β favors exploitation (i.e., choosing the option which has proven to be beneficial). Therefore, a decrease in β may reflect more explorative choice behavior, although a large decrease in β could also indicate a general disruption of behavior, i.e., that the animal chooses more randomly.

Factor π is a stickiness parameter that indicates a preference for the previously chosen ($\pi > 0$; perseveration) or previously unchosen ($\pi < 0$; alternation) option. Here, ϕ is a boolean with $\phi = 1$ if that lever was chosen in the previous trial, and $\phi = 0$ if not. This adds a certain amount of the value of π to the value function of the lever in trial t, in addition to the lever's expected value $Q_{s,t}$.

For the first trial, both lever values were initiated at 0.5. After each trial, the value of the chosen lever was updated based on the trial's outcome according to a Q-learning rule:

$$Q_{s,t} = \begin{cases} Q_{s,t-1} + \alpha^+ \cdot \text{RPE}_{t-1} & \text{for win trials} \\ Q_{s,t-1} + \alpha^- \cdot \text{RPE}_{t-1} & \text{for lose trials} \end{cases} \quad (3)$$

$$\text{with RPE}_{t-1} = \text{outcome}_t - Q_{s,t-1} \quad (4)$$

$$\text{so that RPE}_{t-1} = \begin{cases} 1 - Q_{s,t-1} & \text{for win trials} \\ 0 - Q_{s,t-1} & \text{for lose trials} \end{cases} \quad (5)$$

in which $Q_{s,t-1}$ is the value of the chosen lever. Here, α^+ and α^- indicate the animal's ability to learn from positive (reinforcement; reward delivery), respectively negative (reward omission) feedback. The value of the unchosen side was not updated and thus retained its previous value.

For each individual session, we used maximum a posteriori estimation to determine the best-fit model parameters; fitting procedures are described in the Supplementary Materials and Methods. Supplementary Figure 3 shows the relationship between different variables of the computational model and conventional measures of task performance.

Code accessibility

MedPC script of the probabilistic reversal learning task is available at https://github.com/jeroenphv/ReversalLearning.

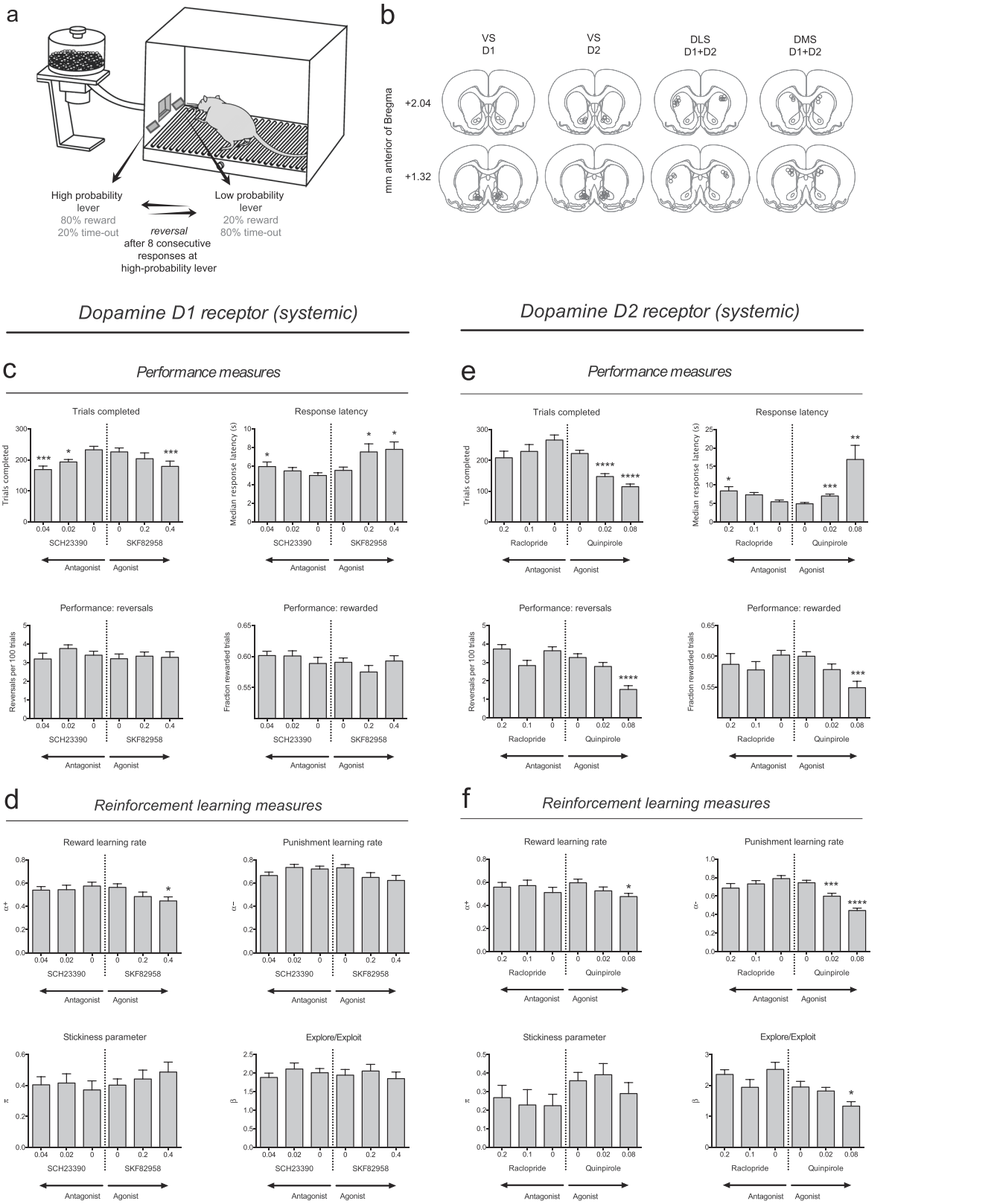

Fig. 1 Task setup and systemic treatment with DA receptor (ant)agonists. **a** Behavioral task. **b** Infusion sites included in the analysis. **c** Effects of systemic treatment with the DA D1 receptor antagonist SCH23390 (0, 0.02 or 0.04 mg/kg) and agonist SKF82958 (0, 0.2 or 0.4 mg/kg) on the behavioral measures of task performance. **d** Effects of systemic treatment with the DA D1 receptor antagonist SCH23390 (0, 0.02 or 0.04 mg/kg) and agonist SKF82958 (0, 0.2 or 0.4 mg/kg) on the computational modeling parameters. **e** Effects of systemic treatment with the DA D2 receptor antagonist raclopride (0, 0.1 or 0.2 mg/kg) and agonist quinpirole (0, 0.02 or 0.08 mg/kg) on the behavioral measures of task performance. **f** Effects of systemic treatment with the DA D2 receptor antagonist raclopride (0, 0.1 or 0.2 mg/kg) and agonist quinpirole (0, 0.02 or 0.08 mg/kg) on the computational modeling parameters. The statistical range is denoted as: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. $n = 24$ rats for SCH23390, SKF82958, and quinpirole, and $n = 18$ rats for raclopride

Statistics

Statistical tests were performed with Prism 6 (GraphPad Software Inc.). For each systemically tested drug, a one-way repeated measures analysis of variance (ANOVA) with Greenhouse-Geisser correction was used to calculate significance. When the ANOVA yielded significant results ($P < 0.05$), a post-hoc Bonferroni test was used to compare the drug doses with vehicle. For the intracranial infusion data, paired, two-tailed *t*-tests were performed in which the tested drugs were compared against vehicle (saline). All statistics are presented in Supplementary Table 1. In all figures, the statistical significance was denoted as follows: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$.

RESULTS

Systemic administration

Treatment with the DA D1 receptor antagonist SCH23390 significantly decreased the number of trials completed (Fig. 1c) and increased the response latency. However, none of the parameters of the reinforcement learning model were significantly affected (Fig. 1d), which was reflected by the lack of effect on the two general performance measures; number of reversals and fraction of rewarded trials (Fig. 1c). Administration of the DA D1 receptor agonist SKF82958 reduced the number of completed trials and increased the response latency (Fig. 1c). In addition, it led to a numerically modest but significant decrease in reward learning rate α^{+} (Fig. 1d), but this had no consequences for the number of reversals made or the fraction of rewarded trials (Fig. 1c). However, win–stay and lose–stay behavior at the high-probability lever were significantly increased by SKF82958 (Supplementary Figure 4). No effects were observed on the value estimates of punishment learning parameter α^{-}, perseveration parameter π, or explore/exploit parameter β (Fig. 1d).

Treatment with the DA D2 receptor antagonist raclopride increased the response latency, without a significant effect on the number of completed trials (Fig. 1e). Furthermore, neither of the measures of task performance were affected (Fig. 1e), which was reflected by the absence of effects on the computational model parameters (Fig. 1f). Injection of the DA D2 receptor agonist quinpirole decreased the number of completed trials and increased response latencies (Fig. 1e). It also impaired task performance, both in terms of the number of reversals and the fraction of rewarded trials (Fig. 1e). Computational analysis revealed that this was associated with a decrease in α^{+}, which was numerically modest, and a profound decrease in α^{-} (Fig. 1f). Moreover, β was significantly decreased, but π was not (Fig. 1f).

Ventral striatum drug administration

Infusion of SCH23390 into the VS did not affect the number of trials completed or the response latencies (Fig. 2a). A significant increase in the number of reversals was observed, but not in the fraction of rewarded trials (Fig. 2a). This increase in the number of reversals was associated by an increase in lose–stay behavior at the high-probability lever (Supplementary Figure 5). However, none of the computational modeling parameters were significantly altered (Fig. 2b), although a trend towards an increase in explore/exploit parameter β was observed ($p = 0.07$; see also Supplementary Table 1). Intra-VS treatment with SKF82958 did not significantly change the number of trials completed, the response latency or the two measures of task performance (Fig. 2a). However, a significant decrease was observed in the value estimate of reward learning parameter α^{+}, without effects on punishment learning rate α^{-}, stickiness parameter π or explore/exploit parameter β.

Infusion of raclopride into the VS significantly increased the animals' response latency, but did not change the number of trials completed (Fig. 2c), the two measures of task performance (Fig. 2c) or any of the computational model parameters (Fig. 2d). In contrast, intra-VS infusion of quinpirole affected different measures of task behavior. First, it strongly decreased the number of trials completed in the task and it increased the response latency of the animals (Fig. 2c). It also reduced the number of reversals achieved, but not the fraction of rewarded trials. Furthermore, win–stay behavior was decreased at both the high-probability and low-probability lever (Supplementary Figure 5). This change in performance was associated with decreases in the value estimates of α^{-} and β, but not by changes in α^{+} or π (Fig. 2d).

Dorsolateral striatum drug administration

Infusion of the SCH23390 or SKF82958 into the DLS had no effect on any of the task measures (Fig. 3a, b). In contrast, infusion of the DA D2 receptor antagonist raclopride significantly reduced the number of completed trials and increased response latencies, but did not affect the two measures of task performance (Fig. 3c). Moreover, none of the computational modeling parameters were significantly changed (Fig. 3d). Intra-DLS infusion of quinpirole increased the animals' response latency, but did not affect the number of trials completed or the conventional performance measures (Fig. 3c and Supplementary Figure 5). It did, however, lead to a significant decrease in the value estimate of explore/exploit parameter β, without any effects on the other computational model parameters (Fig. 3d).

Dorsomedial striatum drug administration

After infusion into the DMS, none of the drugs affected performance in the task or changed the value estimates of the computational model parameters (Fig. 4a-d). Moreover, intra-DMS treatment with SCH23390 or SKF82958 did not affect the trials completed in the task or response latencies (Fig. 4a). Infusion of raclopride increased the response latency of the animals, but did not change the number of trials completed (Fig. 4c). Conversely, infusion of quinpirole decreased the number of trials completed in the task without a significant effect on the animals' response latency (Fig. 4c).

DISCUSSION

In this study, we have used a computational reinforcement learning model to assess how signaling through striatal DA D1 and D2 receptors contributes to subcomponents of value-based learning and decision making, using a probabilistic reversal learning task in rats. This computational modeling approach provides in-depth insights into the behavior of the animals besides the conventional measures of task performance, and can reveal changes in behavioral strategy that do not always become apparent as overt alterations in behavior (e.g., the change in explore/exploit balance after intra-DLS treatment with quinpirole). Interestingly, conventional measures of task performance showed modest, if any, direct correlations with the computational model parameters (Supplementary Figure 3a). Rather, these parameters seemed to interact in quite complicated ways. For example, reward learning rate α^{+} correlated with the fraction of rewarded trials, but only when explore/exploit parameter β was high (i.e., if the animals showed exploitative choice behavior; Supplementary Figure 3b). However, the number of reversals did directly correlate to the value of the stickiness parameter π. This may not be surprising given that the number of reversals depended on perseverative responses (eight in a row) in the high-probability nosepoke hole.

The most important findings of the present study were that negative feedback learning depends on DA D2 receptor signaling, whereas learning from positive feedback depends on DA D1 receptor signaling in the VS. Furthermore, DA D2 receptor function in the VS and DLS is important for the balance between exploitative and explorative choice behavior.

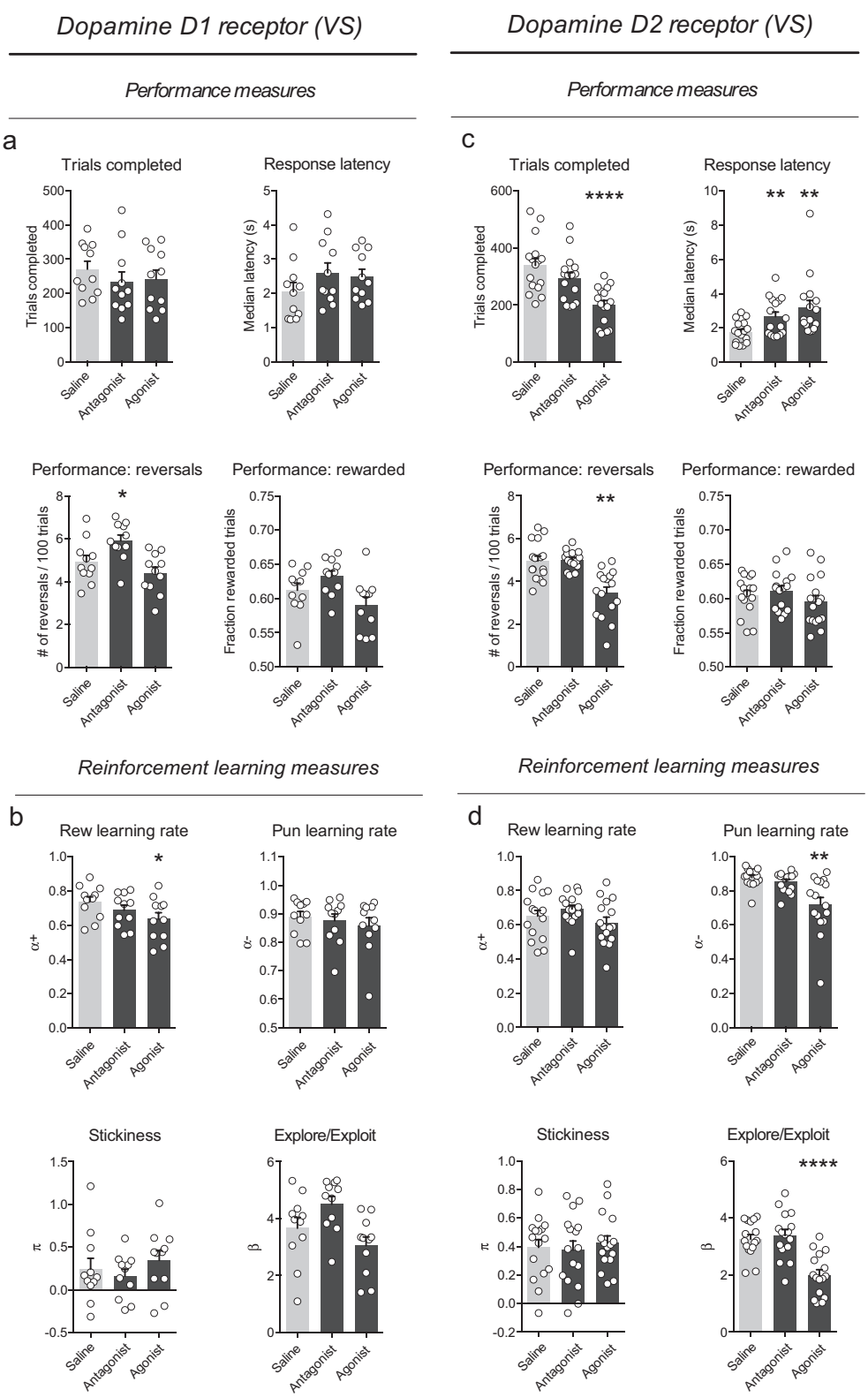

Fig. 2 Ventral striatum infusions. **a** Effects of intra-VS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on the behavioral measures of task performance. **b** Effects of intra-VS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on the computational modeling parameters. **c** Effects of intra-VS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on the behavioral measures of task performance. **d** Effects of intra-VS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on the computational modeling parameters. The statistical range is denoted as: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. $n = 11$ rats for D1 receptor experiments, $n = 16$ rats for D2 receptor experiments

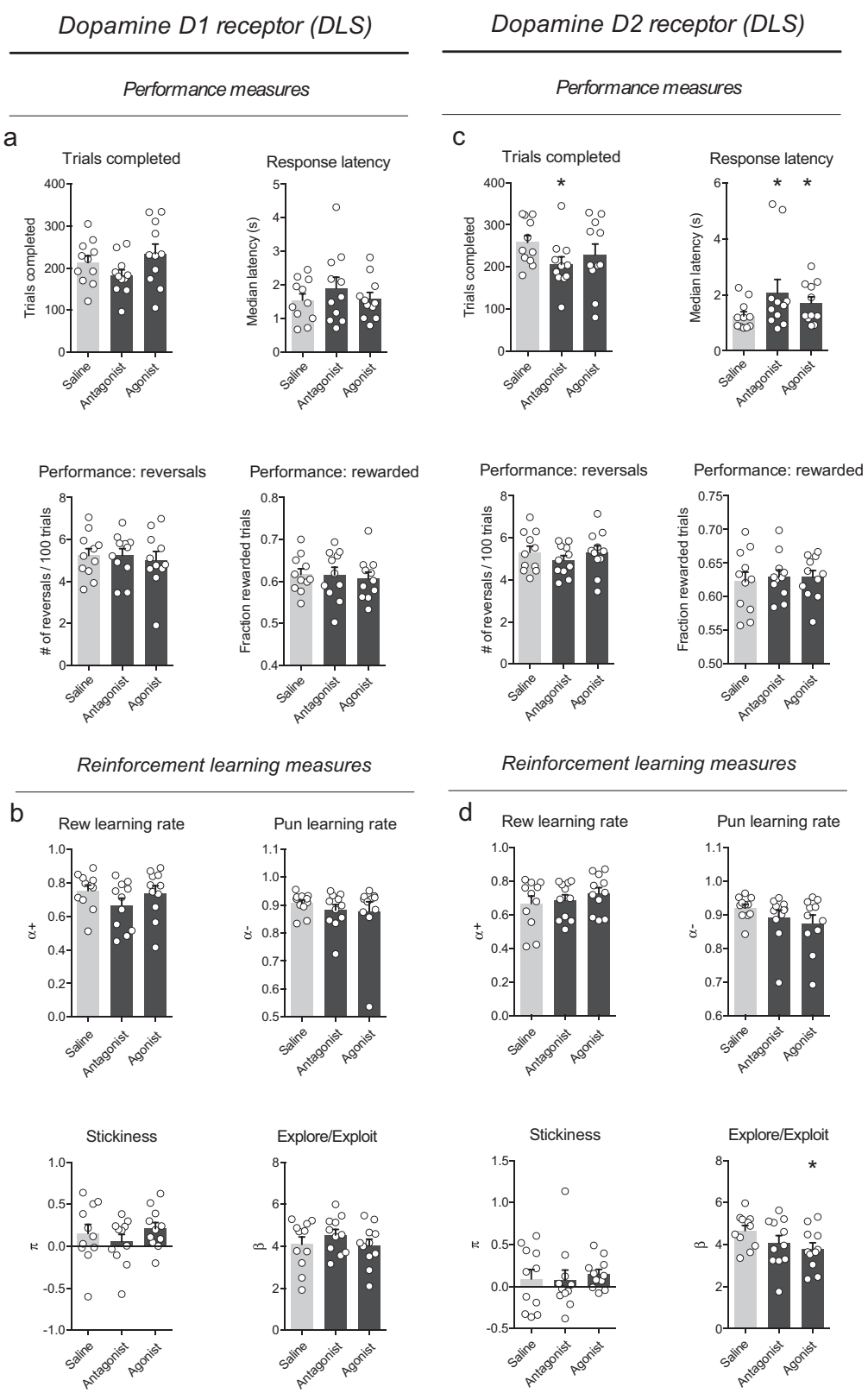

Fig. 3 Dorsolateral striatum infusions. **a** Effects of intra-DLS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on the behavioral measures of task performance. **b** Effects of intra-DLS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on the computational modeling parameters. **c** Effects of intra-DLS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on the behavioral measures of task performance. **d** Effects of intra-DLS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on the computational modeling parameters. The statistical range is denoted as: $^{*}P < 0.05$, $^{**}P < 0.01$, $^{***}P < 0.001$, and $^{****}P < 0.0001$. $n = 11$ rats

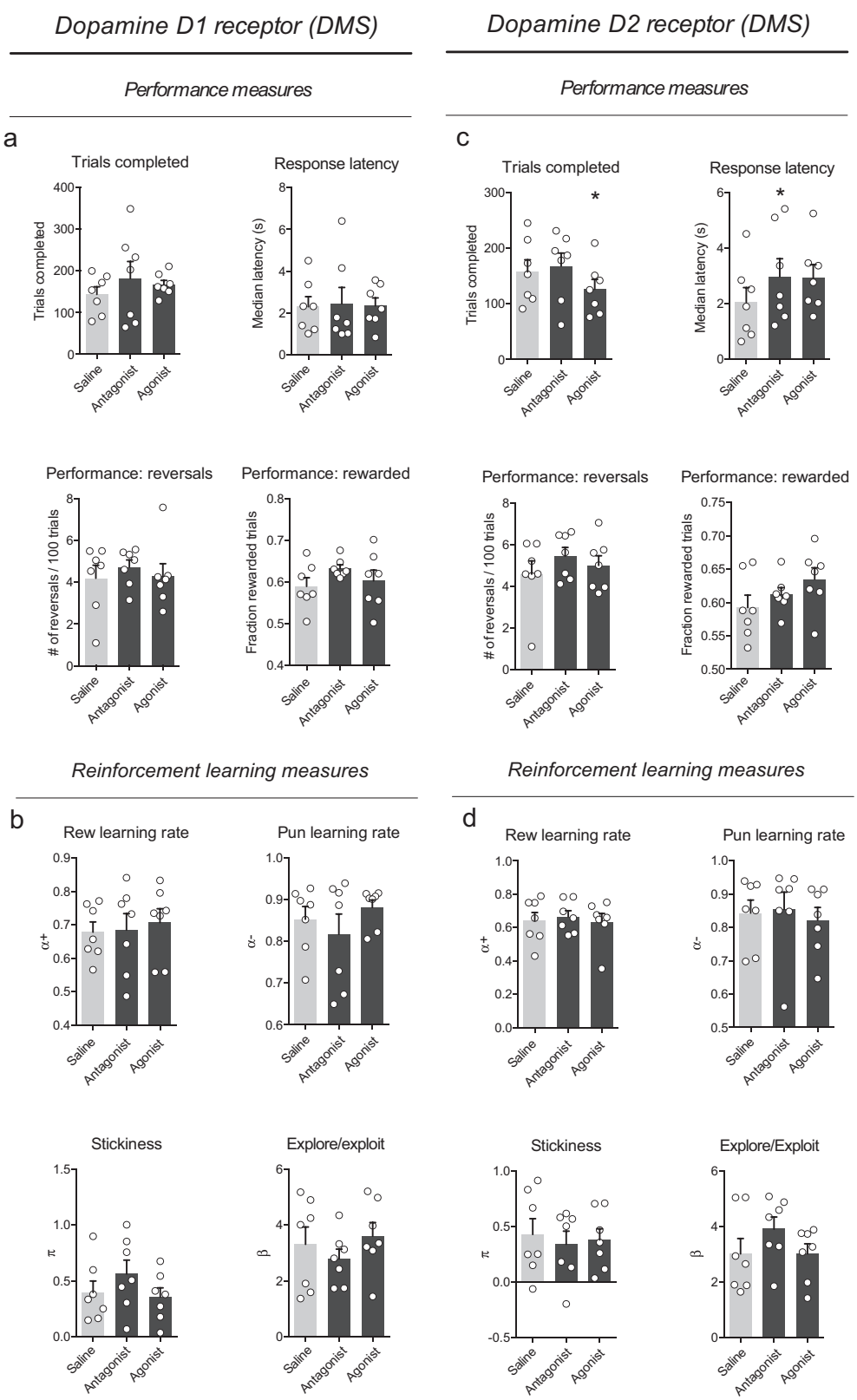

Fig. 4 Dorsomedial striatum infusions. **a** Effects of intra-DMS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on the behavioral measures of task performance. **b** Effects of intra-DMS infusion of the DA D1 receptor antagonist SCH23390 (1 µg/side) and agonist SKF82958 (5 µg/side) on the computational modeling parameters. **c** Effects of intra-DMS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on the behavioral measures of task performance. **d** Effects of intra-DMS infusion of the DA D2 receptor antagonist raclopride (7.5 µg/side) and agonist quinpirole (5 µg/side) on the computational modeling parameters. The statistical range is denoted as: *$P < 0.05$, **$P < 0.01$, ***$P < 0.001$, and ****$P < 0.0001$. $n = 8$ rats

Table 1. Effects of DA receptor (ant)agonists on the computational model (black) and motivational and motoric task parameters (gray)

		Systemic	VS	DLS	DMS
DA D1 antagonist	SCH23390	↓↓ Trials completed ↑ Response latency			
DA D1 agonist	SKF82958	↓ Reward learning ↓↓ Trials completed ↑↑ Response latency	↓ Reward learning		
DA D2 antagonist	Raclopride	↑↑ Response latency	↑↑ Response latency	↓ Trials completed ↑ Response latency	↑ Response latency
DA D2 agonist	Quinpirole	↓ Reward learning ↓↓ Punishment learning ↑ Exploration ↓↓ Trials completed ↑↑ Response latency	↓↓ Punishment learning ↑↑ Exploration ↓↓ Trials completed ↑↑ Response latency	↑ Exploration ↑ Response latency	↓ Trials completed

One arrow denotes significance at a $P < 0.05$ level, two arrows denote significance at a $P < 0.01$ level.

Effects of systemic treatment with DA drugs

After systemic drug treatment (see Table 1), we found a reduction in reward learning after systemic activation of DA D1 and D2 receptors, but not after treatment with their respective antagonists. Punishment learning was solely dependent on the DA D2 receptor, as treatment with the agonist quinpirole decreased this parameter. Furthermore, treatment with quinpirole decreased explore/exploit parameter β, indicating that animals shifted towards a decision-making strategy of exploration, rather than exploitation. No effects were observed on the value estimate of stickiness parameter π, suggesting that choice perseveration is not dependent on DA neurotransmission.

With the exception of quinpirole, which reduced both general performance parameters, treatment with none of the drugs affected the number of reversals or the fraction of rewarded trials. In line with these findings is the notion that the DA D2 receptor has been most strongly implicated in reversal learning, although some studies have also observed reversal deficits following treatment with a D1 receptor agonist [34]. Indeed, the effect of systemic treatment with the DA D1 receptor agonist SKF82958 on reward learning was numerically modest, and did not lead to significant changes in conventional performance measures.

Considering the effects observed after treatment with DA D1 and D2 receptor agonists, it is surprising that no changes in performance or learning rates were observed after treatment with the DA D1 receptor antagonist SCH23390 or the D2 receptor antagonist raclopride, even though we used doses with which effects on cognition have been observed in the past [44, 45]. In fact, the animals became disengaged from the task after systemic treatment with doses higher than 0.04 mg/kg SCH23390 or 0.2 mg/kg raclopride, thereby not completing enough trials to draw reliable conclusions about task performance (data not shown). Importantly, the fact that infusion of these antagonists into the striatum did also not affect reversal learning performance, indicates that baseline DAergic signaling through a single class of DA receptors can support the value-based decision making processes underlying probabilistic reversal learning. In other words, the effects of blockade of one subclass of receptors can be compensated for by functional activity in the other class of receptors. Agonist treatment may, we speculate, more profoundly disrupt signaling in the neural circuit, leading to more obvious changes in behavior.

In addition to the effects on the conventional performance measures and the reinforcement learning parameters, all drugs made the animals' responses significantly slower (i.e., increased response latency). Furthermore, all drugs, except for raclopride, decreased the number of trials completed in the task, indicative of psychomotor slowing or reduced attention [44, 46]. The finding that the pattern of effects on response latency and trials completed was symmetrical for the antagonists and agonists (Fig. 1c, e), suggests that DA signaling normally acts at an optimal level, and that deviations from that optimum impair behavior.

Striatal subregion-specific effects

The striatal infusion experiments suggested that the effect of the DA D1 receptor agonist SKF82958 on reward learning was exerted in the VS (Table 1), and the effects of systemic quinpirole on the computational model parameters were mostly reproduced in the intra-striatal infusion experiments. First, the decrease in punishment learning was also observed after infusion of quinpirole into the VS. Second, the decreased value estimate of explore/exploit parameter β was seen after infusion of quinpirole into the VS and DLS. However, the effect of systemic quinpirole treatment on reward learning was not observed after intra-striatal infusion of this agonist, suggesting that this effect required simultaneous stimulation of DA D2 receptors in multiple (striatal) regions, or that it was driven by DA D2 receptor stimulation elsewhere in the brain, for example through D2 autoreceptors on midbrain DA neurons. Stimulation of these latter receptors inhibits the activity of DA neurons [47, 48], thus preventing a peak in DA release during positive reward prediction errors, which may explain the decrease in reward learning after systemic quinpirole treatment. Furthermore, the effects of systemic treatment with DA D2 receptor-acting drugs on the response latency and trials completed were also seen after infusion of these drugs into the different parts of the striatum. This is consistent with previously reported effects of intra-VS infusion of DA D2 receptor (ant) agonists on attention and task engagement [49, 50], although the involvement of dorsal striatal D2 receptors in these processes is less clear [51]. However, effects on trials completed and response latency were not seen after intra-striatal treatment with DA D1

receptor-acting drugs, suggesting that the effects of these drugs on these parameters were the result of the combined effects of these drugs in the striatal subregions [49, 50], or that the effects arose from other dopaminoceptive brain areas. Finally, the effects of systemic treatment with the D2 agonist quinpirole on the number of reversals was seen after infusion of this drug into the VS, but the effect on the fraction of rewarded trials was not replicated in the local infusion experiments. Interestingly, an apparent increase was observed in the number of reversals obtained after infusion of D1 antagonist SCH23390 into the VS, which needs to be interpreted with caution given that no effects were observed on the computational model parameters, and that this effect was not seen after systemic treatment with SCH23390.

The effects of DA D2 receptor stimulation with quinpirole on explore/exploit parameter β were driven by action of this drug in the VS and DLS. A decrease in the value of β indicates that the choices of the animals were more explorative by nature, thus being less driven by the value of the two levers. As such, the amount of exploration versus exploitation is a descriptor of behavior that is related to value-based decision making, rather than value-based learning. Although relatively little is known about the neural basis of this aspect of decision making [52], it has been shown that in humans, fMRI bold responses in the striatum (as well as in the ventromedial prefrontal cortex) are related to exploitative decisions (i.e., choosing the highest valued option) [53]. Furthermore, it has recently been shown that the balance between exploration and exploitation in human subjects is related to two genes linked to DAergic function [54].

Our study supports previous notions that DA facilitates reward-related and aversion-related behaviors through action of the D1- versus D2-receptors, respectively. One earlier study that delivered evidence for this notion used real-time place preference and intracranial self-stimulation paradigms to show that optogenetic activation of D1- versus D2-receptor expressing medium spiny neurons in the striatum is experienced as rewarding and aversive by mice, respectively [23]. In line with these findings, it had been shown that intrastriatal infusion of DA D1 and D2 receptor (ant) agonists has dissociable effects on tasks that study cognitive flexibility. For example, intra-VS blockade of D1 or activation of D2 receptors evokes perseverative behavior in a set shifting task, and treatment with a D2, but not a D1 receptor agonist impairs deterministic reversal learning [27]. Furthermore, it has been suggested that VS D1 receptors guide initial acquisition of a visual cue-guided reward-learning task, while D2 receptors mediate flexible switching to a new strategy in this same task [28]. Indeed, endogenous DA primarily stimulates D1- and inhibits D2-receptor expressing neurons [24], and one of the most influential theories about the role of DA in cognition is that it facilitates learning through increases and decreases in release from the midbrain in response to unexpected reward and punishment, respectively [16, 18]. These findings were the basis for neurocomputational models of the basal ganglia that implicate striatal D1 receptor-expressing neurons (through the "direct Go-pathway") in learning from reward, and striatal D2 receptor-expressing neurons (through the "indirect NoGo-pathway") in learning from punishment [24, 25, 55, 56], and thus made a specific prediction of the mechanism through which DA D1 and D2 receptors guide reward versus aversion-related behaviors. Here, we provide further evidence to support the notion that the DA D1 and D2 receptors may serve functions of opposing emotional valence, and that, in accordance with these neurocomputational models, this happens by mediating *learning* on the basis of positive versus negative feedback, rather than value-based *decision making*. That said, the fact that effects were only seen after treatment with DA receptor agonists, but not with antagonists, indicates that such claims should be interpreted with caution.

The observed effects of DA D1 and D2 receptors on learning on the basis of positively and negatively valenced feedback support several other observations that have recently been made. For example, our lab has recently shown that an abundance of DA in the VS evokes behavioral changes characterized by insensitivity to loss and punishment, leading to increased risk taking [21]. Here, we provide evidence that this phenomenon is driven by overstimulation of VS DA D2 receptors. In line with this, it has been shown that treatment with a DA D2, but not D1 receptor agonist attenuates risk-taking behavior in rats [57]. Moreover, risky choice behavior has been shown to be mediated by striatal D2 receptor-expressing neurons, likely due to the function of these cells signaling "prior unfavorable outcomes" (i.e., negative feedback) [29].

Finally, although our work provides some important insights into the behavioral topography of the striatum, it is important to note that this structure is highly heterogeneous, with a functional and anatomical gradient along all of its axes [31–33]. One limitation of our study is that our infusions were targeted at the anterior parts of the striatum, and did not distinguish between subregions of the VS. Follow-up studies using contemporary viral vector-based techniques are necessary to further anatomically specify our findings. Another limitation of our work is that we only tested a single dose of the drugs in the intracranial infusion experiments, so that the possibility that a lower or higher dose would have differential effects on behavior cannot fully be excluded.

CONCLUDING REMARKS

Learning and decision making on the basis of emotionally valenced information are fundamental abilities for an organism to thrive and survive in a changeable environment, and DA has been widely implicated in these processes. Here, we used a pharmacological and computational approach to investigate how DA D1 and D2 receptors contribute to four important building blocks of value-based learning and decision making: reward learning, punishment learning, choice perseveration, and exploration versus exploitation. Our research confirms previous notions of a role for the DA D2 receptor in aversion-related behavior and the D1 receptor in reward-related behavior, and show that this may be driven by mediating fundamental value-based learning processes.

FUNDING AND DISCLOSURE

This work was supported by the European Union Seventh Framework Programme under grant agreement number 607310 (Nudge-IT). We thank Mauri van den Heuvel for help with the behavioral experiments. The authors declare no competing interests.

ADDITIONAL INFORMATION

Supplementary Information accompanies this paper at (https://doi.org/10.1038/s41386-019-0454-0).

REFERENCES

1. Dayan P, Daw ND. Decision theory, reinforcement learning, and the brain. Cogn Affect Behav Neurosci. 2008;8:429–53.
2. Rescorla RA, Wagner AR. A theory of Pavlovian conditioning: variations in the effectiveness of reinforcement and nonreinforcement. Class Cond II: Curr Res theory. 1972;2:64–99.
3. Sutton RS, Barto AG. Reinforcement learning: an introduction. MIT press; Cambridge, MA (United States) 1998.
4. Ernst M, Paulus MP. Neurobiology of decision making: a selective review from a neurocognitive and clinical perspective. Biol Psychiatry. 2005;58:597–604.
5. Garon N, Moore C, Waschbusch DA. Decision making in children with ADHD only, ADHD-anxious/depressed, and control children using a child version of the Iowa gambling task. J Atten Disord. 2006;9:607–19.
6. Grant S, Contoreggi C, London ED. Drug abusers show impaired performance in a laboratory test of decision making. Neuropsychologia. 2000;38:1180–7.

7. Johnson SL. Mania and dysregulation in goal pursuit: a review. Clin Psychol Rev. 2005;25:241–62.
8. Murphy FC, Rubinsztein JS, Michael A, Rogers RD, Robbins TW, Paykel ES, et al. Decision-making cognition in mania and depression. Psychol Med. 2001;31:679–93.
9. Noel X, Brevers D, Bechara A. A neurocognitive approach to understanding the neurobiology of addiction. Curr Opin Neurobiol. 2013;23:632–8.
10. Roger RD, Everitt BJ, Baldacchino A, Blackshaw AJ, Swainson R, Wynne K, et al. Dissociable deficits in the decision-making cognition of chronic amphetamine abusers, opiate abusers, patients with focal damage to prefrontal cortex, and tryptophan-depleted normal volunteers: evidence for monoaminergic mechanisms. Neuropsychopharmacology. 1999;20:322–39.
11. Verharen JPH, Adan RAH, Vanderschuren LJMJ. How reward and aversion shape motivation and decision making: a computational account. Neuroscientist. 2019. https://doi.org/10.1177/1073858419834517.
12. Berridge KC. The debate over dopamine's role in reward: the case for incentive salience. Psychopharmacology. 2007;191:391–431.
13. Robbins TW, Everitt BJ. A role for mesencephalic dopamine in activation: commentary on Berridge (2006). Psychopharmacology. 2007;191:433–7.
14. Salamone JD, Correa M. The mysterious motivational functions of mesolimbic dopamine. Neuron. 2012;76:470–85.
15. Alexander GE, Crutcher MD. Functional architecture of basal ganglia circuits: neural substrates of parallel processing. Trends Neurosci. 1990;13:266–70.
16. Keiflin R, Janak PH. Dopamine prediction errors in reward learning and addiction: from theory to neural circuitry. Neuron. 2015;88:247–63.
17. Schultz W. Dopamine reward prediction-error signalling: a two-component response. Nat Rev Neurosci. 2016;17:183–95.
18. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. Science. 1997;275:1593–601.
19. Cools R, Frank MJ, Gibbs SE, Miyakawa A, Jagust W, D'Esposito M. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. J Neurosci. 2009;29:1538–43.
20. Floresco SB. Prefrontal dopamine and behavioral flexibility: shifting from an "inverted-U" toward a family of functions. Front Neurosci. 2013;7:62.
21. Verharen JPH, de Jong JW, Roelofs TJ, Huffels CFM, van Zessen R, Luijendijk MC, et al. A neuronal mechanism underlying decision-making deficits during hyperdopaminergic states. Nat Commun. 2018;9:731.
22. Clatworthy PL, Lewis SJG, Brichard L, Hong YT, Izquierdo D, Clark L, et al. Dopamine release in dissociable striatal subregions predicts the different effects of oral methylphenidate on reversal learning and spatial working memory. J Neurosci. 2009;29:4690–6.
23. Kravitz AV, Tye LD, Kreitzer AC. Distinct roles for direct and indirect pathway striatal neurons in reinforcement. Nat Neurosci. 2012;15:816–8.
24. Surmeier DJ, Ding J, Day M, Wang Z, Shen W. D1 and D2 dopamine-receptor modulation of striatal glutamatergic signaling in striatal medium spiny neurons. Trends Neurosci. 2007;30:228–35.
25. Collins AGE, Frank MJ. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. Psychol Rev. 2014;121:337.
26. Danjo T, Yoshimi K, Funabiki K, Yawata S, Nakanishi S. Aversive behavior induced by optogenetic inactivation of ventral tegmental area dopamine neurons is mediated by dopamine D2 receptors in the nucleus accumbens. Proc Natl Acad Sci USA. 2014;111:6455–60.
27. Haluk DM, Floresco SB. Ventral striatal dopamine modulation of different forms of behavioral flexibility. Neuropsychopharmacology. 2009;34:2041–52.
28. Yawata S, Yamaguchi T, Danjo T, Hikida T, Nakanishi S. Pathway-specific control of reward learning and its flexibility via selective dopamine receptors in the nucleus accumbens. Proc Natl Acad Sci USA 2012;109:12764–9.
29. Zalocusky KA, Ramakrishnan C, Lerner TN, Davidson TJ, Knutson B, Deisseroth K. Nucleus accumbens D2R cells signal prior outcomes and control risky decision-making. Nature. 2016;531:642–6.
30. Floresco SB. The nucleus accumbens: an interface between cognition, emotion, and action. Annu Rev Psychol. 2015;66:25–52.
31. Lammel S, Lim BK, Malenka RC. Reward and aversion in a heterogeneous midbrain dopamine system. Neuropharmacology. 2014;76:351–9.
32. Voorn P, Vanderschuren LJ, Groenewegen HJ, Robbins TW, Pennartz CM. Putting a spin on the dorsal-ventral divide of the striatum. Trends Neurosci 2004;27:468–74.
33. Yin HH, Ostlund SB, Balleine BW. Reward-guided learning beyond dopamine in the nucleus accumbens: the integrative functions of cortico-basal ganglia networks. Eur J Neurosci. 2008;28:1437–48.
34. Izquierdo A, Brigman JL, Radke AK, Rudebeck PH, Holmes A. The neural basis of reversal learning: An updated perspective. Neuroscience. 2016;345:12–26.
35. Dalton GL, Wang NY, Phillips AG, Floresco SB. Multifaceted contributions by different regions of the orbitofrontal and medial prefrontal cortex to probabilistic reversal learning. J Neurosci. 2016;36:1996–2006.
36. Bari A, Theobald DE, Caprioli D, Mar AC, Aidoo-Micah A, Dalley JW, et al. Serotonin modulates sensitivity to reward and negative feedback in a probabilistic reversal learning task in rats. Neuropsychopharmacology. 2010;35:1290–301.
37. Verharen JPH, Kentrop J, Vanderschuren LJMJ, Adan RAH. Reinforcement learning across the rat estrous cycle. Psychoneuroendocrinology. 2019;100:27–31.
38. Gershman SJ. Empirical priors for reinforcement learning models. J Math Psychol 2016;71:1–6.
39. Mohebi A, Pettibone JR, Hamid AA, Wong JT, Vinson LT, Patriarchi T, et al. Dissociable dopamine dynamics for learning and motivation. Nature. 2019;570:65–70.
40. Groman SM, James AS, Seu E, Crawford MA, Harpster SN, Jentsch JD. Monoamine levels within the orbitofrontal cortex and putamen interact to predict reversal learning performance. Biol Psychiatry. 2013;73:756–62.
41. Clarke HF, Hill GJ, Robbins TW, Roberts AC. Dopamine, but not serotonin, regulates reversal learning in the marmoset caudate nucleus. J Neurosci. 2011;31:4290–7.
42. Balleine BW, O'Doherty JP. Human and rodent homologies in action control: corticostriatal determinants of goal-directed and habitual action. Neuropsychopharmacology. 2010;35:48–69.
43. Rigoux L, Stephan KE, Friston KJ, Daunizeau J. Bayesian model selection for group studies—revisited. Neuroimage. 2014;84:971–85.
44. van Gaalen MM, Brueggeman RJ, Bronius PF, Schoffelmeer AN, Vanderschuren LJ. Behavioral disinhibition requires dopamine receptor activation. Psychopharmacology. 2006b;187:73–85.
45. van Gaalen MM, van Koten R, Schoffelmeer AN, Vanderschuren LJ. Critical involvement of dopaminergic neurotransmission in impulsive decision making. Biol Psychiatry. 2006a;60:66–73.
46. Winstanley CA, Zeeb FD, Bedard A, Fu K, Lai B, Steele C, et al. Dopaminergic modulation of the orbitofrontal cortex affects attention, motivation and impulsive responding in rats performing the five-choice serial reaction time task. Behav Brain Res. 2010;210:263–72.
47. Ford CP. The role of D2-autoreceptors in regulating dopamine neuron activity and transmission. Neuroscience. 2014;282:13–22.
48. White FJ, Wang RY. Pharmacological characterization of dopamine autoreceptors in the rat ventral tegmental area: microiontophoretic studies. J Pharmacol Exp Ther. 1984;231:275–80.
49. Pattij T, Janssen MCW, Vanderschuren LJMJ, Schoffelmeer ANM, Van Gaalen MM. Involvement of dopamine D 1 and D 2 receptors in the nucleus accumbens core and shell in inhibitory response control. Psychopharmacology. 2007;191:587–98.
50. Pezze M, Dalley JW, Robbins TW. Differential roles of dopamine D1 and D2 receptors in the nucleus accumbens in attentional performance on the five-choice serial reaction time task. Neuropsychopharmacology. 2006;32:273.
51. Agnoli L, Mainolfi P, Invernizzi RW, Carli M. Dopamine D1-like and D2-like receptors in the dorsal striatum control different aspects of attentional performance in the five-choice serial reaction time task under a condition of increased activity of corticostriatal inputs. Neuropsychopharmacology. 2013;38:701–14.
52. Cohen JD, McClure SM, Yu AJ. Should I stay or should I go? How the human brain manages the trade-off between exploitation and exploration. Philos Trans R Soc Lond B Biol Sci. 2007;362:933–42.
53. Daw ND, O'Doherty JP, Dayan P, Seymour B, Dolan RJ. Cortical substrates for exploratory decisions in humans. Nature. 2006;441:876–9.
54. Gershman SJ, Tzovaras BG. Dopaminergic genes are associated with both directed and random exploration. Neuropsychologia. 2018;120:97–104.
55. Cools R. Role of dopamine in the motivational and cognitive control of behavior. Neuroscientist. 2008;14:381–95.
56. Frank MJ. Dynamic dopamine modulation in the basal ganglia: a neurocomputational account of cognitive deficits in medicated and nonmedicated Parkinsonism. J Cogn Neurosci. 2005;17:51–72.
57. Simon NW, Montgomery KS, Beas BS, Mitchell MR, LaSarge CL, Mendez IA, et al. Dopaminergic modulation of risky decision-making. J Neurosci. 2011;31:17460–70.