HHS Public Access

Author manuscript

Neuroimage. Author manuscript; available in PMC 2023 March 03.

Published in final edited form as:

Neuroimage. 2023 February 15; 267: 119831. doi:10.1016/j.neuroimage.2022.119831.

Striatal dopamine supports reward expectation and learning: A simultaneous PET/fMRI study

Finnegan J Calabro[a,b,*], **David F Montez**[a,d], **Bart Larsen**[c], **Charles M Laymon**[b,e], **William Foran**[a], **Michael N Hallquist**[a,f], **Julie C Price**[e,g], **Beatriz Luna**[a,c]

[a]Department of Psychiatry, University of Pittsburgh, Pittsburgh, PA, USA

[b]Department of Bioengineering, University of Pittsburgh, Pittsburgh, PA, USA

[c]Department of Psychology, University of Pittsburgh, Pittsburgh, PA, USA

[d]Department of Neurology, Washington University School of Medicine, St. Louis, MO, USA

[e]Department of Radiology, University of Pittsburgh, Pittsburgh, PA, USA

[f]Department of Psychology and Neuroscience, University of North Carolina at Chapel Hill, Chapel Hill, NC, USA

[g]Athinoula A. Martinos Center for Biomedical Imaging, Department of Radiology, Massachusetts General Hospital and Harvard Medical School, Charlestown, MA, USA

Abstract

Converging evidence from both human neuroimaging and animal studies has supported a model of mesolimbic processing underlying reward learning behaviors, based on the computation of reward prediction errors. However, competing evidence supports human dopamine signaling in the basal ganglia as also contributing to the generation of higher order learning heuristics. Here, we present data from a large (N = 81, 18–30yo), multi-modal neuroimaging study using simultaneously acquired task fMRI, affording temporal resolution of reward system function, and PET imaging with [^{11}C]Raclopride (RAC), assessing striatal dopamine (DA) D2/3 receptor binding, during

*Corresponding author. fjc20@pitt.edu (F.J. Calabro).

Financial disclosures

All authors confirm they have no financial disclosures or conflicts of interest.

Data and code availability statement

All data was acquired within the Laboratory of Neurocognitive Development, with imaging data acquired at the University of Pittsburgh Medical Center (UPMC) Magnetic Resonance Research Center (MRRC). The PET imaging data used in this study has been previously released through in OpenNeuro with the identifier 10.18112/openneuro.ds002385.v1.0.1. fMRI data will be distributed through the NIMH Data Archive (NDA). All code for custom pre- and post-processing will be made available through the LNCD github page upon publication (https://github.com/LabNeuroCogDevel).

Credit authorship contribution statement

Finnegan J Calabro: Conceptualization, Methodology, Formal analysis, Writing – original draft, Writing – review & editing, Visualization. **David F Montez:** Methodology, Formal analysis. **Bart Larsen:** Conceptualization, Writing – review & editing. **Charles M Laymon:** Methodology, Investigation. **William Foran:** Software, Data curation. **Michael N Hallquist:** Conceptualization, Methodology, Formal analysis. **Julie C Price:** Conceptualization, Methodology, Writing – review & editing. **Beatriz Luna:** Conceptualization, Funding acquisition, Methodology, Formal analysis, Writing – review & editing, Supervision.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.neuroimage.2022.119831.

performance of a probabilistic reward learning task. Both fMRI activation and PET DA measures showed ventral striatum involvement for signaling rewards. However, greater DA release was uniquely associated with learning strategies (i.e., learning rates) that were more task-optimal within the best fitting reinforcement learning model. This DA response was associated with BOLD activation of a network of regions including anterior cingulate cortex, medial prefrontal cortex, thalamus and posterior parietal cortex, primarily during expectation, rather than prediction error, task epochs. Together, these data provide novel, human in vivo evidence that striatal dopaminergic signaling interacts with a network of cortical regions to generate task-optimal learning strategies, rather than representing reward outcomes in isolation.

Keywords

Reinforcement learning; Dopamine; PET/fMRI

1. Introduction

The functional roles of striatal dopamine (DA) signaling have been extensively studied, due in part to their wide-ranging contributions to learning, motivational, and motor processes. Converging evidence from human neuroimaging studies, including Positron Emission Tomography (PET) and functional magnetic resonance imaging (fMRI), as well as electrophysiology, voltammetry, and optogenetic studies, provide compelling evidence that ventral striatal dopamine signals the value of reward outcomes relative to their expectation (Schultz et al., 1993). Critically, this signaling has been found to support updating of internally tracked value states (i.e., learning) in addition to reward reactivity (Glimcher, 2011) but the relative contribution to learning processes in humans is still not fully understood.

Work in animal models has characterized important contributions of DAergic processes in supporting unique aspects of reward processing and learning. Phasic activity of DA neurons in the ventral tegmental area (VTA) signals the difference between reward receipt and the expected value of that reward (Glimcher, 2011; Schultz, 1986), termed reward prediction error (RPE) (Rescorla and Wagner, 1972). More generally, such responses are generated upon the updating of expectation of total future rewards as part of the temporal difference (TD) model (Niv et al., 2005; Sutton, 1999). By this model, any information that causes a revision of future expected rewards generates an error signal, providing the basis for reward learning behaviors. Recent work has begun to dissociate the contributions of VTA and nucleus accumbens (NAcc) to RPE, with VTA identifying the presence of a reward, and NAcc DA release reflecting reward expectation (Mohebi et al., 2019).

Critically, encoding RPEs in this manner enables DAergic processes to play a key role in the learning of reward contingencies. Striatal DA neurons modulate long term potentiation (LTP) and depression (LTD) of synaptic strength (Pawlak and Kerr, 2008; Shen et al., 2008, for review see Gerfen and Surmeier, 2011), and DAergic activation has been shown to regulate dendritic spine growth (Yagishita et al., 2014), providing mechanisms by which RPEs can affect synaptic plasticity, potentially enabling the updating of future reward

expectations. Neuroimaging studies in humans have provided supporting evidence for a similar role of ventral striatal DA as a key signaling mechanism in reward processing. Prediction error responses have been demonstrated in NAcc activation in fMRI studies (Berns et al., 2001; Niv et al., 2015; Rodriguez et al., 2006), and a growing body of PET studies have supported the view that these can be linked directly to DAergic processes (Pappata et al., 2002; Zald et al., 2004).

Interestingly, reward related ventral striatal (VS) fMRI BOLD activation has been found to extinguish when subjects view rewards passively (Hakyemez et al., 2008), and related PET-derived dopamine responses are small when received rewards are not necessary for optimizing future performance (e.g., learning) (Urban et al., 2012). In contrast, VS engagement has been reported when rewards are successfully used to learn reward contingencies, that is, during successful reward learning (Schönberg et al., 2007), suggesting that the greater task context may be important for recruiting DAergic activity. Recent work has suggested that VS RPE responses may directly support learning by adaptively coding the PE as scaled by the variance of the distribution (Diederen et al., 2016). These contextual contingencies of DA signaling may also point to a role in supporting the formation of higher-level learning heuristics. For example, strategy set shifting recruits a network of ventral striatal-prefrontal cortical regions (Floresco et al., 2009, 2006), and lesions of the nucleus accumbens impair this ability while leaving reversal learning intact (Block et al., 2007; Reading and Dunnett, 1991).

Characterizing the contribution of DA to both momentary prediction error signals and higher-order, contextually sensitive learning heuristics requires both temporal and neurochemical sensitivity. However, the relationships between DAergic signaling and neuronal activation, and the relative contribution of these processes to reward reactivity and reward learning in humans, has been hampered by methodological limitations. While PET imaging can provide a direct measure of different DA processes, measurements occur over extended periods of time (e.g., at the whole session level, typically >30 min) limiting our ability to assess what aspects of learning are associated with DA. fMRI can be used to assess trial, epoch, and condition specific responses to rewards and learning but it does not provide a direct measure of DA. Recent advances have allowed the simultaneous acquisition of both PET and fMRI data, suggesting the possibility of characterizing neural activity with both the spatial and temporal precision of fMRI with the molecular specificity provided by PET, though few studies to date have used this approach to study dopaminergic contributions to reward processing. Comparison of fMRI-based activation with [^{11}C]Raclopride PET imaging has supported the formulation of refined models of DA-evoked neuronal activation (Mandeville et al., 2013), and mechanisms underlying functional connectivity (Kullmann et al., 2021). A recent report using such an approach to characterize reward processing in depression has demonstrated differences in D2/D3 receptor density that are not reflected in fMRI-based reward activation responses (Phillips et al., 2022). Other approaches, while not simultaneous, have shown that pre-synaptic DA function via DAT receptor density are associated with fronto-striatal functional connectivity (Kaiser et al., 2018) and expectation-related activation in the nucleus accumbens (Dubol et al., 2018), supporting a link between striatal DA signaling and fMRI-based measures of cortical and subcortical reward processing.

However, studies using simultaneous DA-ergic PET imaging and fMRI are still relatively uncommon due to technical hurdles, in particular when coupling this approach with PET designs to assess DA signaling, rather than baseline receptor density. Here we used a multi-modal, molecular MRI (mMRI) approach to simultaneously obtain task-related changes in raclopride-based PET measures of D2/D3 receptor function and fMRI trial-level blood oxygen level dependent (BOLD) responses during a reward learning task that assesses both reward receipt and learning processes. By simultaneously acquiring these measures as subjects performed a reward learning task, we characterize how these distinct measures of underlying neural activity reflect computational reward learning mechanisms, both in terms of trial-to-trial activation responses, and in contribution to DA-specific signaling mechanisms. Our results support a model in which reward responses are reflected by neuronal activation of ventral striatal regions, but in which individual differences in striatal DA release, in conjunction with the recruitment of a network of prefrontal cortical regions, support the use of task-optimal learning strategies, rather than momentary signaling of prediction errors alone. Our results provide direct evidence in behaving human subjects that distinct aspects of DA physiology and function support active learning processes in addition to reward reactivity. Understanding the specificity of DA function can inform basic DAergic mechanisms as well as their association with impaired function such as in psychopathology (addiction, psychosis, mood disorders).

2. Material and methods

2.1. Participants

Eighty-one participants (ages 18–30, mean age 23.3 +/− 3.6, 41 female) completed the full testing protocol, which included an in-lab session and a combined MRI/PET session that was performed on a separate day subsequent to the behavioral session. Participants were recruited from the local population, and were excluded for major psychiatric illness affecting themselves or a first degree relative; prior neurological illness or injury including loss of consciousness; clinical syndrome levels as assessed by the Adult Self Report (ASR) scale; pregnancy (assessed by urine test), lactation; drug use within the last month; history of alcohol abuse; or contraindications to both MRI (e.g., metal in body) and PET (e.g., prior recent radiation exposure). Participants were consented for both the behavioral and imaging components of the study, and research protocols were approved by the University of Pittsburgh institutional review board, including the radiation safety committee and radioactive drug research committee.

2.2. Behavioral task

Participants performed six 5-minute blocks of a reward learning task while in the scanner. In the task, participants controlled a frog avatar on a 3 × 3 map which consisted of a grid superimposed on three landmark images (see Fig. 1A). Each map location was pre-determined to have a randomized reward probability of 20%, 50%, or 80%. Three grid locations were assigned to each probability level, and the map layout was chosen to avoid common patterns (e.g., three high reward squares in a row). The underlying probabilities were maintained throughout the six task runs. Participants were instructed to explore the ‘map’ to find the cells with most rewards thus encouraging them to learn probability

contingencies in order to maximize their earnings (up to $25) over the course of the entire session.

Participants were instructed "*In this game, you get to be a frog. Your job is to jump around the game board collecting coins to earn points. In this game you will be working towards the $25 bonus.*" On each trial, participants were presented the option to move to one of two different map locations. The two possible locations were pseudo-randomly selected and offered to ensure that the actual reward sequence was matched across participants. This was done so that all participants earned the maximum number of rewards, allowing for equivalent potential for reward receipt in driving DA involvement, and to ensure that all map locations were explored. The sequence of reward outcomes was predetermined such that the choices offered on each trial preserved the temporal reward structure across subjects, while still allowing subjects to make realistic choices (40 reward, 10 non-reward outcomes for the 80% locations; 25 rewarded, 25 non-rewarded for the 50% locations; 10 rewarded, 40 non-rewarded for the 20% outcomes). On each trial, we tracked which location types that still had outcomes remaining for the outcome type assigned to that trial number and presented two map locations randomly selected from those reward probability levels. E.g., if trial *i* was designated a reward trial, and neither the 80% nor 50% had yet reached their allotted number of rewarded trials (40 and 25, respectively), then two locations from among the six 80% and 50% locations would be randomly chosen and presented as choice options. This schema guaranteed that each reward probability level in aggregate would match the desired reward sequence, and that every participant received the full monetary reward as long as they completed the task, while still allowing for each map location to have its own reward probability.

At the start of each trial, two locations were labeled with hash marks ("#") for a random duration from 1 to 6 s, during which subjects were able to consider which square they would move to (thus providing decision-making time, and the opportunity for reward expectation). After this, the symbols were randomly re-labeled as "1" and "2" so that participants could make their response. After selecting a response, the avatar moved to reflect their new location, and feedback was presented via visual and auditory cues to indicate whether they received a reward ("cha-ching!") or not (buzzer), and if so, the visual stimulus indicated whether it was a low reward (single coin, 75% of rewarded trials) or high reward (pile of coins, 25% of rewarded trials).

Thus, the task is a version of previous multi-arm bandit tasks coupled with a restricted, 2-alternative forced choice task, similar to those which have been used successfully even in younger populations (e.g., Schulz et al., 2019). Given the extended PET acquisition (~30 min to obtain a task estimate), the complexity of the task design and learning load ensured that subjects would need to continue learning throughout the acquisition period in order to reach optimal performance. This design also helped enable sufficient inter-subject variability in the overall success of learning, facilitating comparisons between those who did and did not learn the underlying reward structure of the map.

At the end of each block, participants were presented feedback on their total earnings in arbitrary unit points, which they were told would result in a $25 reward at the end of the

study if they reached a certain range of points. Participants were not given direct feedback about monetary earnings until after the scan was completed. At the completion of the six blocks, participants were tested on their knowledge of which map locations were the most profitable. In this post scan "map test", participants were presented 50 pairs of map locations and were instructed to report which of the two squares was better. No feedback was given during this phase, allowing it to be used as a pure measure of the degree to which participants had learned the map (i.e., there was no exploratory value, as during the main task).

On a portion of trials, after receiving feedback participants performed either a rewarded or unrewarded anti-saccade task. Although this data was modeled in fMRI analyses described below, results are not presented here, as this was designed to probe cognitive processes for a separate study. In order to learn the mechanics of the task, participants were trained using an abbreviated (single 5-min block) version of the task both during the in-lab session which occurred prior to the scan day, and again immediately prior to entering the scanner. Participants were instructed that the maps used during these training sessions would be different than the map used during the scan. Across all analyses incorporating behavioral data, data points more than 3 standard deviations from the mean were considered statistical outliers, and were excluded.

2.3. Reinforcement learning model

To assess performance during the task, we fit a reinforcement learning (RL) model to each participant's trial-by-trial responses. The model tracked the expected value (EV) of each map location following the participant's moves (S), and updated the EV on each trial following feedback via the learning rule

$$V_{i,j}(t+1) = V_{i,j}(t) + \upsilon * \left(R - V_{i,j}(t)\right) \quad (1)$$

where $(i, j) \in S$ where V is the expected value of location (i,j), t is the trial number, $\boldsymbol{\upsilon}$ is the learning rate and R is the reward received on the current trial (0 for no reward, 0.5, for a low reward, and 1 for a high reward; low reward values were determined based on the group mean parameter estimate from a preliminary model using a free parameter bounded by 0,1). To assess model likelihood, probability that the subject would choose location "1″ was modeled via the softmax function, as

$$P_1 = \frac{1}{1 + e^{-\beta(EV_1 - EV_2)}} \quad (2)$$

where β is the temperature parameter, and EV_1 and EV_2 are the expected values of the two map choices for the given trial.

Within this model framework, we considered a number of additional parameters based on a Variational Bayesian Analysis (VBA) using the VBA toolbox in MATLAB (Daunizeau et al., 2014), as described in the Supplemental Material. The consensus model incorporated separate learning rates for positive and negative prediction error outcomes, similar to previous reports (Chase et al., 2010). Overall model likelihood was computed as the

cumulative probability of the choices made across all trials. Free parameters were determined for each subject by minimizing the overall log likelihood of the model. To validate this modeling approach, we performed posterior simulations for each participant based on their best fit model parameters from the winning model (see Supplemental Material and Supplemental Figure S3). These simulations showed that the model parameters produced simulated data which was significantly associated with actual performance among RL-learners (performance * group interaction $p = 0.01$; $p = 0.0002$ among RL-learners), and showed similar improvement across task blocks (block by learner group interaction, $p = 0.014$; among RL Learners, a significant main effect of block, $p = 9.1 \times 10^{-6}$) as observed in our main data, indicating that the winning model and best fit parameters were capable of reproducing key aspects of our behavioral data.

Based on the consensus model, we performed simulations of our task using a variety of model parameters to quantify the association between RL parameters and performance as measured by the proportion of trials with moves to higher reward probability locations. These simulations showed that the softmax temperature parameter, which captures how often the higher EV square was chosen, was a strong driver of overall performance. Additionally, the positive prediction error learning rate exhibited a quadratic relationship with performance (Fig. 1B), consistent with a “sweet spot” for performance, reflecting that at low learning rates, the model failed to accumulate enough information to make optimal choices, whereas at high learning rates, updates were so large that previously learned information was neglected in favor of recent outcomes. Thus, for this task, an optimal learning rate was associated with learning rates near 0.4. A similar pattern was observed for negative PE learning rates. Simulation results were used to define an “optimality” measure of both the positive and negative learning rate, defined as the percentile performance of the rate being used (peak learning rate=100%, minimally performing learning rate=0%).

2.4. MR data acquisition

MRI and PET data were collected simultaneously over 90 min on a 3T Siemens Biograph molecular Magnetic Resonance (mMR) PET/MRI scanner. Participants’ heads were immobilized using pillows placed inside the head coils, and participants were fitted with earbuds for auditory feedback and to minimize scanner noise. A 12-channel head coil was used. Structural images were acquired using a T1 weighted magnetization-prepared rapid gradient-echo (MPRAGE) sequence (TR, 2300 ms; echo time [TE] = 2.98 ms; flip angle, 9°; inversion time [TI] = 900 ms, voxel size = 1.0 × 1.0 × 1.0 mm). Functional images were acquired using a blood oxygen level dependent (BOLD) signal from an echoplanar sequence (TR, 1500 ms; TE, 30 ms; flip angle, 50°; voxel size, 2.3 × 2.3 mm in-plane resolution) with contiguous 2.3mm-thick slices aligned to maximally cover the entire brain.

2.5. MR data analysis

Structural MRI data was preprocessed to extract the brain from the skull, and was warped to the MNI standard brain using both linear (FLIRT) and non-linear (FNIRT) transformations. Task fMRI images were processed using a pipeline designed to minimize the effects of head motion (Hallquist et al., 2013), including 4D slice-timing and head motion correction, wavelet despiking (Patel and Bullmore, 2016), co-registration to the structural image and

non-linear warping to MNI space, local spatial smoothing with a 5 mm Gaussian kernel based on the SUSAN algorithm (Smith and Brady, 1997), intensity normalization, and high pass filtering ($f > 0.0125$ Hz). Frame-wise motion estimates were computed, and volumes containing frame-wise displacement (FD) > 0.9 mm or DVARS (a measure of total brain signal change) > 21 were excluded from analyses.

First level analysis was performed by modeling all trial events in AFNI's 3dDeconvolve (Cox, 1996). Event timing was aligned to the hash mark interval (decision making and expectation, when the available squares were demarcated, but before they were numbered and participants were able to respond), the feedback event (in which participants made their selection and received feedback,), and the preparatory and outcome phases of the embedded anti-saccade task (modeled with GAM HRFs). Both expectation and outcome phases were modeled parametrically based on the reinforcement learning model. Given that the behavioral RL model results suggested separate learning rates for positive and negative outcomes, we modeled these event types separately in the fMRI data. To obtain parametric regressors, we used group mean RL model parameters to generate trial-by-trial estimates of reward expectation and prediction error (Wilson and Niv, 2015). This provided per-trial estimates of expectation and prediction error, which we then used these as the basis of a parametric fMRI analysis, in which we estimated both constant and linear terms for expectation, positive prediction error, and negative prediction error, producing 6 total contrast maps, each modeled with a canonical GAM hemodynamic response function (HRF). The constant term for each of these captured the mean response at the trial interval (epoch) which was independent of the magnitude of the expectation or PE on the given trial, while the linear term captured the extent to which BOLD activation scaled linearly with expectation or PE, respectively. We note that we did not orthogonalize these regressors to facilitate interpretation of each activation map. In all analyses, we included a 2nd order polynomial regressor separately for each run to account for baseline shifts and drift.

Group average activation for each of the 6 contrast maps was assessed by a 1-sample *t*-test using AFNI's 3dTtest++ function. Cluster correction was performed by assessing the spatial autocorrelation of the residuals from each subject's first level analysis and using the mean smoothness parameters as inputs to AFNI's 3dClustsim, using the ACF estimation (Cox et al., 2017). Based on this, we identified clusters of significant activation for each contrast map given a voxel-wise $p<0.01$, and a Bonferroni corrected cluster-wise significance of $\alpha<0.05$ based on 6 comparisons ($\alpha<0.0083$ uncorrected). To assess the relationship to learning and DA release, we then performed a secondary analysis of the data using log(ΔAIC) (an estimate of RL model support for successful task learning) and VST%ΔBP (PET-derived estimates of DA release) as covariates for each of the 6 contrast maps. We restricted this analysis to regions that showed a significant main effect of task activation, and again performed cluster simulations within these restricted maps to identify significant clusters of activation that scaled with learning and DA release, respectively. Cluster significance was assessed for each map based on a voxel-wise $p<0.05$ and a cluster-wise $\alpha<0.05$ corrected based on 6 comparisons ($\alpha<0.0083$ uncorrected). Since our group mean RL parameters included non-RL learners, who were defined based as being poorly fit by the RL model, we repeated this analysis using RL parameters derived from RL-learners only to ensure that inclusion of the non-RL learners did not introduce noise into

the parametric fits. This produced highly correlated activation maps (all voxelwise t>0.8), detailed in Supplemental Table S3.

2.6. PET data acquisition and modeling

The PET acquisition methods and modeling approach have been previously reported (Larsen et al., 2020). Briefly, we used a 90 min bolus+infusion administration of [^{11}C]Raclopride (RAC) consisting of a 33–40 mCi dose. Subjects were at rest for the first 35–40 min of the scan, at which point they began the reward learning task. We used a modified simplified reference tissue model (SRTM) model with a cerebellar reference region to quantify both baseline BPnd (binding potential prior to performing the task), as well as the change in BPnd during the task, modeled as a step function. The quantity γ by which the BPnd decreased during the task is taken as a measure of DA release, since a reduction in BPnd reflects increased DA occupancy. Details of both the PET acquisition and modeling are reported in the Supplemental Material. In addition to voxelwise analyses, anatomical regions were extracted from an atlas defined and validated based on structural T1 and PET ([11C]PHNO) data (Tziortzi et al., 2011). From this atlas, we considered bilateral ventral striatum, pre/postcommisural caudate, and pre/postcommisural dorsal and ventral putamen, to identify regions which overlapped most strongly with the foci of voxelwise results.

3. Results

3.1. Behavioral performance

Performance on our reward learning task was assessed within-task, based on both the proportion of moves made to the higher probability map location and by reinforcement learning (RL) model fit, as well as post-task, based on a post task map learning assessment. RL model fits were characterized based on the ΔAIC compared to a “guessing” model and exhibited a bimodal distribution (see Supplemental Figure S2), which we used to define ‘RL learners’ (ΔAIC>10) and ‘non-RL learners’ (ΔAIC<10). This criteria is based on previous recommendations (Burnham and Anderson, 2002), and provided an approximate median split, with 41 participants classified as RL learners and 40 as non-RL learners (see Supplemental Table S1 for subject characteristics). These groups did not differ on age (RL learners 23.6+/−3.2, non-RL learners 22.9+/−4.1, p = 0.37), gender (RL learners 49% female, non-RL learners 54% female, p = 0.66), or IQ (RL learners 108.5+/−9.4, non-RL learners 108.8+/−10.6, p>0.9). Since RL learners and non-RL learners were defined by whether the RL model explained choice patterns, as expected we found that RL learners showed a greater proportion of optimal moves (defined as choices to the higher true probability location) than non- RL learners (t = 3.3, χ^2 = 11.1, p = 0.0008, see Fig. 2A). Not surprisingly, this difference was not present in the first block (t = 0.57, p = 0.57) when participants had not yet had the opportunity to learn the task, but emerged beginning with the second block (t = 2.36, p = 0.02) as subjects had time to explore and learn the map. Supporting our RL group definitions, during the post-task map learning assessment RL learners identified 77.2% of map pairs correctly, while non- RL learners identified 62.3% of pairs correctly (F = 50.5, p<0.10^{-9}, see Fig. 2B). Finally, both groups showed equivalent response times (RT; t = 0.095, χ^2 = 0.019, p = 0.88), with faster responses with later blocks (t=−6.64, χ^2 = 99.2, p<0.0001) and no group by block RT interaction (t=−0.79, χ^2 = 0.62, p

= 0.42, see Fig. 2C), indicating that both groups were engaged to a similar degree in making task responses. Additionally, performance on the embedded response inhibition antisaccade task did not differ by group in accuracy (t=−0.99, p = 0.33), latency (t = 0.21, p = 0.83), or inter-trial performance variability (t=−0.039, p = 0.97), supporting that subjects in both groups were engaged and attentive throughout the scan session to a similar extent.

3.2. Striatal da release

PET [^{11}C]Raclopride data was analyzed to assess dopamine (DA) release over the entire task period (approximately 30 min). A decreased non-displaceable binding potential (BP_{ND}) as measured during the task period compared to baseline indicates fewer available receptors, consistent with the release and binding of DA which displaces the tracer, thus providing a measure of task-dependent DA release. We assessed DA release voxelwise across the entire striatum to identify striatal regions which had significant changes in their [^{11}C]RAC BP_{ND}. We found four significant clusters which survived cluster correction (family wise p<0.01) of decreased BP_{ND}: bilateral clusters centered on the ventral striatum (VST) regions extending slightly into the precommisural caudate nucleus (CN), and bilateral clusters contained within the precommisural dorsal putamen (PUT) (see Fig. 3A–C).

To characterize the association of dopamine release with reward learning performance, we compared task-related change in RAC BP_{ND} with individual RL model parameters, based on the consensus RL model. We found a significant difference between RL learners and non-RL learners in VST DA release, as indexed by%ΔBP_{ND} (t = 2.11, p = 0.038, see Fig. 3E). This effect was specific to DA release as no such effects were observed for baseline RAC BP (Fig. 3D), and to the ventral striatum as effects were not present in other striatal regions which showed an overall task effect on DA release (Fig. 3E). Furthermore, we found that VST DA release (%ΔBP_{ND}) exhibited a significant quadratic relationship with learning rate for positive outcomes ("+PE"; linear term, t = 2.88, p = 0.0053; mean-centered quadratic term, t=−2.58, p = 0.012, Fig. 3H), such that higher DA release was observed for those with learning rates near 0.5; learning rates that were either very low or very high were associated with diminished DA release. This effect persisted when controlling for learner category (main effect of quadratic term, t=−2.50, p = 0.015, main effect of learner category, t=−3.20, p = 0.002), and did not show a significant interaction with learner category (p>0.6). This quadratic relationship exhibited an inverted U shape, peaking for intermediate learning rate values, which was consistent with a proportional relationship between DA release and the overall optimality of the +PE learning rate (see Supplemental Figure S12; main effect of optimality, t = 3.17, p = 0.002; main effect of learner category t = 2.77, p = 0.007; no significant interaction, p>0.7). We did not find any significant relationship between DA release and either the softmax temperature parameter ("beta", t = 0.54, p = 0.59, Fig. 3G) or the negative outcome learning rate ("−PE"; t = 0.09, p = 0.93, Fig. 3I). Of note, although DA release was associated with the learning rate, it did not significantly correlate with the overall proportion of optimal response (t=−2.9, p>0.7).

3.3. BOLD reward response

fMRI data was analyzed using a parametric activation analysis based on trial-by-trial expectation and prediction error estimates derived from the RL model. Since the behavioral

analysis suggested separate learning rates for positive and negative PE outcomes, we modeled these separately in the fMRI data as well. This produced a total of 6 contrast maps: constant (i.e., mean) and linear (i.e., proportional) terms for each of the expectation, positive PE and negative PE terms. The constant terms captured the mean activation of each epoch across all trials, while the linear term captured the extent to which BOLD activation scaled with the level of expectation or PE respectively among trials. Overall, these 6 contrast maps captured widespread cortical and sub-cortical activation that was largely consistent with previous reports (see Fig. 4 for striatal activation, and Supplemental Figure S5 for whole brain activation maps). Notably given our *a priori* interest in the striatum as a region of interest, we found significant activation for both positive and negative PE throughout the striatum associated with the constant PE term (i.e., present on all trials, Fig. 4, top row). Additionally, we found bilateral activation in the putamen on positive PE trials which scaled with the PE magnitude. Overall, mean activation at the expectation epoch tended to be negative (i.e., deactivations) throughout the striatum, but we identified a cluster in the ventral portion of the caudate which showed increased activation proportional to the trial-by-trial reward expectation.

Within task-activated regions, we performed a follow-up analysis to assess the association of BOLD activation with successful reward learning on our task. For each of the 6 contrast maps, we identified regions with individual differences in activation that scaled with per-subject estimates of ΔAIC relative to a null model, which we used as a proxy for RL-based learning. Notably, we did not identify any significant clusters within the striatum in which activation in any of the six contrasts scaled with learning. We confirmed this result using an alternative fMRI analysis in which we contrasted rewarded to non-rewarded trials in *a priori* anatomically-defined striatal regions of interest, rather than relying on the RL model for parametric estimates (see Supplemental Figures S6 and S7 and Supplemental Table S2 for full activation details). Furthermore, based on this analysis, no association was found between VST BOLD and any of the RL model parameters (Supplemental Figure S9), nor with overall task performance ($t = 0.13$, $p>0.8$). We did identify several cortical clusters whose activation scaled with learning in the model-based analysis of epoch-specific BOLD responses. This included regions of the left supramarginal gyrus, superior medial gyrus, middle temporal gyrus, and inferior temporal gyrus, whose activation at the expectation epoch was inversely proportional to individual differences in learning, as well as clusters in the right angular gyrus and precuneus, whose activation at the feedback epoch on positive PE trials positively correlated to learning (see Table 1 and Fig. 5). Of note, in both cases, correlations were seen with the constant term BOLD response, that is, activation which occurred on all trials regardless of the magnitude of reward expectation or prediction error (see Supplemental Figure S10 for per-cluster correlations).

3.4. Association of bold activation with vst dopamine release

Given the association between VST DA release and model-based learning parameters, we were interested in assessing the relationship of VST DA release and simultaneously obtained BOLD task activation. Thus, we performed a voxelwise correlation analysis between fMRI activation within the task-activated regions described above and the VST DA release assessed based on the change in RAC BP_{ND} during the task. Based on this

analysis, we found that only one fMRI contrast, the mean expectation-related activation, produced cluster-corrected significant activation which was associated with VST DA release. Specifically, we identified four significant clusters which survived cluster correction as well as Bonferroni correction based on the six contrast maps tested. These included regions of the left Rolandic operculum and superior medial gyrus, as well as midline thalamus and anterior cingulate clusters (see Fig. 6 and Table 2). In all four cases, the sign of the association was positive, indicating higher BOLD activation in subjects with greater DA release (see Supplemental Figure S11). Neither positive nor negative PE contrasts produced any significant clusters associated with VST DA release.

4. Discussion

This study provides direct evidence from simultaneous PET-fMRI imaging for dopaminergic contributions to reward learning processes. We leveraged a simultaneous PET-fMRI acquisition to characterize the role of DA during a probabilistic reward learning task. Methodologically, this approach is valuable for linking temporally sensitive fMRI measures to transmitter-specific PET responses. Since learning is a dynamic process, obtaining this information within a single scan session provides an unprecedented view of the functional and neurophysiological processes underlying reward learning. Here, we leveraged a comparison between these imaging measures and learning performance on a probabilistic reward learning task to identify neurophysiological processes associated with reward learning.

To characterize the computational strategies employed by participants in our task, we tested a large number of candidate reinforcement learning (RL) models and found that the most parsimonious account of performance arose from a 3-parameter model which incorporated separate learning rates for positive and negative prediction error outcomes, and a temperature parameter which governed the reliability with which participants selected the higher expected value map location (e.g., Supplemental Figure S1). As aforementioned, we further performed model simulations that showed a reasonable reproduction of participant behavior, supporting the validity of the RL model (Supplemental Figure S3). Still, although this model was able to reproduce key aspects of behavior, as always there remains space for improving the fit with performance. Thus, we consider that this approach indicates that the use of independent learning strategies for positive and negative reward outcomes are among the key parameters governing participants' decision making strategies in our task, supporting the use of valence-dependent learning strategies.

Consistent with previous studies, our results revealed that performing a reward learning task elicits both a robust BOLD activation (Fig. 4 and Supplemental Figures 5 & 6) as well as PET DA response (Fig. 3) in the ventral striatum. Interestingly, while striatal BOLD responses to both reward expectation and prediction errors were present to a similar degree independent of learning performance, there was a significant learning dependent, task-related DA response, which was significantly greater among participants that exhibited reward learning performance consistent with the use of RL learning strategies (Fig. 3E). Further, this DA response was correlated with RL model fit parameters (Figs. 3H and S12), such that greater DA release was associated with the use of a more optimal learning

rate. We operationalize optimality here to refer exclusively to the learning rates utilized by participants based on model simulations of our task (Fig. 1B): very slow learning rates lead to sub-optimal performance since they result in very small updates on each trial, such that participants would need many trials for internal expected value estimates to reach true probability levels, whereas very high learning rates would result in excessive updating on each trial, effectively causing previously learned outcomes to be neglected. Our model simulations confirmed that for this task, an optimal learning rate occurred for learning rate values near 0.4, with performance falling as learning rates varied in either direction, though we note that this optimal value is highly task-specific, and that different choices of reward probabilities, temporal reward dynamics, and other task parameters can significantly affect the value. This value mirrored DA release results, in which greatest DA responses were seen in participants with learning rates near the optimal value (Fig. 3H), as confirmed by a relationship between DA release and the overall optimality of the positive outcome learning rate (Figure S12).

Interestingly, while differences in DA release were related to the optimality of the RL learning rate parameter, they were not directly associated with task performance metrics. The specificity of this effect to learning rate, and not overall performance, may be explained by the role of the temperature parameter in these associations. That is, overall task performance depends on both the learning rate, which sets how participants acquire new information, and the temperature parameter, which determines how reliably participants use this information to pick the “better” choice (Fig. 1B). But while the temperature parameter has a substantial effect on subject performance, it was not directly associated with DA release, confounding associations between DA release and overall task performance. Instead, our results suggest that DA release is most strongly and specifically linked to the effectiveness of the learning rate employed, independent of the resulting reward outcomes, thus pointing to a DA response mechanism that is specifically linked to the optimization of learning processes. This is consistent with previous reports linking DA signaling to balancing the utility of known rewards with opportunity costs (Le Heron et al., 2020) and in encoding action policy uncertainty (Gershman and Uchida, 2019), since both the learning rate and temperature parameter can be seen as a means of balancing expected, known rewards with the utility of exploration. The optimal value of RL parameters varies depending on the specific task being used, such that this association with DA may reflect a role in learning the overall structure of the tasks, rather than just reactivity to reward contingencies of individual map locations.

Of note, individual differences in DA release were not associated with differences in ventral striatal BOLD activation associated with either prediction error or reward expectation. This mirrors recent reports in clinical studies of depression which identified differences in PET-based DA signaling absent any differences in ventral striatal BOLD reward responses (Phillips et al., 2022), and which instead found differences in functional connectivity associated with DA-related group differences. Similarly, recent work from our group has suggested that changes in DAergic function through adolescent development drive changes in fronto-striatal functional connectivity (Parr et al., 2021). These studies suggest potential interactions whereby striatal DA function might more directly modulate activation of other extrastriate regions. Indeed, we found that individual differences in DA release were

uniquely correlated to BOLD activation in a network of cortical regions, including the dorsomedial PFC, ACC, Rolandic operculum, and thalamus (see Fig. 6). Several of these regions have well established links to striatal DA signaling. The thalamus is known as a hub mediating striatal connectivity to the cortex (Alexander et al., 1986). This pathway is critical to mediating well known effects of DAergic inputs on prefrontal cortex activity supporting cognitive control (Ott and Nieder, 2019) via modulation of the ACC (Holroyd and Yeung, 2012) and others, as well as for conditioning stimuli affecting activation in the superior medial gyrus (e.g., Diaconescu et al., 2010; Hartwell et al., 2011), both of which were also identified by our analysis. We note that this analysis cannot assess causality at the time scales we are measuring, since our PET measure is defined over the course of the entire task period (~30 min), while BOLD responses are aggregated over specific epochs distributed throughout the scan. Thus, while it remains equally possible the striatal DA release drives future expectation-related activity, or that prefrontally-mediated expectation responses facilitate striatal DA release, these results indicate that striatal DA may work in tandem with cortical learning mechanisms.

We additionally identified a set of cortical regions that showed differential activation associated with our learning criteria (Fig. 5). Interestingly, activation during the expectation epoch of our task was systematically lower during reward expectation among RL Learners (see also Supplemental Figure S10), while positive outcome PE responses were higher among RL learners. Contrary to our expectation, the strongest associations between BOLD activation and both learner criterion and DA release were present in the ‘constant’, rather than ‘linear’ (i.e., proportional), activation maps. This suggests that DA-driven learning was mostly strongly associated with a change in activation overall (i.e., on all trials), rather than proportionally to the reward expectation or outcome (prediction error). This is similar to our observation of BOLD activation correlated to DA release occurring across all trials, further supporting that DAergic learning in our task may be associating with learning strategies and heuristics, which are applied uniformly across trials, rather than signaling the trial-specific magnitude of reward outcomes alone. Taken together, these results provide support for a model by which NAcc DA release supports reward learning by modulation of cortical BOLD responses, primarily during the reward expectation phase of the task, rather than by directly affecting either positive or negative prediction error responses, suggesting a possible cortical circuitry of subcortically coupled learning.

Striatal DA has been associated with a number of distinct functions, including motor processes, reward response, and learning (Berke, 2018). Dissociating these contributions has been difficult, especially in human studies, in part because these functions typically co-occur. The characterization of learning-related behaviors in our study allows us the opportunity to partially disambiguate these processes. First, since the sensorimotor aspects of the task were matched across all participants, it is relatively unlikely that these could directly contribute to the learning-related differences we observed. In addition, in the fMRI BOLD analyses, we were able to assess how activation scaled with the magnitude of expectation and prediction error. There was no difference in the motor aspects of these contrasts, since the reward classification did not occur until the reward was presented, which was after the completion of the motor response. Thus, motor differences are unlikely to contribute to the parametric BOLD responses presented here. Nor are motor effects likely to

be related to the mean or parametric expectation terms, since no motor response occurred at this task epoch. Motor responses could be present in the mean activation for positive and negative prediction error contrasts, but we note that these contrasts did not show any association with either learning nor VST DA release, and are thus unlikely to be account for these aspects of our results. Equivalent decreases in response time through the task and performance of inhibitory control task in both groups suggests that motivation and engagement was similar across groups and did not underlie the brain functional differences. Finally, we matched the reward schedule across participants, such that all participants received the same pattern of reward feedback and with matched timing. Taken together, these indicate that visuomotor and attentional engagement differences are unlikely to contribute to individual differences in DA release that we have reported.

Separating the contributions of reward receipt and reward learning is even more challenging, since these are often inextricably linked. Work from rodent models has long indicated the reward prediction errors drive DAergic activity in the ventral striatum (Schultz, 2016, 1998), paralleling fMRI studies (Berns et al., 2001; Diederen et al., 2016; Niv et al., 2015; Rodriguez et al., 2006). However, while we observed trial-wise BOLD activation in the ventral striatum, individual differences in this measure were not related to per-subject DA release. One possibility for these results, supported by previous work, is that fMRI may be particularly sensitive to post-synaptic glutamatergic signaling (Attwell and Laughlin, 2016; Logothetis, 2003), and relatively insensitive to DA activity. However, another possibility arises based on our PET finding that greater DA release was associated with learning strategies on our task. Based on this, VST DA release may be playing a different role than VST BOLD, wherein DA sets higher order task strategies and heuristics through interactions with cortical regions as we have discussed, while BOLD activation reflects expectation and prediction error responses, which scale with reward magnitude across trials.

Interestingly, overall response during the reward expectation epoch was seen as a BOLD deactivation relative to baseline, although responses increased (smaller deactivation) as reward probability levels increased. This was likely not due to limitations in the HRF model given the use of temporal jittering and the relatively long ISIs included in our task design, and supported by the observation that the modeled HRF shows both an initial baseline value of 0, and a clear return to 0 after ~16 s (see Supplemental Figure S8) as is typical for task-evoked HRFs. Instead, deactivation may reflect that at each trial only 2 map locations were available for responding vs 7 locations that were not available, though still visually present. Evidence from rodent electrophysiology has shown that a number of VS/NAc neurons show suppression of firing rates during reward expectation, either continuously throughout the task (Taha and Fields, 2006), or specifically during cue sampling (Roesch et al., 2009). Thus, the overall response may reflect suppression of location-specific responses for the 7 unavailable locations, with a comparatively small additive factor from the 2 active choices (which scales parametrically with the reward level of these two locations, as we observed). Anecdotally, we have also observed that subjects tended to have ‘favorite’ map locations. Since only two of nine map locations are available for selection on each trial, most trials would not have included their preferred choice. Thus, the mean expectation response may include a negative response reflecting the failure to receive preferred choices (i.e., a

de facto cue-related prediction error signal), which is then modulated (upwards) if other promising choices are included.

There are several important methodological considerations and limitations when drawing inferences from this study. First, although a number of PET studies have proposed modeling of single-session task designs, such approaches are still relatively less common. A particular concern is that these models may systematically overestimate pre-task binding potential, or mis-estimate other parameters (e.g., k2') of the PET compartment model, creating false positive task effects. We think this is unlikely to account for our results for several reasons. For one, we began our task relatively late in the session (35–40 min), as indicated by prior simulation studies (Wang et al., 2017) and which is relatively conservative compared to other recent approaches (Hamilton et al., 2018), such that pre-task BPnd estimates are less likely to be significantly biased. For another, while any such biases could undermine the main effect of task, they are unlikely to explain the learning-related differences we have seen, since both RL learners & non-RL learners began the task at the same time within their scan sessions. Similarly, a mis-estimation of reference tissue model parameters, such as k2' (the reference tissue rate constant), would affect all voxels, making it less likely that we would see effects constrained to specific clusters within the striatum. Although continued efforts are needed to more fully define optimal modeling approaches and to continue to quantify potential sources of bias (e.g., see Levine et al., 2022), such approaches have the potential to greatly expand our ability to characterize the contribution of DA to behavioral and cognitive processes. A second limitation arises from the modeling approach we have employed with our fMRI data. Specifically, we used a canonical hemodynamic response function (HRF) in order to measure the amplitude scaling of activation across trials in accordance with reward expectation and outcome (prediction error). However, previous work has suggested that the shape of the HRF itself may change depending on reward outcomes (e.g., Li and Jasanoff, 2020), which is consistent with observations we have made during the expectation epoch of our task (see Supplemental Figure S8). Further work is needed to better parametrize these effects in order to measure systematic changes in HRF shape as a function per-trial reward measures in order to model these effects simultaneously.

Finally, our interpretations of associations between DA release and the optimality of the learning rate are limited by the nature of the learning rate estimates: RL model parameters are inextricably contingent upon the particular form of RL model employed, and as discussed above, it remains possible (and indeed likely) that there exist other, untested models that could perform as well or better in explaining our data. Such formulations would likely render different estimates of learning rates, and may identify different associations with DA release. Thus, the associations we present should be interpreted within the context of the best-fitting RL model we have identified, and future work to continue to characterize RL mechanisms employed during reward learning tasks may be valuable in further specifying and refining the nature of the relationships we have described. Regardless, we believe that better fitting models would refine and extend these results, rather than invalidate them. In particular, non-RL learners do not appear to employ strategies consistent with the RL models we have tested. While it is possible that there exist other models that would capture what these participants are doing in the task, any such model would appear to be ineffective: overall performance of these subjects is barely above chance, does not

meaningfully increase across blocks, and does not appear to result in strong recollection of reward contingencies post-scan. Thus, future efforts not only to further characterize RL learning strategies, but to assess individual differences in both RL parametersand the model strategies themselves (e.g., Piray et al., 2019), will be necessary for characterizing the diversity of learning approaches participants employ.

In sum, we suggest that these data support a DAergic contribution to learning that is separate, or in addition to, its role in signaling momentary prediction errors. This process of identification of task structure and tailoring of optimal learning rates may be mediated by an interaction between striatal DA and the activation of a network of executive cortical regions. These results have important implications in understanding the role of dopamine in reward contexts. Dopamine is still often considered in terms of its role in reward reactivity. However, our data provide in vivo evidence from simultaneous PET and fMRI supporting a growing body of literature suggesting that reward receipt alone is not sufficient to account for ventral striatal DA responsiveness (e.g., Hakyemez et al., 2008), but depends critically on the use of rewards as the basis for learning. There are important implications to considering the role of DA in learning beyond reward reactivity for clinical conditions with DA dysfunction that show abnormal reward processing such as in Parkinson's Disease (Skvortsova et al., 2017), which may reflect effects on motivated learning in addition or instead of reward reactivity. Adolescent peaks in sensation seeking (Spear, 2000) are frequently deemed to be underlied by elevated reward reactivity (Luna et al., 2015; Luna and Wright, 2016; Shulman et al., 2016), but these same changes in neurophysiology may impact unique aspects of learning during this time. Our results provide compelling new evidence for multiple roles of dopaminergic function in reward reactivity and learning that can inform comprehensive models of motivation and impact our understanding of lifespan dopaminergic development and clinical dysfunction.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This work was supported by National Institute of Mental Health (NIMH) grant MH080243 to BL and funding from the Staunton Farm Foundation. We would like to thank Evan Morris and Shuo Wang for help in designing the PET paradigm and analysis approach. Data collection was expertly performed by Julia Lecht, Matt Missar, Jen Fedor, Jess Graves, and Laurie Thompson. We are grateful to personnel in both the Magnetic Resonance Research Center (MRRC) and the PET center at UPMC Presbyterian, especially James Ruszkiewicz, Tae Kim, Chan Moon, Brian Lopresti, and Hoby Hetherington, for their valuable assistance in planning, implementing, and performing the multi-modal imaging acquisitions. We thank the University of Pittsburgh Clinical and Translational Science Institute (CTSI) for their support in recruiting participants, as well as their support by the National Institutes of Health through Grant Number UL1TR001857. Portions of the PET data presented in this study have been previously reported in prior publications (Levine et al., 2022; Luna et al., 2020).

Data availability

Data will be made available on request.

References

Alexander GE, DeLong MR, Strick PL, 1986. Parallel organization of functionally segregated circuits linking basal ganglia and cortex. Annu. Rev. Neurosci. 9, 357–381. [PubMed: 3085570]

Attwell D, Laughlin SB, 2016. An energy budget for signaling in the grey matter of the brain. J. Cereb. Blood Flow Metab. doi:10.1097/00004647-200110000-00001.

Berke JD, 2018. What does dopamine mean? Nat. Neurosci. 21, 787. doi:10.1038/s41593-018-0152-y. [PubMed: 29760524]

Berns GS, McClure SM, Pagnoni G, Montague PR, 2001. Predictability modulates human brain response to reward. J. Neurosci. 21, 2793–2798. doi:10.1523/JNEUROSCI.21-08-02793.2001. [PubMed: 11306631]

Block AE, Dhanji H, Thompson-Tardif SF, Floresco SB, 2007. Thalamic-prefrontal cortical-ventral striatal circuitry mediates dissociable components of strategy set shifting. Cereb. Cortex 17, 1625–1636. doi:10.1093/cercor/bhl073. [PubMed: 16963518]

Burnham KP, Anderson DR, 2002. Model Selection and Multimodel Inference: A Practical Information-Theoretic Approach, 2nd ed. Springer-Verlag, New York doi:10.1007/b97636.

Chase HW, Frank MJ, Michael A, Bullmore ET, Sahakian BJ, Robbins TW, 2010. Approach and avoidance learning in patients with major depression and healthy controls: relation to anhedonia. Psychol. Med. 40, 433–440. doi:10.1017/S0033291709990468. [PubMed: 19607754]

Cox RW, 1996. AFNI: software for analysis and visualization of functional magnetic resonance neuroimages. Comput. Biomed. Res. 29, 162–173. doi:10.1006/cbmr.1996.0014. [PubMed: 8812068]

Cox RW, Chen G, Glen DR, Reynolds RC, Taylor PA, 2017. FMRI clustering in AFNI: false-positive rates redux. Brain Connect. 7, 152–171. doi:10.1089/brain.2016.0475. [PubMed: 28398812]

Daunizeau J, Adam V, Rigoux L, 2014. VBA: a probabilistic treatment of non-linear models for neurobiological and behavioural data. PLoS Comput. Biol 10. doi:10.1371/journal.pcbi.1003441.

Diaconescu AO, Menon M, Jensen J, Kapur S, McIntosh AR, 2010. Dopamine-induced changes in neural network patterns supporting aversive conditioning. Brain Res. 1313, 143–161. doi:10.1016/j.brainres.2009.11.064. [PubMed: 19961836]

Diederen KMJ, Spencer T, Vestergaard MD, Fletcher PC, Schultz W, 2016. Adaptive prediction error coding in the human midbrain and striatum facilitates behavioral adaptation and learning efficiency. Neuron 90, 1127–1138. doi:10.1016/j.neuron.2016.04.019. [PubMed: 27181060]

Dubol M, Trichard C, Leroy C, Sandu A−L, Rahim M, Granger B, Tzavara ET, Karila L, Martinot J−L, Artiges E, 2018. Dopamine transporter and reward anticipation in a dimensional perspective: a multimodal brain imaging study. Neuropsychopharmacol 43, 820–827. doi:10.1038/npp.2017.183.

Floresco SB, Ghods-Sharifi S, Vexelman C, Magyar O, 2006. Dissociable roles for the nucleus accumbens core and shell in regulating set shifting. J. Neurosci. 26, 2449–2457. doi:10.1523/JNEUROSCI.4431-05.2006. [PubMed: 16510723]

Floresco SB, Zhang Y, Enomoto T, 2009. Neural circuits subserving behavioral flexibility and their relevance to schizophrenia. Behav. Brain Res. 204, 396–409. doi:10.1016/j.bbr.2008.12.001. [PubMed: 19110006]

Gerfen CR, Surmeier DJ, 2011. Modulation of striatal projection systems by dopamine. Annu. Rev. Neurosci. 34, 441–466. doi:10.1146/annurev-neuro-061010-113641. [PubMed: 21469956]

Gershman SJ, Uchida N, 2019. Believing in dopamine. Nat. Rev. Neurosci. 20, 703–714. doi:10.1038/s41583-019-0220-7. [PubMed: 31570826]

Glimcher PW, 2011. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. Proc. Natl Acad. Sci. 108, 15647–15654. doi:10.1073/pnas.1014269108. [PubMed: 21389268]

Hakyemez HS, Dagher A, Smith SD, Zald DH, 2008. Striatal dopamine transmission in healthy humans during a passive monetary reward task. Neuroimage 39, 2058–2065. [PubMed: 18063390]

Hallquist MN, Hwang K, Luna B, 2013. The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures

functional connectivity. Neuroimage 82, 208–225. doi:10.1016/j.neuroimage.2013.05.116. [PubMed: 23747457]

Hamilton JP, Sacchet MD, Hjørnevik T, Chin FT, Shen B, Kämpe R, Park JH, Knutson BD, Williams LM, Borg N, Zaharchuk G, Camacho MC, Mackey S, Heilig M, Drevets WC, Glover GH, Gambhir SS, Gotlib IH, 2018. Striatal dopamine deficits predict reductions in striatal functional connectivity in major depression: a concurrent 11C-raclopride positron emission tomography and functional magnetic resonance imaging investigation. Transl. Psychiatry 8. doi:10.1038/s41398-018-0316-2.

Hartwell KJ, Johnson KA, Li X, Myrick H, LeMatty T, George MS, Brady KT, 2011. Neural correlates of craving and resisting craving for tobacco in nicotine dependent smokers. Addict. Biol. 16, 654–666. doi:10.1111/j.1369-1600.2011.00340.x. [PubMed: 21790899]

Holroyd CB, Yeung N, 2012. Motivation of extended behaviors by anterior cingulate cortex. Trends Cogn. Sci. 16, 122–128. doi:10.1016/j.tics.2011.12.008. [PubMed: 22226543]

Kaiser RH, Treadway MT, Wooten DW, Kumar P, Goer F, Murray L, Beltzer M, Pechtel P, Whitton A, Cohen AL, Alpert NM, El Fakhri G, Normandin MD, Pizzagalli DA, 2018. Frontostriatal and dopamine markers of individual differences in reinforcement learning: a multi-modal investigation. Cereb. Cortex 28, 4281–4290. doi:10.1093/cercor/bhx281. [PubMed: 29121332]

Kullmann S, Blum D, Jaghutriz BA, Gassenmaier C, Bender B, Häring H−U, Reischl G, Preissl H, la Fougère C, Fritsche A, Reimold M, Heni M, 2021. Central insulin modulates dopamine signaling in the human striatum. J. Clin. Endocrinol. Metab. 106, 2949–2961. doi:10.1210/clinem/dgab410. [PubMed: 34131733]

Larsen B, Olafsson V, Calabro F, Laymon C, Tervo-Clemmens B, Campbell E, Minhas D, Montez D, Price J, Luna B, 2020. Maturation of the human striatal dopamine system revealed by PET and quantitative MRI. Nat. Commun. 11, 846. doi:10.1038/s41467-020-14693-3. [PubMed: 32051403]

Le Heron C, Kolling N, Plant O, Kienast A, Janska R, Ang Y−S, Fallon S, Husain M, Apps MAJ, 2020. Dopamine modulates dynamic decision-making during foraging. J. Neurosci. 40, 5273–5282. doi:10.1523/JNEUROSCI.2586-19.2020. [PubMed: 32457071]

Levine MA, Mandeville JB, Calabro F, Izquierdo-Garcia D, Chonde DB, Chen KT, Hong I, Price JC, Luna B, Catana C, 2022. Assessment of motion and model bias on the detection of dopamine response to behavioral challenge. J. Cereb. Blood Flow Metab. doi:10.1177/0271678X221078616, 0271678X221078616.

Li N, Jasanoff A, 2020. Local and global consequences of reward-evoked striatal dopamine release. Nature 580, 239–244. doi:10.1038/s41586-020-2158-3. [PubMed: 32269346]

Logothetis NK, 2003. The underpinnings of the BOLD functional magnetic resonance imgaging signal. J. Neurosci. 23, 3963–3971. [PubMed: 12764080]

Luna B, Marek S, Larsen B, Tervo-Clemmens B, Chahal R, 2015. An integrative model of the maturation of cognitive control. Annu. Rev. Neurosci. 38, 151–170. doi:10.1146/annurev-neuro-071714-034054. [PubMed: 26154978]

Luna B, Parr A, Larsen B, Calabro F, Foran W, 2020. Specialization of the dopaminergic and frontostriatal systems through adolescence. Biol. Psychiatry 87, S33. doi:10.1016/j.biopsych.2020.02.107.

Luna B, Wright C, 2016. Adolescent brain development: implications for the juvenile criminal justice system. In: Heilbrun K, DeMatteo D, Goldstein NES (Eds.), APA Handbook of Psychology and Juvenile Justice. American Psychological Association, Washington, DC, US, pp. 91–116.

Mandeville JB, Sander CYM, Jenkins BG, Hooker JM, Catana C, Vanduffel W, Alpert NM, Rosen BR, Normandin MD, 2013. A receptor-based model for dopamine-induced fMRI signal. Neuroimage 75, 46–57. doi:10.1016/j.neuroimage.2013.02.036. [PubMed: 23466936]

Mohebi A, Pettibone JR, Hamid AA, Wong J−MT, Vinson LT, Patriarchi T, Tian L, Kennedy RT, Berke JD, 2019. Dissociable dopamine dynamics for learning and motivation. Nature 570, 65. doi:10.1038/s41586-019-1235-y. [PubMed: 31118513]

Niv Y, Daniel R, Geana A, Gershman SJ, Leong YC, Radulescu A, Wilson RC, 2015. Reinforcement learning in multidimensional environments relies on attention mechanisms. J. Neurosci. 35, 8145–8157. doi:10.1523/JNEUROSCI.2978-14.2015. [PubMed: 26019331]

Niv Y, Duff MO, Dayan P, 2005. Dopamine, uncertainty and TD learning. Behav. Brain Funct. 1, 6. doi:10.1186/1744-9081-1-6. [PubMed: 15953384]

Ott T, Nieder A, 2019. Dopamine and cognitive control in prefrontal cortex. Trends Cogn. Sci. 23, 213–234. doi:10.1016/j.tics.2018.12.006. [PubMed: 30711326]

Pappata S, Dehaene S, Poline JB, Gregoire MC, Jobert A, Delforge J, Frouin V, Bottlaender M, Dolle F, Di Giamberardino L, Syrota A, 2002. In vivo detection of striatal dopamine release during reward: a PET Study with [11C]Raclopride and a single dynamic scan approach. Neuroimage 16, 1015–1027. doi:10.1006/nimg.2002.1121. [PubMed: 12202089]

Parr AC, Calabro F, Larsen B, Tervo-Clemmens B, Elliot S, Foran W, Olafsson V, Luna B, 2021. Dopamine-related striatal neurophysiology is associated with specialization of frontostriatal reward circuitry through adolescence. Prog. Neurobiol. 201, 101997. doi:10.1016/j.pneurobio.2021.101997. [PubMed: 33667595]

Patel AX, Bullmore ET, 2016. A wavelet-based estimator of the degrees of freedom in denoised fMRI time series for probabilistic testing of functional connectivity and brain graphs. Neuroimage 142, 14–26. doi:10.1016/j.neuroimage.2015.04.052. [PubMed: 25944610]

Pawlak V, Kerr JND, 2008. Dopamine receptor activation is required for corticostriatal spike-timing-dependent plasticity. J. Neurosci. 28, 2435–2446. doi:10.1523/JNEUROSCI.4402-07.2008. [PubMed: 18322089]

Phillips RD, Walsh E, Zürcher NR, Lalush D, Kinard J, Tseng C−E, Cernasov P, Kan D, Cummings K, Kelley L, Campbell D, Dillon D, Pizzagalli DA, Izquierdo-Garcia D, Hooker J, Smoski M, Dichter GS 2022. A simultaneous [11C]Raclopride positron emission tomography and functional magnetic resonance imaging investigation of striatal dopamine binding in anhedonia. doi:10.1101/2022.07.21.22277878

Piray P, Dezfouli A, Heskes T, Frank MJ, Daw ND, 2019. Hierarchical Bayesian inference for concurrent model fitting and comparison for group studies. PLoS Comput. Biol. 15, e1007043. doi:10.1371/journal.pcbi.1007043. [PubMed: 31211783]

Reading PJ, Dunnett SB, 1991. The effects of excitotoxic lesions of the nucleus accumbens on a matching to position task. Behav. Brain Res. 46, 17–29. doi:10.1016/s0166-4328(05)80093-2. [PubMed: 1786111]

Rescorla R, Wagner AR 1972. 3 A Theory of Pavlovian Conditioning: variations in the Effectiveness of Reinforcement and Nonreinforcement.

Rodriguez PF, Aron AR, Poldrack RA, 2006. Ventral–striatal/nucleus–accumbens sensitivity to prediction errors during classification learning. Hum. Brain Mapp. 27, 306–313. doi:10.1002/hbm.20186. [PubMed: 16092133]

Roesch MR, Singh T, Brown PL, Mullins SE, Schoenbaum G, 2009. Ventral striatal neurons encode the value of the chosen action in rats deciding between differently delayed or sized rewards. J. Neurosci. 29, 13365–13376. doi:10.1523/JNEUROSCI.2572-09.2009. [PubMed: 19846724]

Schönberg T, Daw ND, Joel D, O'Doherty JP, 2007. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. J. Neurosci. 27, 12860–12867. doi:10.1523/JNEUROSCI.2496-07.2007. [PubMed: 18032658]

Schultz W, 2016. Dopamine reward prediction error coding. Dialogues Clin. Neurosci. 18, 23–32. [PubMed: 27069377]

Schultz W, 1998. Predictive reward signal of dopamine neurons. J. Neurophysiol. 80, 1–27. [PubMed: 9658025]

Schultz W, 1986. Responses of midbrain dopamine neurons to behavioral trigger stimuli in the monkey. J. Neurophysiol. 56, 1439–1461. doi:10.1152/jn.1986.56.5.1439. [PubMed: 3794777]

Schultz W, Apicella P, Ljungberg T, 1993. Responses of monkey dopamine neurons to reward and conditioned stimuli during successive steps of learning a delayed response task. J. Neurosci. 13, 900–913. [PubMed: 8441015]

Schulz E, Wu CM, Ruggeri A, Meder B, 2019. Searching for rewards like a child means less generalization and more directed exploration. Psychol. Sci. 30, 1561–1572. doi:10.1177/0956797619863663. [PubMed: 31652093]

Shen W, Flajolet M, Greengard P, Surmeier DJ, 2008. Dichotomous dopaminergic control of striatal synaptic plasticity. Science 321, 848–851. doi:10.1126/science.1160575. [PubMed: 18687967]

Shulman EP, Smith AR, Silva K, Icenogle G, Duell N, Chein J, Steinberg L, 2016. The dual systems model: review, reappraisal, and reaffirmation. Dev. Cogn. Neurosci. 17, 103–117. doi:10.1016/j.dcn.2015.12.010. [PubMed: 26774291]

Skvortsova V, Degos B, Welter M−L, Vidailhet M, Pessiglione M, 2017. A selective role for dopamine in learning to maximize reward but not to minimize effort: evidence from patients with Parkinson's disease. J. Neurosci. 37, 6087–6097. doi:10.1523/JNEUROSCI.2081-16.2017. [PubMed: 28539420]

Smith S, Brady J, 1997. SUSAN - a new approach to low level image processing. Int. J. Comput. Vis. 23, 45–78.

Spear LP, 2000. The adolescent brain and age-related behavioral manifestations. Neurosci. Biobehav. Rev. 24, 417–463. [PubMed: 10817843]

Sutton RS, 1999. Reinforcement learning: past, present and future. In: McKay B, Yao X, Newton CS, Kim J-H, Furuhashi T (Eds.), Simulated Evolution and Learning, Lecture Notes in Computer Science. Springer, Berlin Heidelberg, pp. 195–197.

Taha SA, Fields HL, 2006. Inhibitions of nucleus accumbens neurons encode a gating signal for reward-directed behavior. J. Neurosci. 26, 217–222. doi:10.1523/JNEUROSCI.3227-05.2006. [PubMed: 16399690]

Tziortzi AC, Searle GE, Tzimopoulou S, Salinas C, Beaver JD, Jenkinson M, Laruelle M, Rabiner EA, Gunn RN, 2011. Imaging dopamine receptors in humans with [11C]-(+)-PHNO: dissection of D3 signal and anatomy. Neuroimage 54, 264–277. doi:10.1016/j.neuroimage.2010.06.044. [PubMed: 20600980]

Urban NBL, Slifstein M, Meda S, Xu X, Ayoub R, Medina O, Pearlson GD, Krystal JH, Abi-Dargham A, 2012. Imaging human reward processing with positron emission tomography and functional magnetic resonance imaging. Psychopharmacology 221, 67–77. doi:10.1007/s00213-011-2543-6. [PubMed: 22052081]

Wang S, Kim S, Cosgrove KP, Morris ED, 2017. A framework for designing dynamic lp-ntPET studies to maximize the sensitivity to transient neurotransmitter responses to drugs: application to dopamine and smoking. Neuroimage 146, 701–714. doi:10.1016/j.neuroimage.2016.10.019. [PubMed: 27743899]

Wilson RC, Niv Y, 2015. Is model fitting necessary for model-based fMRI? PLoS Comput. Biol. 11, e1004237. doi:10.1371/journal.pcbi.1004237. [PubMed: 26086934]

Yagishita S, Hayashi-Takagi A, Ellis-Davies GCR, Urakubo H, Ishii S, Kasai H, 2014. A critical time window for dopamine actions on the structural plasticity of dendritic spines. Science 345, 1616–1620. doi:10.1126/science.1255514. [PubMed: 25258080]

Zald DH, Boileau I, El-Dearedy W, Gunn R, McGlone F, Dichter GS, Dagher A, 2004. Dopamine transmission in the human striatum during monetary reward tasks. J. Neurosci. 24, 4105–4112. [PubMed: 15115805]

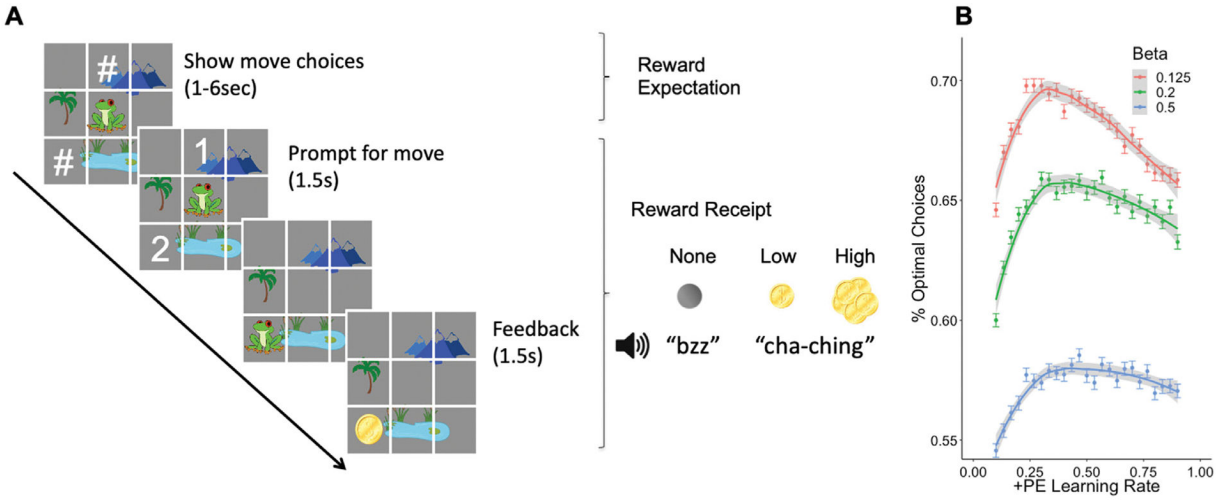

Fig. 1.

(A) Task schematic. On each trial, subjects were presented two map locations as possible movement choices, indicated by hash marks ("#"). Following a randomized delay, these symbols were replaced by the numbers "1″ and "2″ such that the subject could make a button press response. Once they responded, the map updated to show their move, and feedback was given as both visual (black circle, single coin, pile of coins) and auditory (flat tone, "cha-ching" sound) feedback to indicate whether a reward was received, and whether it was small (low reward) or large (high reward). (B) Simulated reinforcement learning model data for combinations of the positive prediction error learning rate and softmax temperature parameter (Beta) in successfully making optimal movement choices.

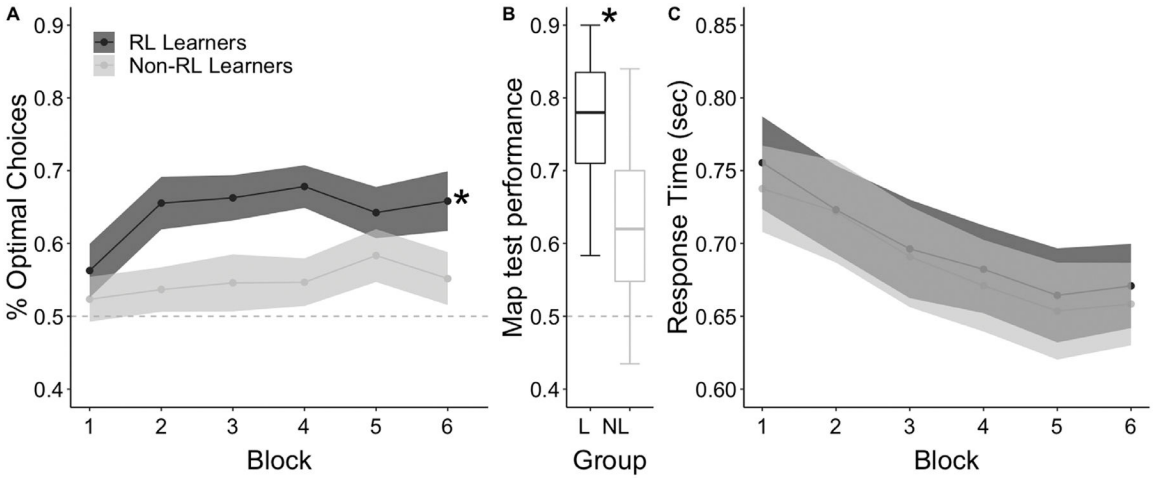

Fig. 2.

Task performance by learner group. Performance as quantified by (A) proportion of optimal trials by blocks, (B) distribution of post-task assessments for RL learners (L) and non-RL learners (NL), and (C) response time by block, split by learner category. Dashed horizontal line indicates chance performance.

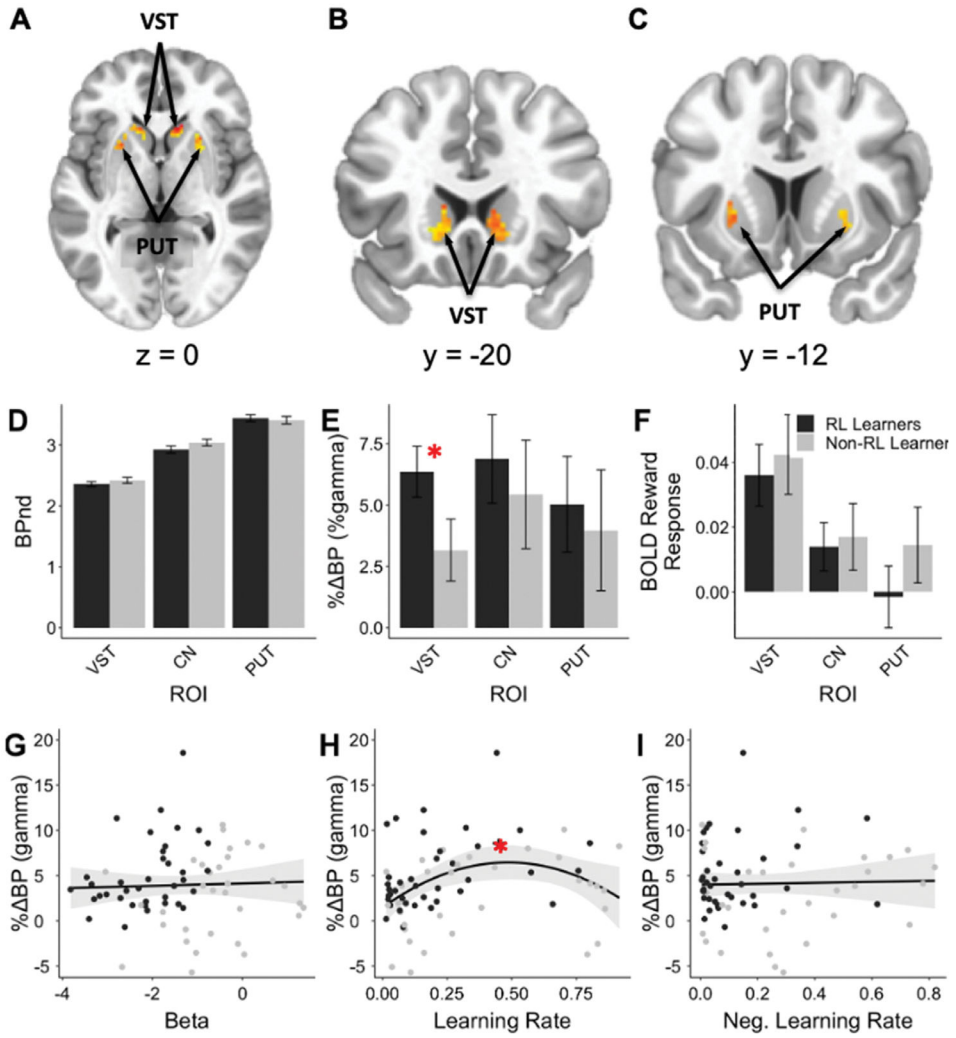

Fig. 3.

(A-C) Cluster corrected maps of significant task-related change in RAC BP. Slices are shown at (A) $z = 0$ (axial) (B) $y = -20$ (coronal) (C) $y = -12$ (coronal). Four clusters of activation were observed (left slice), including a bilateral set focused primarily on the ventral striatum (VS, panel B), and a bilateral set focused on the precommisural dorsal putamen (PUT, panel C). (D-F) Comparisons of RL learners and non-RL learners across regions containing task-dependent DA responses (ventral striatum, VST; precommisural caudate nucleus, CN; precommisural dorsal putamen, PUT), for (D) baseline (pre-task) RAC BPnd, (E) task-dependent change in BPnd, and (F) aggregate BOLD reward response (contrast of rewarded vs. non-reward trial outcomes). (G-I) Association of task-related DA release in the VST and RL model parameters, including (G) softmax temperature parameter, (H) learning rate for positive PE trials, and (I) learning rate for negative PE trials.

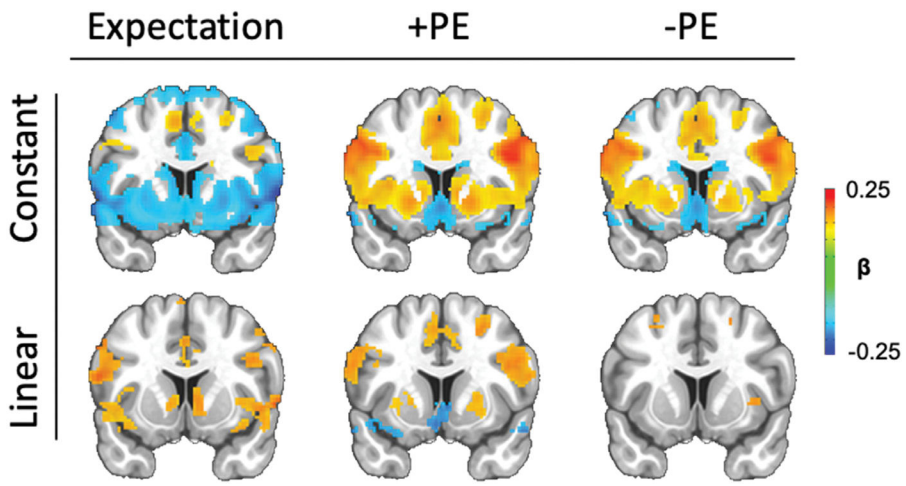

Fig. 4.

Striatal activation for each of the six fMRI contrasts, including expectation (left), positive prediction error (middle) and negative prediction error (right). For each, both mean activation at each trial epoch ('constant', top row) and activation proportional to reward expectation/PE respectively ('linear', bottom row) are shown. Activation maps are cluster corrected and additionally Bonferroni corrected for six comparisons.

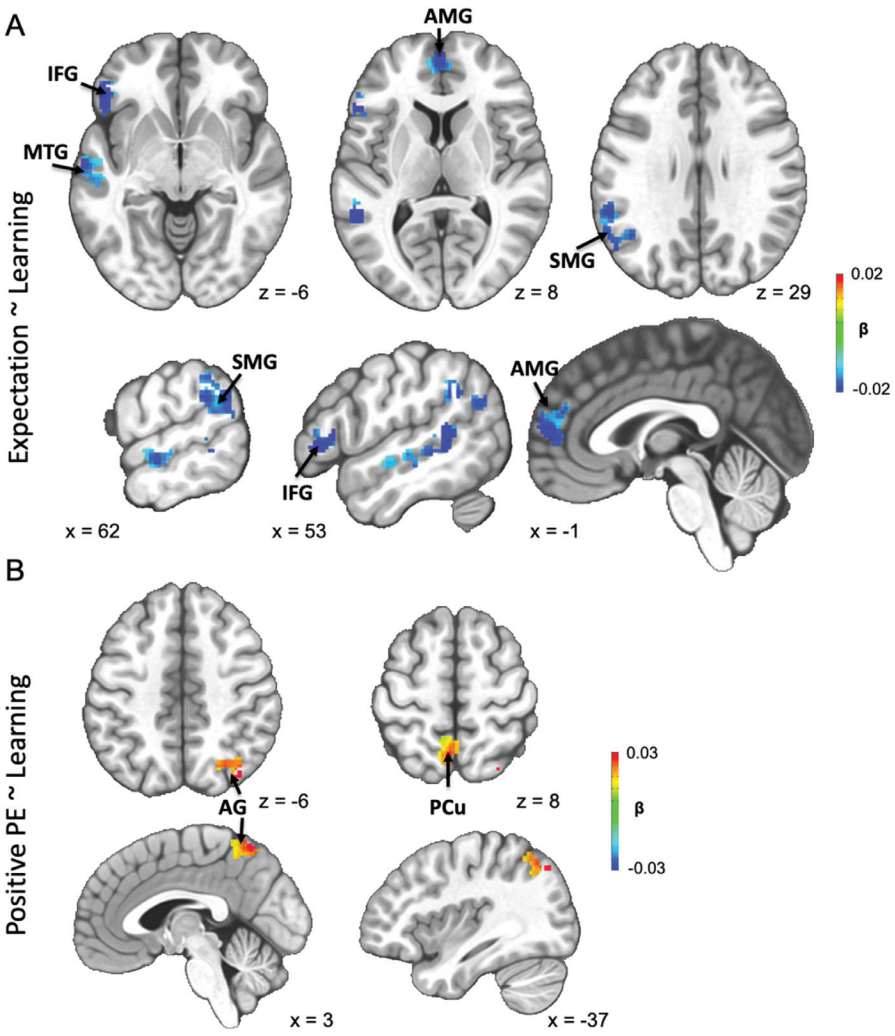

Fig. 5.

Cluster corrected correlation of task fMRI activation with reinforcement learning performance. Axial (top) and sagittal (bottom) views of significant clusters, for mean (A) expectation and (B) positive prediction error contrasts. Activation clusters were identified in the supramarginal gyrus (SMG), anterior medial gyrus (AMG), middle temporal gyrus (MTG), inferior frontal gyrus (IFG), angular gyrus (AG), and precuneus (PCu) (see Table 1).

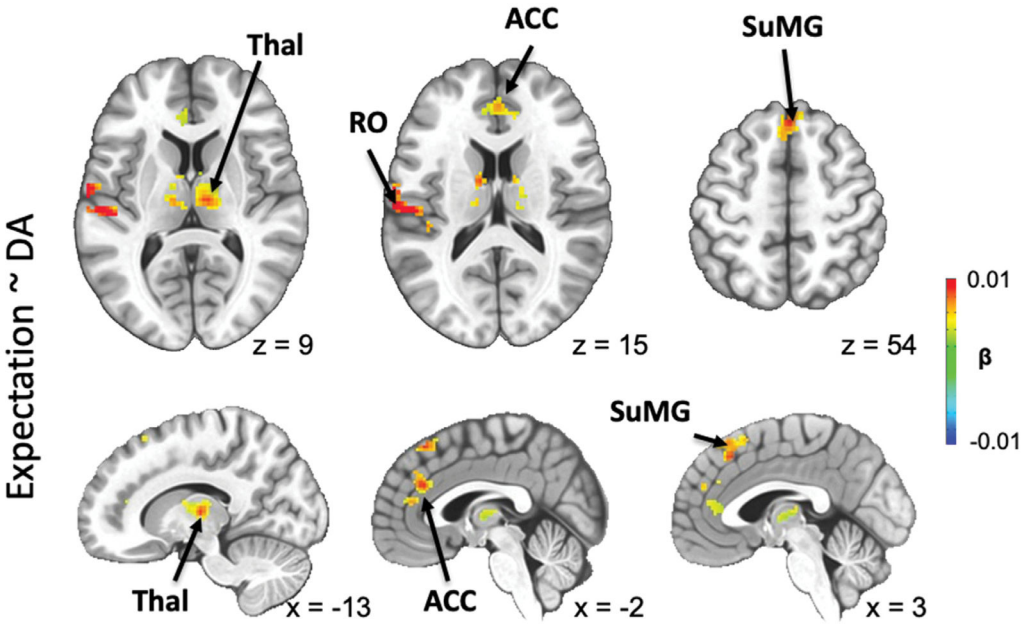

Fig. 6.

Correlation of expectation-related BOLD activation with VST DA release. Axial (top) and sagittal (bottom) views of significant clusters. Significant clusters of activation were found in the thalamus (Thal), anterior cingulate cortex (ACC), superior medial gyrus (SuMG), and Rolandic operculum (RO) (see Table 2).

Table 1

Clusters of activation which scaled with learning performance. For each of the six fMRI contrasts, significant clusters in which BOLD activation was correlated with individual AAIC (RL model support relative to a null model). Cluster coordinates in MNI space are given for the center of mass, and cluster size (k) reflects number of voxels (Bonferroni corrected cluster size threshold given for each contrast). N/A indicates contrasts which had no significant clusters. Excluded L/R indicates a bilateral midline region.

	Area	L/R	MNI Coordinates			*k*	Peak β	Peak Z
			x	*y*	*z*			
Expectation - Constant Term *(k > 165)*								
Supramarginal Gyrus (SMG)		L	−58	−39	28	284	−0.023	−3.04
Anterior Medial Gyrus (AMG)		L	−1	54	7	272	−0.027	−3.52
Middle Temporal Gyrus (MTG)		L	−53	−44	8	269	−0.028	−3.31
Inferior Frontal Gyrus (IFG)		L	−51	26	−6	172	−0.034	−2.98
Expectation - Linear Term *(k>60)*								
N/A								
Negative PE - Constant Term *(k>143)*								
N/A								
Negative PE - Linear Term *(k>52)*								
N/A								
Positive PE - Constant Term *(k> 150)*								
Angular Gyrus (AG)		R	36	−70	47	170	0.051	3.35
Precuneus (PCu)			−3	−56	61	151	0.037	2.17
Positive PE - Linear Term *(k>101)*								
N/A								

Neuroimage. Author manuscript; available in PMC 2023 March 03.

Table 2

Clusters of activation which scaled with VST DA release. For each of the six fMRI contrasts, significant clusters in which BOLD activation was correlated with individual VST%ΔBP (task-dependent change in DA binding). Cluster coordinates in MNI space are given for the center of mass, and cluster size (*k*) reflects number of voxels (Bonferroni corrected cluster size threshold given for each contrast). N/A indicates contrasts which had no significant clusters. Excluded L/R indicates a bilateral midline region.

	Area	L/R	MNI Coordinates			*k*	Peak *β*	Peak Z
			x	*y*	*z*			
Expectation - Constant Term *(k>165)*								
Rolandic Operculum (RO)		L	−57	−20	18	328	0.017	3.53
Thalamus (Thal)			2	−11	9	314	0.009	3.38
Anterior Cingulate Cortex (ACC)			−1	38	18	190	0.009	4.08
Superior Medial Gyrus (SuMG)		L	1	30	53	172	0.009	3.98
Expectation - Linear Term *(k>60)*								
N/A								
Negative PE - Constant Term *(k>143)*								
N/A								
Negative PE - Linear Term *(k>52)*								
N/A								
Positive PE - Constant Term *(k> 150)*								
N/A								
Positive PE - Linear Term *(k>52)*								
N/A								