Social Cognitive and Affective Neuroscience, 2024, **19**(1), nsae050

DOI: https://doi.org/10.1093/scan/nsae050
Advance Access Publication Date: 11 July 2024
Original Research – Neuroscience

D2 dopamine receptor expression, reactivity to rewards, and reinforcement learning in a complex value-based decision-making task

Cristina Banuelos[1,2,3], Kasey Creswell[1], Catherine Walsh[4], Stephen B. Manuck[4], Peter J. Gianaros[4,5,‡], Timothy Verstynen[1,2,3,6,*,‡]

[1]Department of Psychology, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[2]Carnegie Mellon Neuroscience Institute, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[3]Center for the Neural Basis of Cognition, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[4]Department of Psychology, University of Pittsburgh, Pittsburgh, PA 15260, United States
[5]Center for the Neural Basis of Cognition, University of Pittsburgh, Pittsburgh, PA 15260, United States
[6]Department of Biomedical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, United States
[*]Corresponding author. Department of Psychology, Carnegie Mellon University, 340 U Baker Hall, Pittsburgh, PA 15213, United States.
E-mail: timothyv@andrew.cmu.edu
[‡]The last two authors are co-senior authors.

Abstract

Different dopamine (DA) subtypes have opposing dynamics at postsynaptic receptors, with the ratio of D1 to D2 receptors determining the relative sensitivity to gains and losses, respectively, during value-based learning. This effective sensitivity to different reward feedback interacts with phasic DA levels to determine the effectiveness of learning, particularly in dynamic feedback situations where the frequency and magnitude of rewards need to be integrated over time to make optimal decisions. We modeled this effect in simulations of the underlying basal ganglia pathways and then tested the predictions in individuals with a variant of the human dopamine receptor D2 (*DRD2*; *−141C Ins/Del* and *Del/Del*) gene that associates with lower levels of D2 receptor expression ($N = 119$) and compared their performance in the Iowa Gambling Task to noncarrier controls ($N = 319$). Ventral striatal (VS) reactivity to rewards was measured in the Cards task with fMRI. DRD2 variant carriers made less effective decisions than noncarriers, but this effect was not moderated by VS reward reactivity as is hypothesized by our model. These results suggest that the interaction between DA receptor subtypes and reactivity to rewards during learning may be more complex than originally thought.

Keywords: dopamine; basal ganglia; DRD2; reinforcement learning; decision making

Introduction

Consider the problem of choosing where to get your lunch: do you go with the food truck that always serves consistent mediocre food or the truck that sometimes serves amazing food, but at other times is simply unpalatable? Formally this represents a reinforcement learning problem (Sutton and Barto 1998) with dynamic, or nonstationary, feedback schedules (Daw et al. 2006), which requires updating the estimated value of each action based on the gains (e.g. deliciousness) or losses (e.g. unpalatable) experienced in the past. From an algorithmic perspective, learning from these gains and losses happens in the form of temporal-difference (TD) learning (Sutton and Barto 1998) that updates the expected value of any given action for any given state of the world. Over time this TD learning can lead to the locally optimal solution for determining action value, known as the Bellman solution (Bellman 1956).

In the brain, TD learning is implemented by phasic dopamine (DA) signals in cortico-basal ganglia-thalamic (CBGT) pathways (Fig. 1a, see Fig. 1b for model equation). The CBGT pathways are organized as a set of computational loops, where each loop can be conceptually thought of as an independent decision (or action) channel (Fig. 1b, Mink 1996, Bogacz 2007, Bogacz and Gurney 2007, Klaus et al. 2017). The goal of the CBGT loops is to integrate information from competing cortical sources to bias downstream selection systems toward one decision or another and then use feedback signals to promote learning that modifies this bias for future decisions (Mink 1996). The canonical model of CBGT pathways relies on three dissociable control pathways:

Received: 24 October 2023; **Revised**: 24 April 2024; **Accepted**: 10 July 2024

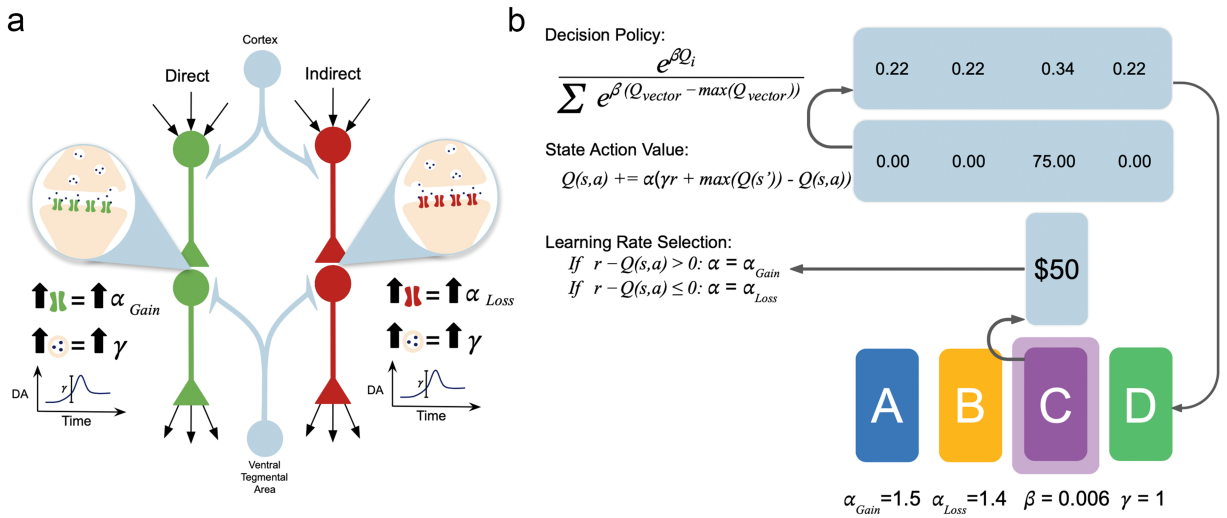

Figure 1. (a) Diagram of the two main pathways in the basal ganglia: direct facilitator pathway with D1 receptor neurons, and the indirect suppression pathway with D2 receptor neurons. (b) The Q-Learning Agent uses an error-driven learning algorithm coupled with the softmax exploration strategy to complete the IGT. The learning paraments of the agent are abbreviated as: r = reward, RPE = r—Q(a), α_{Gain} = Learning Rate when RPE > 0, α_{Loss} = Learning Rate when RPE ≤ 0, β= Inverse Temperature (degree of randomness), γ = Reward strength (representative of DA reactivity).

the direct (facilitation), indirect (suppression), and hyperdirect (braking) pathways. At any given moment, the instantaneous competition between the direct and indirect pathways reflects the strength of bias for a given decision (Collins and Frank 2014, Dunovan and Verstynen 2016, Mikhael et al. 2016, Bariselli et al. 2019). Reward feedback influences this biasing signal via phasic DA signaling (Schultz 1998), which modifies the sensitivity of cortical signals on striatal spiny projection neurons. Specifically, the phasic DA signal is thought to reflect something akin to a reward prediction error (RPE; Schultz et al. 1992, 1997). Positive RPEs (α_{Gain}), i.e. greater than expected gains, sensitize the D1-expressing cells of the direct pathway and depress the D2-expressing cells of the indirect pathway. Negative RPEs (α_{Loss}) do the opposite, enhancing the sensitivity of the indirect pathway while depressing those of the direct pathway (Gurney et al. 2015). This opposing plasticity between the two pathways means that gains reinforce the appropriately selected action over less rewarding alternatives, while losses reduce the saliency of the selected action and allow for more competition between the action channels, pushing the network into a more exploratory state (Cools et al. 2009, Collins and Frank 2014, Stauffer et al. 2014, Vich et al. 2020).

There is another factor that influences how gains and losses impact learning: initial sensitivity to D1 versus D2 in the first place. Particular *DRD2* gene polymorphism variants have been found to associate with functional modulation of DA receptor expression throughout the brain, including the striatum (Zhang et al. 2007). These variants have been shown to have a detectable influence on behavior, particularly learning. For example, individuals with these variants show blunted probabilistic learning in simple bandit-like tasks with strict probabilistic feedback (Frank et al. 2007, Klein et al. 2007, Foll et al. 2009, Frank and Hutchison 2009, Jocham et al. 2009, Gorwood et al. 2012, Klaus et al. 2019). This research indicates that lower striatal D2 receptor density is linked with decreased learning rates, particularly in response to negative reward prediction errors. This suggests a potential mechanistic pathway through which variations in DRD2 expression contribute to differences in learning behavior. By further elucidating the associations between DRD2 variants, striatal D2 receptor density, evoked reward response, and learning parameters to gains and losses, we can gain deeper insights into the underlying neural mechanisms that shape individual differences in feedback-based learning processes.

Previous studies have not demonstrated the possible interaction between striatal D2 receptor density and variations of phasic DA signaling response (approximating γ) to rewards in feedback-based learning. Here we investigate how asymmetries in feedback sensitivity, driven by inherited differences in D2 receptor expression, might interact with phasic DA signals when learning to make value-based decisions in an environment, where reward feedback is dynamic and, in some cases, deceptive. We hypothesize that the presence of the *DRD2* polymorphism variant that associates with lower striatal D2 receptor density interacts with the magnitude of the evoked reward response measured using functional MRI to impact sensitivity to losses during learning in the Iowa Gambling Task (IGT). Due to the lack of individual trial data, we will use a simulation-based evaluation to test our predictions on the effect of the interaction between D2 receptor expression variation and VS reward reactivity variations on effective learning from losses.

Materials and methods

Participants

We used an already collected sample of neurologically healthy adults from southwestern Pennsylvania taken from the University of Pittsburgh's Adult Health and Behavior project, Phase II. The sample consisted of 438 participants (228 females, 210 males, 81.7% White, non-Hispanic) between the ages of 30 and 54 years (M = 42.67, SD = 7.36). Every participant had their blood drawn and genotyped for the presence of *DRD2 –141C Ins/Ins*, *Ins/Del*, or *Del/Del* variants [see Lerman et al. (2005) for a detailed description of this method]. Carriers were defined as having at least one deletion allele (i.e. *DRD2 –141C Ins/Del*, or *Del/Del* variants). The sample consisted of 119 carriers (55 males, 97 *Ins/Del*, 22 *Del/Del*), and 319 noncarriers (155 males). These participants were part of a larger project that included the completion of many other tasks, some of which were completed within the magnetic resonance imaging (MRI) scanner. This research project was approved by the

institutional review boards at both the University of Pittsburgh and Carnegie Mellon University.

Ventral striatal reactivity task

The Cards task was used to assess the reactivity of the ventral striatum (VS) to negative and positive feedback cues associated with monetary gain (Hariri et al. 2006, Gianaros et al. 2011). The task consisted of 45 trials, divided into 9 separate blocks with 5 trials each. Within each trial, the participant was shown a "?" in the center of the card for 3 s, which indicated that the participant needed to now guess whether the following card would be less than or greater than 5. Their choice was indicated by a button press. An index finger press signaled less than 5, and a middle finger press signaled greater than 5. After the guess was made, the number in question was presented for 500 ms, followed by feedback based on the congruence of their response for 500 ms. The number presented was selected by the task based on the block's predetermined positive feedback rate. The feedback was either a green up arrow for positive feedback for a correct response or a red down arrow for negative feedback for an incorrect response. The end of the trial was then signaled with a crosshair presented for 1.5 s. The total length of a trial was 5.5 s.

Participants were instructed that their performance would determine the monetary reward at the end of the task. However, performance was predetermined based on the conditions of the block they were on. Each of the blocks was one of three different conditions: win, loss, or control. In the win condition, there was an 80% positive feedback rate (4 out of 5 correct responses) and a 20% negative feedback rate (1 out of 5 incorrect responses). The opposite was true for the loss condition. In the control condition, instead of receiving feedback or being asked to guess, they were presented with an "x" for 3 s and then instructed to press with either their index or middle finger in response. After pressing, they were then presented with an "*" for 500 ms and then a yellow circle for 500 ms. The block type varied by presenting "Guess Number" for 3 s at the start of each block for the win and loss conditions or "Press Button" for the control condition. The length of the task in total was 350 s.

Participants were scanned on a 3 T Trio TIM whole-body scanner (Siemens, Erlangen, Germany) using a 12-channel phased-array head coil (FOV) = 200 × 200 mm, matrix = 64 × 64, repetition time (TR) = 2000 ms, echo time (TE) = 29 ms and flip angle (FA) = 90° (for more information see Verstynen et al. 2020). While in the MRI scanner, participants completed a computerized reward task paradigm (for preprocessing information see Verstynen et al. 2020). After preprocessing, linear contrast images, reflecting relative BOLD signal changes (i.e. win blocks versus loss blocks), were estimated for each participant using general linear modelestimation. The mean BOLD contrast parameter estimates were extracted from a predefined VS region of interest (ROI) (Gianaros et al. 2011, Verstynen et al. 2020). For more information on the estimation process and creating the a priori ROI mask see Verstynen et al. (2020) and Gianaros et al. (2011), respectively.

Iowa Gambling Task

To measure decision-making in a dynamic and deceptive feedback environment, participants completed a computerized version of the Iowa Gambling Task (IGT). The IGT is a common task for assessing executive function in healthy and clinical populations (Buelow and Suhr, 2009). The participants receive a loan of $2000 and are instructed that the goal of the task is to maximize profits. Although participants were instructed to maximize their overall monetary net gain, they were not provided extra money based on their performance (Verstynen et al. 2020). In the IGT, participants are asked to select a card from any of the four decks presented with a varying amount of reward or punishment (Bechara et al. 1994). The participants specifically select one card at a time from any of the 4 decks for a total of 100 card selections. The exact value and order of each of the cards within the four decks have been predetermined by the experimenters without the participant's knowledge. They are allowed to switch between any of the decks at any time and as often as they wished. The participants are not aware of any of the deck specifications and are only informed that each deck was different. With each selection from Decks A or B (the "disadvantageous decks"), participants have a net loss of money. With each selection from Decks C or D (the "advantageous decks"), participants have a net gain of money or a net zero, respectively. The amount of reward or punishment varies between decks and the position within a deck. Deck A and Deck B both have the same amount of overall net loss. However, in Deck A the reward is less frequent and higher in magnitude, while in Deck B the reward is more frequent and higher in magnitude. Similarly, Decks C and D have the same overall net gain. In Deck C the reward is less frequent and lower in magnitude, while in Deck D the reward is more frequent and higher in magnitude. Furthermore, Decks A and C result in higher frequency losses and Decks B and D result in lower frequency losses. Overall, choosing from Decks A and B results in short-term gains with long-term losses, and choosing from Decks C and D results in short-term losses with long-term gains. From the selections made by the participants, their overall Payoff score (i.e. $Payoff = (C + D) - (A + B)$), and the Sensitivity score (i.e. sensitivity to frequency of rewards, $Sensitivity = (B + D) - (A + C)$ was calculated.

Statistical analysis

Group differences in VS reactivity and Payoff score as well as Sensitivity score were first evaluated using *t*-tests. Follow-up regression models measured how VS reactivity, carrier status, and their interaction associated with the Payoff score as well as with the Sensitivity score. Of particular interest are any potential race effects on gene—behavior associations. However, the non-white portion of the sample in the dataset was small (18.26%), yet made up a significant portion of the carriers (43.70%), making any independent racial group analysis severely underpowered in this dataset. Nonetheless, we used model comparison procedures to determine whether age, self-reported gender, and racial identity needed to be included in the final regression model. The model with the control factors (Akaike information criterion, AIC = 4191, Bayesian information criterion, BIC = 4220) was found to explain a negligible amount of additional information compared to the simpler model (AIC = 4211, BIC = 4228, Bayes factor = 18.370). Thus, for our final regression model, we opted to not include age, gender, and racial identity as control factors.

Reinforcement learning agent

In order to simulate how different reward reactivities and learning rates impacted decision-making, we simulated IGT performance using a standard Q-learning agent with *a* softmax decision policy (Sutton and Barto, 1998). Q-learning is a specific form of TD learning, where updates influence the subjective value of individual actions, as opposed to individual environmental states. The model equations are shown in Fig. 1b. The logic of this model follows a similar structure as the opponent actor learning (OpAL) model, where learning on gains and rewards is independent, but we do not model separate opposing pathways leading to the decision as in the OpAL framework (Collins and Frank 2014). Briefly, on

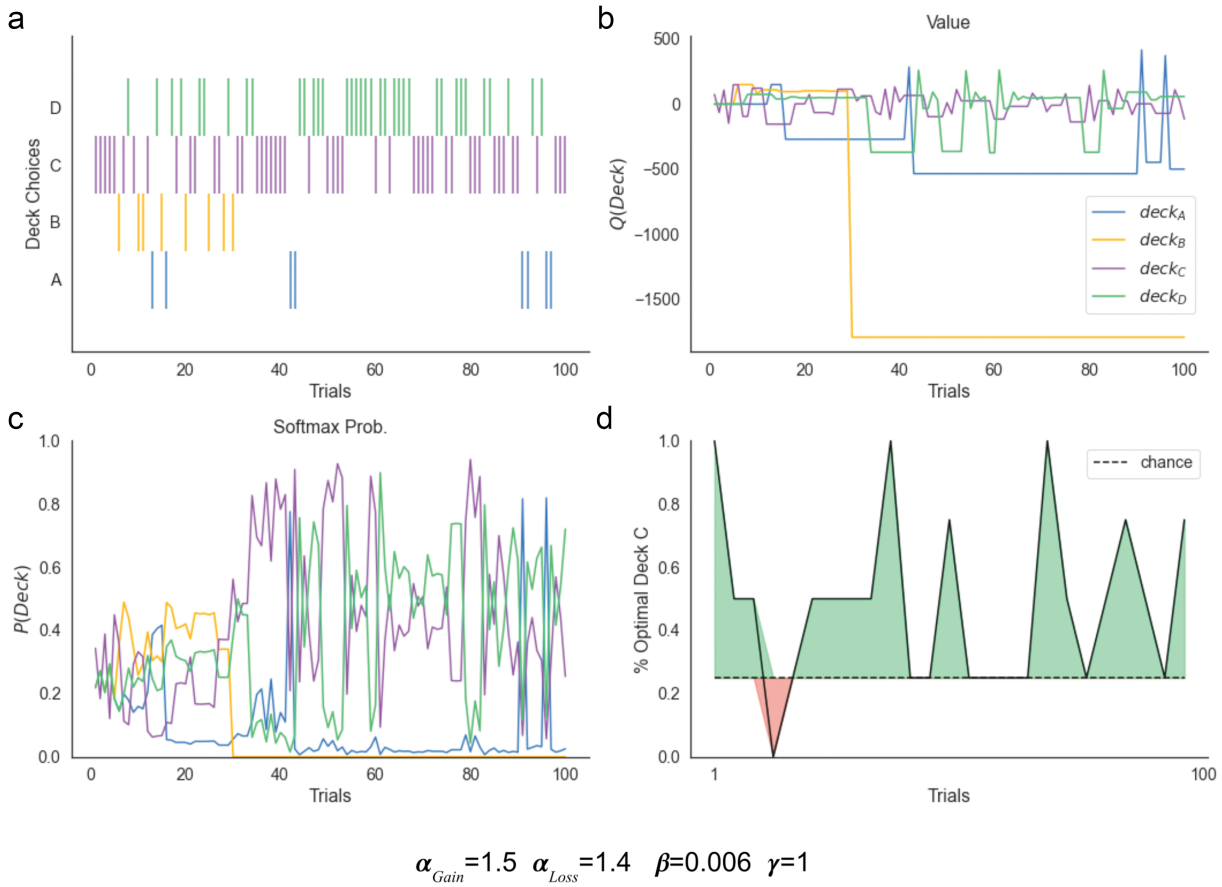

Figure 2. The parameters the agent was given are listed at the bottom of the figure. The agent completed 100 trials or card selections over Decks A–D in which the deck choices were tracked in a raster plot. (a) The predicted deck values were tracked in a line plot, (b) the probability of selecting each deck was tracked in a line plot (c), and the percent optimal Deck C chosen is shown in the area plot (d).

any given trial, the model selects one of four decks using a softmax decision policy (P). The inverse temperature parameter (β), which determines the greediness or randomness of the decision policy, was set to 0.006 for the final model, producing a moderately exploratory agent. After selection, a reward (r) is generated according to the feedback schedule of the IGT. Reactivity to reward is approximated by the scaling term γ, which is directly applied to r. The difference between the experienced reward, γr, and the expected reward, $Q(a)$, produces the RPE that is used to update $Q(a)$ on the next trial. The learning rate (α), determines how much the RPE influences the update of $Q(a)$. To approximate asymmetries in learning, we put a contingency on α. If the RPE is greater than 0, then $\alpha = \alpha_{Gain}$. Otherwise, for negative RPEs, $\alpha = \alpha_{Loss}$.

For each simulation run, an agent was generated with a specific set of values for $\beta, \gamma, \alpha_{Gain}$, and α_{Loss}. The agent completed 100 trials of the IGT with a predetermined deck, where the optimal deck is C. The trial-wise selection of decks across trials was used to calculate Payoff and Sensitivity scores in the same way as estimated for human participants.

Results

Model simulations

In order to understand how asymmetries in learning from gains versus losses can impact the efficiency of decision-making in the IGT, we used a standard Q-learning agent (see "Materials and Methods" section), where we varied the parameters specified in the decision and learning processes. The parameters of this model approximate the differences in decision-making between individuals in the IGT, specifically more exploratory or exploitative decision policies that impact long-term payoffs or losses. These parameters include learning rate from positive RPEs (α_{Gain}), learning rate from negative RPEs (α_{Loss}), "greediness" of the decision policy (β), and overall reactivity to reward (γ; see "Materials and Methods" section and Fig. 1b). The difference between learning on gains versus losses can be reflected as asymmetrical, such as when learning is stronger for cases where the reward value is greater than the expected value (RPE > 0; gains) and weaker when the reward value is less than the expected value (RPE ≤ 0; loss).

Figure 2 shows an example run of one of the agents. We see that over time the deck selections for this agent become more strategic, with a preference for Decks C and D, and with optimal Deck C chosen the majority of the time (Fig. 2a). This preference is reflected as an increase in state-action value (Q) for Deck C in later trials (Fig. 2b), which increases the probability that this deck will be selected over the others (Fig. 2c). As a result, the optimal choices made by the agent increase over time as it effectively uses and learns from feedback, as seen by the percentage of choosing optimal Deck C being above chance consistently after about 40 trials (Fig. 2d).

In order to illustrate how the relative ratio of α_{Gain} and α_{Loss}impacts decision effectiveness (i.e. Payoff score), we ran a series of agents with different learning rate asymmetries and sensitivities to reward (Fig. 3a–c). As expected, the heat maps in

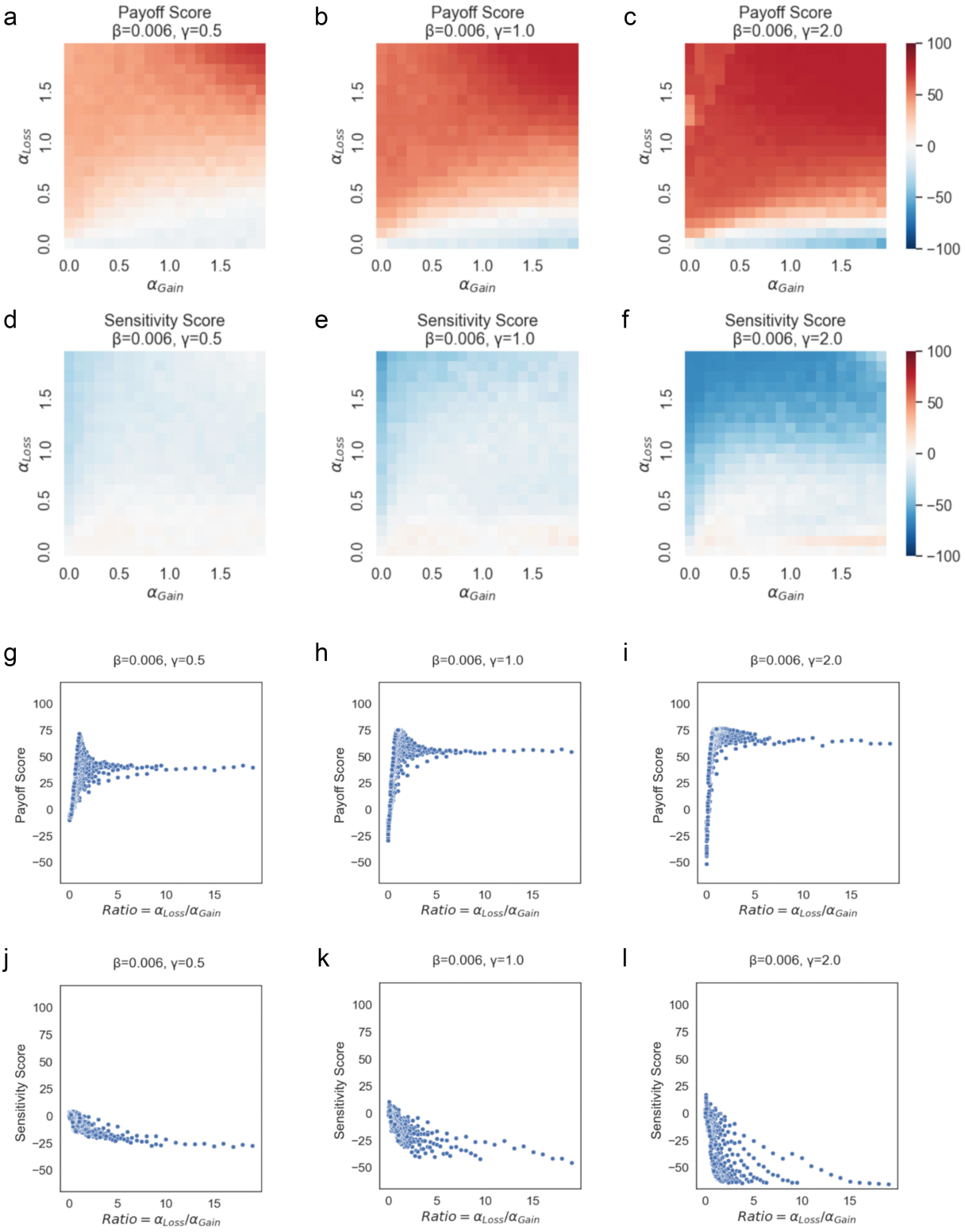

Figure 3. Payoff (a–c) and sensitivity (d–f) scores were plotted in the heatmap for agents with α gain and α loss from 0 to 2, β of 0.006, and γ of 0.5, 1, and 2 (left to right). Payoff (g–i) and Sensitivity (j–l) scores in the IGT were plotted against alpha ratio (α loss/α gain) for β of 0.006, and γ of 0.5, 1, and 2 (left to right).

Fig. 3a–c show that below a relatively low ratio of α_{Loss} to α_{Gain}, the average Payoff score is negative. Payoff scores were greatest when α_{Loss} reached a level that allowed for the agents to learn from their mistakes, when α_{Loss} was greater than 1.5 and the ratio of α_{Loss} to α_{Gain} approximates 1, depending on the overall reactivity to rewards. This relationship between the ratio of α_{Loss} to α_{Gain} and performance in the task was amplified with an increase in reactivity to rewards, γ (Fig. 3g–i), whereas reactivity

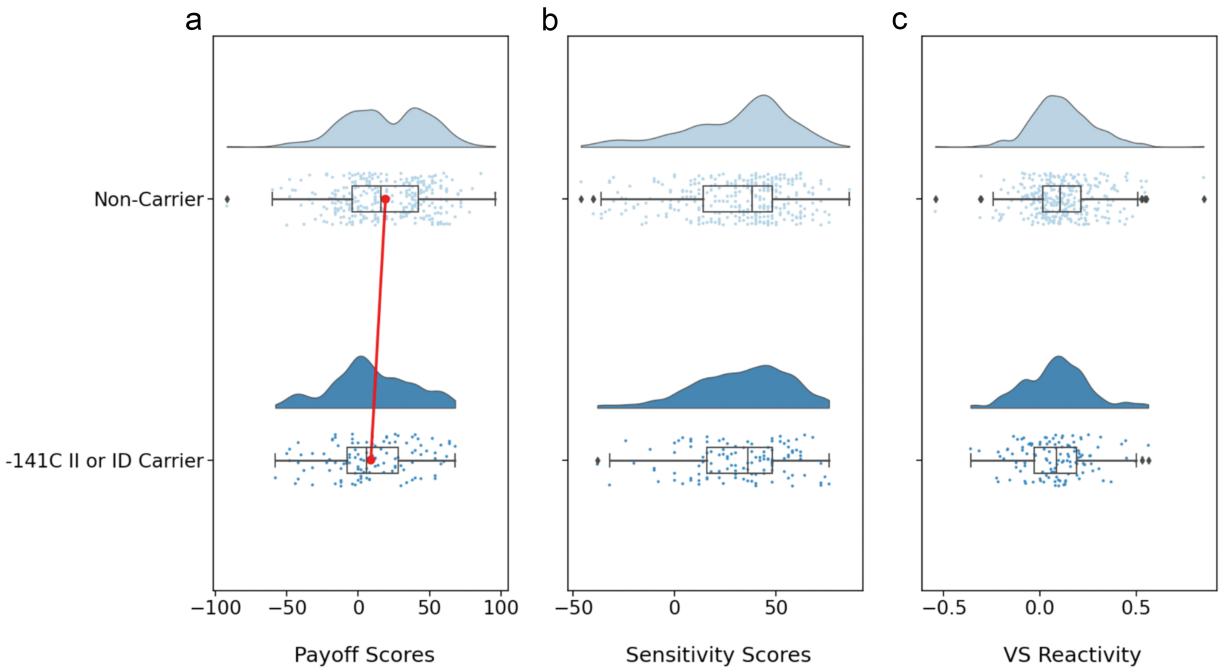

Figure 4. The distributions of measures for IGT (a) Payoff and (b) Sensitivity scores, and (c) VS reactivity for each DRD2 group for noncarriers and carriers were plotted for comparison. The vertical line between mean Payoff scores between noncarriers and carriers demonstrates a significant difference in means.

to frequency of rewards diminished with a greater α_{Loss} to α_{Gain} ratio (Fig. 3j–l). In contrast, the heat maps for the Sensitivity score (Fig. 3d–f) of these agents show an inverted pattern with a slightly positive sensitivity to frequency of rewards below the same threshold of relatively low ratio of α_{Loss} to α_{Gain}. Although the magnitude of this effect of α_{Loss} to α_{Gain} ratio is weaker for Sensitivity scores than for Payoff scores.

These simulations predict that in individuals with reduced learning from losses (i.e. reduced learning from negative RPEs compared to positive RPEs), the overall Payoff scores in the IGT should be lower. The effect should be opposite for Sensitivity scores in the IGT, albeit weaker. Thus, we expect a strong main effect of DA carrier group on Payoff scores, with a possible weaker main effect for Sensitivity scores. If the groups also vary in rate of learning from different reward feedback, this should also result in an interaction between the group and independent measures of reward reactivity.

Empirical data

Our model simulations show that reduced learning from negative feedback signals (α_{Loss}) should reduce Payoff scores and increase Sensitivity scores, and this effect should be scaled by how sensitive, or reactive, an individual is to rewards overall. To empirically test this we first looked at overall group differences in both Payoff and Sensitivity scores and VS reactivity in our sample of human participants. Consistent with our model, Payoff, the measure of effective use of feedback to make decisions, was significantly lower in the DRD2 carrier group than in noncarrier controls (Fig. 4a; $t[436] = -3.230$, $P = .001$). Overall the noncarrier group had a mean Payoff score of 19.16, with a slightly bimodal distribution, whereas the carrier group had a more unimodal distribution, with a mean of 8.89. In contrast, the difference between both the noncarrier and carrier groups in the Sensitivity scores and VS reactivity were not statistically significant, centered just above 31 (Fig. 4b; $t[436] = 0.539$, $P = .590$) and 0 (Fig. 4c; $t[436] = -1.771$, $P = .077$), respectively. Thus, individuals with expected lower D2 receptor density perform worse in the IGT than controls but do not show reliable differences in VS reactivity to rewards or sensitivity to frequency of rewards.

The model predicted that the impact of carrier status (approximating α_{Loss}) on Payoff scores should be stronger in individuals with stronger VS reactivity (approximating γ). Given the lack of group differences in VS reactivity, we did not expect a reliable group-by-VS reactivity interaction on Payoff scores, even though one could still be possible. Figure 5a shows that both carriers and noncarriers have positive-trending slopes for the association between VS reactivity and Payoff score, consistent with our model (see above), with carriers having a slightly shallower trendline compared to the noncarriers.

In regard to sensitivity to high frequency of rewards, Fig. 5d shows a relatively flat slope for both carriers and noncarriers for the association between VS reactivity and Sensitivity score. Indeed, this lack of an interaction effect is born out in a linear regression analysis shown in Tables 1 and 2, respectively. We see that carrier status has a statistically significant, negative effect on Payoff score, replicating the *t*-test results. In contrast, an individual's VS reactivity score correlated with an overall higher average Payoff score. However, the interaction term between the group and VS reactivity was not statistically significant. Thus, we do not see that higher VS reactivity amplifies the effect of group status on Payoff scores.

One assumption in our model simulations is that the effect of reactivity to rewards on the expression of asymmetric learning is monotonic and linear. It is possible that this is not a valid assumption. Therefore, we took a closer look at the interaction between VS reactivity and DRD2 carrier status by binning participants according to VS reactivity quartiles, measuring the main effect of the group in each quartile separately. The four binned groups showed a negative effect of group status on Payoff scores, consistent with

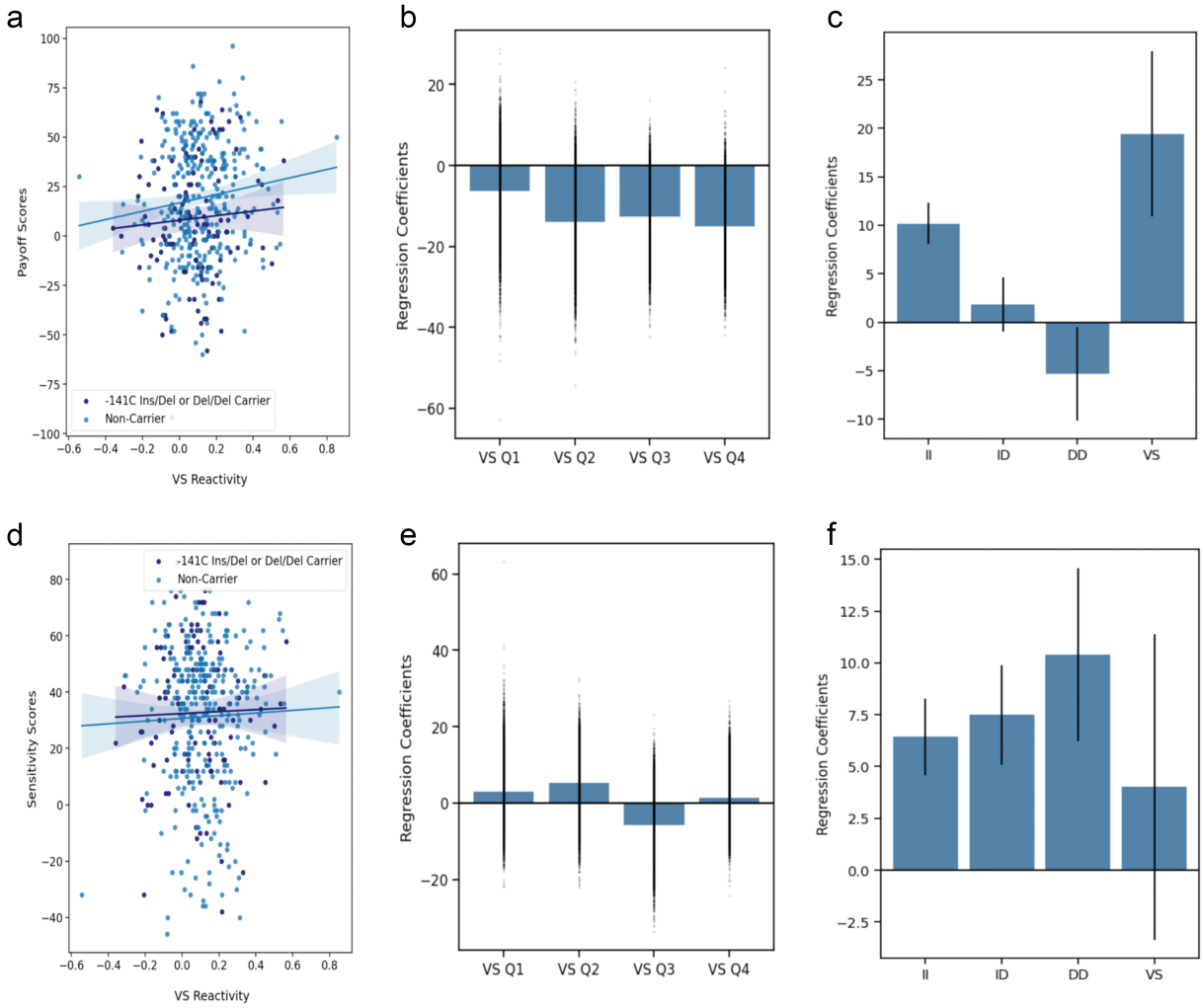

Figure 5. A linear regression model was used for the visualization of the relationships between both Payoff (a) and Sensitivity (d) scores with VS reactivity for both carriers (II, ID) and noncarriers (DD). Bootstrapped regression model coefficients for carrier status on Payoff (b) and Sensitivity (e) scores, binned by VS reactivity quartiles. Regression model coefficients for Payoff (c) and Sensitivity (f) scores versus DRD2 carrier status *–141C Ins*/Ins (II), *Ins/Del* (ID), *Del/Del* (DD), and VS reactivity (VS). Error bars reflect the standard error of the mean.

Table 1. The linear regression model for Payoff was composed of only the main effect variables, the DRD2 group assignment, VS reactivity, and their interaction.

	Coef.	Std. Err.	t	P > |t|	[0.025	0.975]	Summary statistics	
Intercept	16.704	2.020	8.269	0.000	12.733	20.674	Adj. R^2	0.028
DRD2	–8.791	3.634	–2.419	0.016	–15.932	–1.649	AIC	4211
VS	21.140	10.020	2.110	0.035	1.447	40.833	BIC	4228
DRD2:VS	–9.573	18.830	–0.508	0.611	–46.582	27.436	Log-Likelihood	–2101.700
							F-statistic	5.162

The summary results for the linear regression model were also included such as the AIC and BIC scores.

our overall effects. However, this did not increase as VS reactivity increased (Fig. 5b). There was no clear trend between Sensitivity scores in the IGT and VS reactivity (Fig. 5e). This enticing finding suggests that how reactivity to reward interacts with asymmetries in learning may be more complicated than assumed in our simple reinforcement learning model. However, we cannot make any strong conclusions about this relationship due to the high variability across quantiles.

In addition to the binary classification of carriers and noncarriers, we also tested the linear regression model that included all three different *DRD2* polymorphism variants, *–141C Ins/Ins*, *Ins/Del*, and *Del/Del*. To look at this, we plotted the regression coefficients of the Payoff model and the Sensitivity model for the different groups, as well as VS reactivity, in Fig. 5c and f, respectively. Breaking down the main effect model into these three subgroups we see a tiering effect, where an increase in Del alleles is associated with a decrease in Payoff scores (Fig. 5c) and an increase in Sensitivity scores (Fig. 5f). It is important to note that the sample sizes of each respective *DRD2* polymorphism group (*Ins/Del* $N = 97$, *Del/Del* $N = 22$) leave us with low statistical power

Table 2. The linear regression model for Sensitivity was composed of only the main effect variables, the DRD2 group assignment, VS reactivity, and their interaction.

	Coef.	Std. Err.	t	P > |t|	[0.025	0.975]	Summary statistics	
Intercept	30.661	1.746	17.563	0.000	27.230	34.092	Adj. R^2	−0.005
DRD2	1.735	3.140	0.553	0.581	−4.437	7.906	AIC	4083
VS	4.783	8.658	0.552	0.581	−12.234	21.801	BIC	4100
DRD2:VS	−1.308	16.272	−0.080	0.936	−33.289	30.673	Log-Likelihood	−2037.700
							F-statistic	0.219

The summary results for the linear regression model were also included such as the AIC and BIC scores.

to look at more complex models. However, the pattern of main effects across the different carrier subtypes clearly discerns a reliable effect that compliments the main effects seen in the original regression models.

Discussion

Here we investigated how differences in striatal D2 receptor expression could interact with reactivity to rewards to impact feedback-based learning in situations where reward feedback is dynamic and deceptive. Using a simple reinforcement learning model with a simulation-based evaluation, we first showed how asymmetries in learning from gains versus losses should impact performance in the IGT, and how this can be moderated by reward reactivity. Our experimental data in human participants support part of our computationally driven hypothesis, in that individuals with a genetic predisposition for lower D2 receptor expression performed worse on the IGT than controls in terms of overall payoff. However, this effect was not scaled by individual differences in reward reactivity as predicted by our model. Nonetheless, our results further bolster the observation from both the computational modeling and behavioral literature that sensitivities to losses are a critical component in effective long-term learning, particularly in dynamic and deceptive feedback environments.

Our findings align, in part, with the prior literature demonstrating that genetic variants influencing D2 receptor expression also impact goal-directed behavior. For example, Zhang et al. (2007) looked at two single nucleotide polymorphisms (SNPs) that regulate the D2 receptor and found that carriers of the SNPs, which downregulate D2 receptor expression as in our study, showed altered striatal responses during an N-Back working memory task (Zhang et al. 2007). In contrast, we did not see that striatal responses in the functional imaging task differed between carrier groups. This may simply be due to the difference in the cognitive process being measured in the scanner (e.g. working memory versus reward reactivity). Although the BOLD responses during the Cards task, used as a measure of VS reward reactivity, do not directly measure DA neurotransmission, the relative pattern has been found to be consistent with in vivo human striatal DA synthesis as measured positron emission tomography (Siessmeier et al. 2006). Whether this evoked response represents a good proxy for the true phasic DA response requires further testing.

Along these same lines, Klein et al. (2007) showed that individuals with the *TaqA1* polymorphism variant, which is also believed to reduce D2 receptor expression, showed reduced reactivity to errors in a probabilistic avoidance task (Klein et al. 2007). This effect has been replicated multiple times (Frank et al. 2007, Frank and Hutchison 2009). Our work extends this by showing how this reduced reactivity to negative feedback, as a result of genetic predispositions for D2 receptor expression, impacts the efficiency of using feedback in more complex reinforcement scenarios, where frequency and magnitude of gains or losses need to be integrated over time in order to make an optimal decision. Later, Jocham et al. (2009) showed that *TaqA1* polymorphism variant carriers had deficits in reversal learning that consisted of a decreased ability to sustain the newly rewarded response after a reversal and a decreased tendency to stick to the rewarding response in general (Jocham et al. 2009). Taken together with our current work, these findings provide clear evidence that a reduction in D2 efficiency or expression can lead to deficiencies in the integration of feedback from positive and negative signals.

Understanding how our findings relate to the broader behavioral genetics literature on the influence of D2 pathways in cognition should be tempered by the heterogeneity of genetic polymorphisms on the underlying DA pathways. The *TaqA1* (SNP ID: rs1800497), *C957T* (SNP ID: rs6277), and the *−141C Ins/Del* (SNP ID: rs1799732) polymorphism variants are some of the most well understood genetic factors impacting the D2 pathway, although there are many others being discovered and studied (Foll et al. 2009, Gorwood et al. 2012). Both the *−141C Ins/Del* and *TaqA1* polymorphism variants are believed to primarily impact dopaminergic signaling by lowering the D2 receptor density in the striatum (Pohjalainen et al. 1998, Jönsson et al. 1999), whereas the *C957T* polymorphism is believed to be impacting DA D2 receptor availability by affecting the receptor affinity to DA (Hirvonen et al. 2009a, 2009b, Smith et al. 2017). Furthermore, the *−141C Ins/Del* and *C957T* polymorphisms are believed to be directly on the DRD2 gene (Arinami et al. 1997, Hirvonen et al. 2004), whereas the *TaqA1* polymorphism is located downstream of the *DRD2* gene in the ankyrin repeat and kinase domain containing 1 (*ANKK1*) gene (Neville et al. 2004). Trying to integrate findings across these disparate genetic markers in order to come to a mechanistic understanding of how DA, particularly D2, pathways influence behavior requires a careful accounting of the different influences these mutations have on the underlying neural circuitry.

It is also worth noting that while our results and prior work suggest that mutations impacting D2 pathways impact high-level decisions, not all evidence points in this direction. A recent meta-analysis by Klaus *et al.* was not able to establish significant associations between the *TaqA1* and *C957T* polymorphism variants and any of the executive function domains tested, which included a variety of batteries measuring working memory, response inhibition, and cognitive flexibility (Klaus et al. 2019). The results of this systematic review suggest that the presence of *TaqA1* and *C957T* polymorphism variants and their impact on DA D2 receptor signaling may have a limited effect on high-level executive function. Although, it is also possible that the neuropsychological batteries covered in the review by Klaus and colleagues may

not be sensitive to subtle variations in cognitive ability that may be driven by differences in DA receptor expression.

So where do the present results, and the broader literature, leave us in terms of understanding the role of different DA systems during learning, particularly in contexts where feedback is dynamic and deceptive? Of primary importance for future work determining the mechanism by which reactivity to feedback signals interacts might reward reactivity. Our reinforcement learning model, as well as general intuition, shows clearly that these two factors should interact, yet we failed to find this in our data [but see Verstynen et al. (2020)]. One possibility could be the need to find a better or more specific, marker of phasic DA responses. This would likely require moving to more invasive methods, likely in nonhuman model populations. Another possibility is that feedback signal reactivity may have a nonlinear interaction with VS reactivity, and a nonlinear interaction model would need to be tested with a much larger sample size capable of discerning this interaction. Another open question is the role of D1 pathways in this learning process. Our model assumes that learning on gains and losses are equally important for effective long-term value learning, but our work, as well as prior work, only looks at the role of D2 pathways and learning from losses. Integrating findings across genetic markers for the different DA pathways would help to fully elucidate the nature of this process. Of course, these concerns can be tested both experimentally and theoretically using biologically realistic models of corticostriatal plasticity during learning (Gurney et al. 2015, Vich et al. 2020). Working out these precise mechanisms of both learning on gains and losses as well as their possible interaction with reward reactivity is left to future work.

Conflict of interest

None declared.

Funding

This work was supported in part by the National Institutes of Health [P01HL040962, R01089850, R01DA053014].

Data availability

All code and data used in this article can be found at the following Github repository: https://github.com/cbanuelos/DRD2_Abnormal_Learning.

References

Arinami T, Gao M, Hamaguchi H *et al*. A functional polymorphism in the promoter region of the dopamine D2 receptor gene is associated with schizophrenia. *Hum Mol Genet* 1997;**6**:577–82. https://doi.org/10.1093/hmg/6.4.577

Bariselli S, Fobbs WC, Creed MC *et al*. A competitive model for striatal action selection. *Brain Res* 2019;**1713**:70–79. https://doi.org/10.1016/j.brainres.2018.10.009

Bechara A, Damasio AR, Damasio H *et al*. Insensitivity to future consequences following damage to human prefrontal cortex. *Cognition* 1994;**50**:7–15. https://doi.org/10.1016/0010-0277(94)90018-3

Bellman R. Dynamic programming and lagrange multipliers. *Proc Natl Acad Sci USA* 1956;**42**:767–69. https://doi.org/10.1073/pnas.42.10.767

Bogacz R. Optimal decision-making theories: linking neurobiology with behaviour. *Trends Cogn Sci* 2007;**11**:118. https://doi.org/10.1016/j.tics.2006.12.006

Bogacz R, Gurney K. The basal ganglia and cortex implement optimal decision making between alternative actions. *Neural Comput* 2007;**19**:442–77. https://doi.org/10.1162/neco.2007.19.2.442

Buelow MT, Suhr JA. Construct validity of the Iowa gambling task. *Neuropsychology review* 2009;**19**:102–114.

Collins AGE, Frank MJ. Opponent actor learning (OpAL): modeling interactive effects of striatal dopamine on reinforcement learning and choice incentive. *Psychol Rev* 2014;**121**:337–66. https://doi.org/10.1037/a0037015

Cools R, Frank MJ, Gibbs SE *et al*. Striatal dopamine predicts outcome-specific reversal learning and its sensitivity to dopaminergic drug administration. *J Neurosci* 2009;**29**:1538–43. https://doi.org/10.1523/JNEUROSCI.4467-08.2009

Daw ND, O'Doherty JP, Dayan P *et al*. Cortical substrates for exploratory decisions in humans. *Nature* 2006;**441**:876–79. https://doi.org/10.1038/nature04766

Dunovan K, Verstynen T. Believer-skeptic meets actor-critic: rethinking the role of basal ganglia pathways during decision-making and reinforcement learning. *Front Neurosci* 2016;**10**:1–15. https://doi.org/10.3389/fnins.2016.00106

Foll BL, Le Foll B, Gallo A *et al*. Genetics of dopamine receptors and drug addiction: a comprehensive review. *Behav Pharmacol* 2009;**20**:1–17. https://doi.org/10.1097/fbp.0b013e3283242f05

Frank MJ, Hutchison K. Genetic contributions to avoidance-based decisions: striatal D2 receptor polymorphisms. *Neuroscience* 2009;**164**:131–40. https://doi.org/10.1016/j.neuroscience.2009.04.048

Frank MJ, Moustafa AA, Haughey HM *et al*. Genetic triple dissociation reveals multiple roles for dopamine in reinforcement learning. *Proc Natl Acad Sci USA* 2007;**104**:16311–16. https://doi.org/10.1073/pnas.0706111104

Gianaros PJ, Manuck SB, Sheu LK *et al*. Parental education predicts corticostriatal functionality in adulthood. *Cereb Cortex* 2011;**21**:896–910. https://doi.org/10.1093/cercor/bhq160

Gorwood P, Le Strat Y, Ramoz N *et al*. Genetics of dopamine receptors and drug addiction. *Hum Genet* 2012;**131**:803–22. https://doi.org/10.1007/s00439-012-1145-7

Gurney KN, Humphries MD, Redgrave P *et al*. A new framework for cortico-striatal plasticity: behavioural theory meets in vitro data at the reinforcement-action interface. *PLoS Biol* 2015;**13**:e1002034. https://doi.org/10.1371/journal.pbio.1002034

Hariri AR, Brown SM, Williamson DE *et al*. Preference for immediate over delayed rewards is associated with magnitude of ventral striatal activity. *J Neurosci* 2006;**26**:13213–17. https://doi.org/10.1523/JNEUROSCI.3446-06.2006

Hirvonen M, Laakso A, Någren K *et al*. C957T polymorphism of the dopamine D2 receptor (DRD2) gene affects striatal DRD2 availability in vivo. *Mol Psychiatry* 2004;**9**:1060–61. https://doi.org/10.1038/sj.mp.4001561

Hirvonen MM, Laakso A, Någren K *et al*. C957T polymorphism of dopamine D2 receptor gene affects striatal DRD2 in vivo availability by changing the receptor affinity. *Synapse* 2009a;**63**:907–12. https://doi.org/10.1002/syn.20672

Hirvonen MM, Lumme V, Hirvonen J *et al*. C957T polymorphism of the human dopamine D2 receptor gene predicts extrastriatal dopamine receptor availability in vivo. *Prog Neuro Psychopharmacol Biol Psychiatry* 2009b;**33**:630–36. https://doi.org/10.1016/j.pnpbp.2009.02.021

Jocham G, Klein TA, Neumann J *et al*. Dopamine DRD2 polymorphism alters reversal learning and associated neural activity.

J Neurosci 2009;**29**:3695–704. https://doi.org/10.1523/JNEUROSCI.5195-08.2009

Jönsson EG, Nöthen MM, Grünhage F *et al*. Polymorphisms in the dopamine D2 receptor gene and their relationships to striatal dopamine receptor density of healthy volunteers. *Mol Psychiatry* 1999;**4**:290–96. https://doi.org/10.1038/sj.mp.4000532

Klaus A, Martins GJ, Paixao VB *et al*. The spatiotemporal organization of the striatum encodes action space. *Neuron* 2017;**96**:949. 10.1016/j.neuron.2017.10.031

Klaus K, Butler K, Curtis F *et al*. The effect of ANKK1 Taq1A and DRD2 C957T polymorphisms on executive function: a systematic review and meta-analysis. *Neurosci Biobehav Rev* 2019;**100**:224–36. https://doi.org/10.1016/j.neubiorev.2019.01.021

Klein TA, Neumann J, Reuter M *et al*. Genetically determined differences in learning from errors. *Science* 2007;**318**:1642–45. https://doi.org/10.1126/science.1145044

Lerman C, Jepson C, Wileyto EP *et al*. Role of functional genetic variation in the dopamine D2 receptor (DRD2) in response to bupropion and nicotine replacement therapy for tobacco dependence: results of two randomized clinical trials. *Neuropsychopharmacol* 2005;**31**:231–42. https://doi.org/10.1038/sj.npp.1300861

Mikhael JG, Bogacz R, Blackwell KT. Learning reward uncertainty in the basal ganglia. *PLoS Comput Biol* 2016;**12**:e1005062. 10.1371/journal.pcbi.1005062

Mink JW. The basal ganglia: focused selection and inhibition of competing motor programs. *Prog Neurobiol* 1996;**50**:381–425.

Neville MJ, Johnstone EC, Walton RT. Identification and characterization of ANKK1: a novel kinase gene closely linked to DRD2 on chromosome band 11q23.1. *Hum Mutat* 2004;**23**:540–45. https://doi.org/10.1002/humu.20039

Pohjalainen T, Rinne JO, Någren K *et al*. The A1 allele of the human D 2 dopamine receptor gene predicts low D 2 receptor availability in healthy volunteers. *Mol Psychiatry* 1998;**3**:256–60. https://doi.org/10.1038/sj.mp.4000350

Schultz W. Predictive reward signal of dopamine neurons. *J Neurophysiol* 1998;**80**:1–27. https://doi.org/10.1152/jn.1998.80.1.1

Schultz W, Apicella P, Scarnati E *et al*. Neuronal activity in monkey ventral striatum related to the expectation of reward. *J Neurosci* 1992;**12**:4595–610. https://doi.org/10.1523/JNEUROSCI.12-12-04595.1992

Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science* 1997;**275**:1593–99. https://doi.org/10.1126/science.275.5306.1593

Siessmeier T, Kienast T, Wrase J *et al*. Net influx of plasma 6-[18F]fluoro-L-DOPA (FDOPA) to the ventral striatum correlates with prefrontal processing of affective stimuli. *Eur J Neurosci* 2006;**24**:305–13. https://doi.org/10.1111/j.1460-9568.2006.04903.x

Smith CT, Dang LC, Buckholtz JW *et al*. The impact of common dopamine D2 receptor gene polymorphisms on D2/3 receptor availability: C957T as a key determinant in putamen and ventral striatum. *Transl Psychiatry* 2017;**7**:e1091. https://doi.org/10.1038/tp.2017.45

Stauffer WR, Lak A, Schultz W. Dopamine reward prediction error responses reflect marginal utility. *Curr Biol* 2014;**24**:2491–500. https://doi.org/10.1016/j.cub.2014.08.064

Sutton RS, Barto AG. *Introduction to Reinforcement Learning*. Cambridge: MIT Press, 1998.

Verstynen T, Dunovan K, Walsh C *et al*. Adiposity covaries with signatures of asymmetric feedback learning during adaptive decisions. *Soc Cogn Affect Neurosci* 2020;**15**:1145–56. https://doi.org/10.1093/scan/nsaa088

Vich C, Dunovan K, Verstynen T *et al*. Corticostriatal synaptic weight evolution in a two-alternative forced choice task: a computational study. *Commun Nonlinear Sci Numer Simul* 2020;**82**:105048. https://doi.org/10.1016/j.cnsns.2019.105048

Zhang Y, Bertolino A, Fazio L *et al*. Polymorphisms in human dopamine D2 receptor gene affect gene expression, splicing, and neuronal activity during working memory. *Proc Natl Acad Sci USA* 2007;**104**:20552–57. https://doi.org/10.1073/pnas.0707106104

Social Cognitive and Affective Neuroscience, 2024, 19(1), nsae050, DOI: https://doi.org/10.1093/scan/nsae050, Advance Access Publication Date: 11 July 2024, Original Research – Neuroscience

Received: 24 October 2023; **Revised**: 24 April 2024; **Accepted**: 10 July 2024