

Reward Prediction Error Signaling as a Neurocomputational Correlate of Co-Occurring Anxiety and Depression in Youth

Marta Korom, PhD¹, Diana Burk, PhD², Tahjane Givens, BA¹, Daniel S. Pine, MD¹, Melissa A. Brotman, PhD¹, Simone P. Haller, PhD¹, Bruno B. Averbeck, PhD², Ellen Leibenluft, PhD¹, & Katharina Kircanski, PhD^{1,3}

¹ Emotion and Development Branch, National Institute of Mental Health, Bethesda, MD

² Section on Learning and Decision Making, National Institute of Mental Health, Bethesda, MD

³ Department of Psychology, University of Southern California, Los Angeles, CA

This work was supported by the National Institute of Mental Health Intramural Research Program projects ZIA-MH002781(DSP), ZIA-MH002928(BA), and 1ZIAMH002969(MAB). The contributions of the NIH author(s) are considered Works of the United States Government. The findings and conclusions presented in this paper are those of the author(s) and do not necessarily reflect the views of the NIH or the U.S. Department of Health and Human Services.

The research was performed with permission from the National Institutes of Health.

Disclosure: Korom, Burk, Givens, Pine, Brotman, Haller, Averbeck, Leibenluft, and Kircanski reported no biomedical financial interests or potential conflicts of interest.

The authors would like to thank all the participants and their families for continuing to dedicate their time to this research. The authors also thank the many research assistants who helped with data collection over the years.

Correspondence to: Marta Korom, PhD, Section on Development and Affective Neuroscience, National Institute of Mental Health. Email: marta.korom@nih.gov.

Highlights

1. Co-occurring anxiety and depression symptoms (CADS) linked to altered reward prediction error (RPE) signaling.
2. Anxiety and depression show distinct RPE-related associations in frontal and cingulate cortex
3. CADS severity predicts interactive effects on RPE signals in ventral striatum and amygdala
4. Results support interactive, not additive, CADS effects on reward signaling in signals

Abstract

Co-occurring depressive and anxiety symptoms (CADS) is a chronic and impairing condition that affects 1 in 6 youth. This underscores the need to identify neurocognitive mechanisms that characterize CADS. One candidate mechanism is altered reward prediction error (RPE), a reinforcement-learning signal indexing the discrepancy between expected and received outcomes. Prior work has largely focused on RPE signaling in youth with anxiety or depression diagnoses, with limited attention to co-occurring dimensional symptoms. To address this gap, we used computational modeling and neuroimaging to examine RPE-modulated brain activation in relation to CADS. A transdiagnostic sample of 88 youths (Mage=14.43 years, range=8.81–18.57; 49% female) completed a child-friendly two-armed bandit reinforcement-learning task during fMRI and completed anxiety and depressive symptom measures. Planned ROI-based regressions and exploratory whole-brain analyses tested the main and interactive effects of symptoms on RPE-modulated activation in reward circuitry. ROI analyses showed no significant main effects of depressive symptoms, but parent-reported anxiety was negatively associated with right amygdala RPE-modulated activation. Whole-brain analyses identified three clusters with parent-reported symptoms: anxiety showed a positive association with RPE-modulated activation in left superior frontal cortex, whereas depression was negatively associated with two clusters in left superior frontal and middle cingulate cortices. ROI-based interactions indicated that elevated child- and parent-reported CADS predicted RPE-modulated activation in right ventral striatum and amygdala, respectively, such that positive associations between anxiety and RPE-modulated activation weakened as depressive symptoms increased. Findings suggest that elevated CADS exert interactive, rather than additive, effects on RPE signaling.

Keywords: reward prediction error, anxiety, depression, fMRI, reinforcement learning, striatum, amygdala

Introduction

Co-occurring anxiety and depression symptoms (CADS)—operationalized as concurrently elevated dimensional symptoms of anxiety and depression—are highly prevalent in youth and have a distinct risk profile that confers greater psychosocial impairment, service use, suicide risk, and chronicity than either symptom domain alone (1,2). Although transdiagnostic interventions target overlapping regulatory vulnerabilities across co-occurring symptoms, meta-analytic work suggests that benefits are often short-lived (3). Despite this burden and the limited long-term effectiveness of interventions, the neurocognitive mechanisms underlying CADS remain poorly understood.

A key question in examining mechanisms that sustain CADS is whether anxiety and depressive symptoms exert additive or interactive effects on cognitive correlates. For clinical features, diagnostic approaches support an additive model (1,4,5), wherein comorbid anxiety and depression diagnoses are associated with greater impairment, and single-diagnostic presentations that disregard subthreshold symptoms show intermediate impairment profiles. In contrast, dimensional and Research Domain Criteria (RDoC)—aligned perspectives emphasize a partially dissociable, interactive model of neurocognitive correlates (2,6–11), wherein the severity of anxiety and depressive symptoms synergistically interact as they both increase to shape cognitive correlates in CADS. Consistent with the RDoC framework, this interactive model has received translational support and offers greater mechanistic precision than diagnostic approaches. This modeling distinction is particularly relevant for neurocognitive mechanisms implicated in CADS, such as reinforcement learning (RL). In computational psychiatry, RL models describe how the brain updates expectations after reward or non-reward feedback, using parameters like learning rates (how quickly one updates expectations from expected and unexpected events). Meta-analytic and computational work indicate that, both neurally and behaviorally, anxiety and depression are associated with heightened punishment sensitivity or learning (how strongly one detects, reacts to, and is motivated by rewards) (12–15), whereas blunted reward sensitivity is more specific to depressive and anhedonic dimensions (16). Importantly, nearly all prior work has modeled anxiety and depressive symptoms either categorically or as isolated single dimensions, leaving anxiety-by-depression interactions

on reward signaling untested in fMRI. Guided by this gap in the literature, this study tests whether the interaction between anxiety and depressive symptoms predicts reward prediction error (RPE) signaling within the reward network during RL in a transdiagnostic sample of youth across varying levels of CADS severity.

Reward prediction errors (RPEs) provide a mechanistic entry point into RL processes. RPEs are a core computational signal of reward learning encoded within cortico-striatal-limbic circuits, including the ventral striatum and amygdala, and are robustly measurable during RL (17–20). RPEs index the discrepancy between expected and received outcomes, with positive RPEs reflecting better-than-expected outcomes, negative RPEs indicating worse-than-expected outcomes, and zero RPEs signaling outcome confirmation (21–24). RPEs guide approach and avoidance behavior by reinforcing rewarding and discouraging non-rewarding behaviors based on prior outcomes. In depression, fMRI studies report blunted RPE-related BOLD activity in the striatum and altered coupling with midbrain and prefrontal regions during reward learning (12,15,25) (but see (26) for null effects). Computational work further suggests dimension-specific effects: anxiety relates to tighter coupling of negative prediction errors to punishment-related signals and thus more exaggerated avoidance learning, whereas depression relates to reduced reward signaling, indexed by blunted reward network activation, without impaired reward learning per se (27,28).

Computational models of reinforcement learning can estimate each participant's subject-specific RPEs on every trial from their behavioral response to a series of choices and feedback, and these RPEs can then be used as trial-specific predictors of BOLD signal magnitude (18,19,29). This computational approach quantifies internal learning signals that cannot be directly observed from behavior alone. By fitting models to trial-by-trial choices, we derive RPEs and relate brain activation to how outcomes differ from expectations (better, the same, or worse than expected), rather than just to outcome type (win vs. no-win). Applying these methods in youth with CADS allows us to test whether anxiety and depressive symptoms relate to how youth learn from feedback, and whether symptoms have distinct or interacting effects on RPE-related activation in core reward circuits.

In the present study, we completed planned ROI-based analyses in the bilateral ventral striatum and amygdala, followed by exploratory whole-brain analyses during probabilistic RL. We selected the ventral striatum and amygdala as a priori ROIs because they are core hubs of RPE signaling: the ventral striatum encodes signed reward RPEs tied to reinforcement learning (30), and the amygdala preferentially encodes aversive/non-reward prediction errors and affective learning (31), rendering both regions particularly relevant for investigating RL mechanisms in pediatric mood and anxiety psychopathology. We tested whether elevated parent- and child-reported CADS are associated with the greatest attenuation—or qualitatively altered patterns—of RPE-modulated BOLD. The analyses were exploratory, yet we hypothesized that the interactive, rather than the additive, model would yield significant effects. Specifically, we hypothesized that CADS would interact to shape blunted RPE-modulated brain activation in the bilateral striatum and amygdala, and that this blunting would not emerge when only the main effects are examined.

Of note, the current report presents a distinct analysis of data that were also included in a prior publication. That prior work focused on pediatric irritability rather than anxiety or depressive symptoms and examined RL both before and after a frustrating task that successfully induced frustration in the scanner. In that study, irritability-related RPE-modulation effects largely emerged during the post-frustration runs ($p = .01$), with weak to no associations at baseline (no significant association with the right ventral striatum and a positive association with the left ventral striatum; $p = .042$). The current report focuses only on RL at baseline (i.e., before frustration induction), because frustration was not conceptualized as part of the pathogenesis of CADS as it was for irritability. To account for potential confounding by irritability, we conducted sensitivity analyses where we included irritability as a covariate in our ROI models that examined the CADS interaction effects.

Methods

Participants and Procedures

A sample of 106 participants recruited from the community (51 females; age 8.55–18.57 years, $M=14.10$, $SD=2.63$) completed a reinforcement learning (RL) neuroimaging task at the National Institute of Mental Health (NIMH). Participants were recruited via community advertisements, postcard mailings, and clinician referrals. A total of 18 participants did not meet data quality thresholds. The final sample included youth with primary attention deficit hyperactivity disorder ($n=24$), major depressive disorder ($n=20$), disruptive mood dysregulation disorder ($n=18$), anxiety disorders (generalized, social, separation anxiety disorder; $n=12$), and no psychiatric diagnosis ($n=14$), demonstrating variability in anxious and depressive symptoms both transdiagnostically and dimensionally. Exclusion criteria were diagnosis of pervasive developmental, psychotic, neurological, or current posttraumatic stress disorders, recent substance use (past 3 months), cardinal bipolar symptoms, an estimated $IQ<70$ (matrix reasoning and vocabulary subtests of the Wechsler Abbreviated Scale of Intelligence assessed by a trained research assistant), or unstable medical conditions. Parents provided written informed consent, and the youth provided assent. Procedures were approved by the NIMH Institutional Review Board, and data collection took place between 2018 and 2025. Demographics are summarized in Table 1.

Parent- and child-report clinical measures were completed within three months of imaging. Before the imaging session, participants completed training on each fMRI task phase (see “Carnival Paradigm” below and the Supplementary section 1 for the full description of the training trials and the full task). Participants were debriefed after the imaging session.

Measures

Screen for Child Anxiety Related Emotional Disorders (SCARED): The SCARED is a dimensional youth- and parent-report measure used to evaluate anxiety-related symptoms among individuals ages 8–18 (Birmaher et al., 1997). The instrument includes 41 items capturing symptom frequency across multiple anxiety domains, such as generalized anxiety, separation-related concerns,

social anxiety, and specific fears in various situations. Participants and the accompanying parent are asked to indicate how often each statement applies to the participant using a 3-point response scale (0 = Not true or rarely true, 1 = Somewhat or sometimes true, 2 = Very or often true). Total scores are calculated by summing item responses, with higher values reflecting greater anxiety symptomatology. In the present sample, the scale demonstrated excellent internal consistency (child report: Cronbach's $\alpha = .956$; parent report: Cronbach's $\alpha = .944$).

Mood and Feelings Questionnaire Short Form: The MFQ Short Form is a self- and parent-report measure used to assess depressive symptoms in children and adolescents (32). The questionnaire comprises 13 descriptive statements that capture the presence of core features of depression, including low mood, anhedonia, fatigue, and negative self-evaluation over the past two weeks. Participants and parents indicate how well each statement describes the participant's feelings recently using a 3-point Likert Scale (0 = Not true, 1 = Sometimes true, 2 = True). Item responses are summed to generate a total score, with higher values reflecting greater symptom severity. The MFQ has demonstrated excellent internal consistency in the current sample (child report: Cronbach's $\alpha = .926$; parent report: Cronbach's $\alpha = .934$).

Additional self and parent-reported measures: We also administered the Affective Reactivity Index (ARI). ARI is a 6-item measure of temper outbursts and irritable mood across the previous six months and produces a total score ranging from 0 to 12. In this sample, the ARI demonstrated excellent internal consistency (Cronbach's $\alpha = 0.9$).

Table 1
Demographic information

Variable	Value, N (%) or M (SD)
Sex	Female, n = 43 (48.86%) Male, n = 45 (51.14%)
Age	14.43 (2.52) years
Race	American Indian or Alaskan Native = 2 (2.27%) Asian = 5 (5.68%) Black or African American = 8 (9.09%) Multiple Races = 16 (18.18%) White or Caucasian = 54 (61.36%) Unknown = 3 (3.41%)

Child-reported measures:

SCARED ¹	20.023 (15.824)
MFQ ²	6 (6.118)
ARI ³	2.86 (2.98)

Parent-reported measures:

SCARED ¹	16.159 (13.057)
MFQ ²	4.591 (5.514)
ARI ³	3.432 (3.32)
IQ (n = 70)	115.843 (14.66)
Motion in the scanner ⁴	0.12 (0.07)

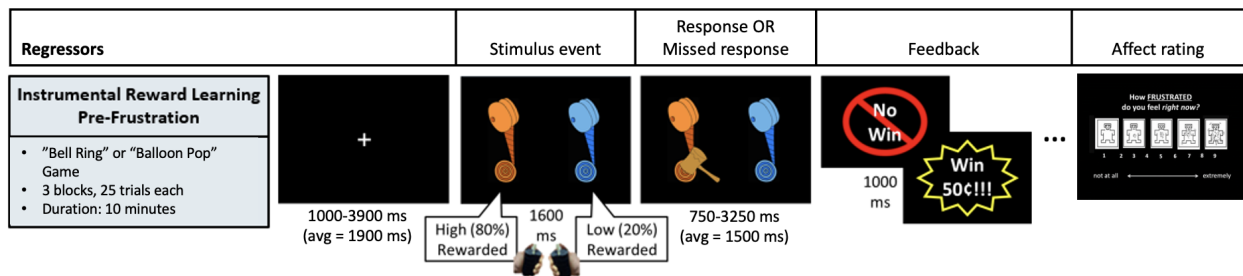
¹Screen for Child Anxiety Related Emotional Disorders; ²Mood and Feelings Questionnaire; ³Affective Reactivity Index; ⁴Calculated as the mean Euclidean distance of framewise volume shift after censoring.

fMRI task

The Carnival paradigm is a two-armed bandit reinforcement learning task that assesses RPE encoding before, during, and after frustration. For the present study, we only examine the pre-frustration runs, and we refer to them as ‘*baseline*’ runs. Our rationale for removing the frustration manipulation and post-frustration runs for this study was to isolate the computational properties of reinforcement learning at baseline, rather than the modulation by transient affective states. The baseline phase included three runs (“Bell Ring” or “Balloon Pop” visuals, counterbalanced; 25 trials run = 75 trials; duration: 10 minutes), with an 80:20 reward:nonreward ratio for the better option and 20:80 for the worse. Incorrect trials always resulted in non-reward. Participants used grip force dynamometers in each hand to indicate responses. Following the baseline phase, participants rated their frustration, excitement, and unhappiness on a nine-point Likert scale, anchored by five animated figures to reflect increasing emotional intensity. See Figure 1 for a depiction of the baseline run of the Carnival Task.

Figure 1

Overview of the baseline runs of the Carnival instrumental learning task with the regressors displayed at the top.



Computational Modeling and Behavioral Data Analysis

Behavioral data analysis

Reinforcement Learning Models. To characterize trial-by-trial changes in expected value, we implemented four reinforcement learning (RL) models, each derived from the Hull/Rescorla-Wagner framework. In all models, value updating was governed by a reward prediction error (RPE, δ), defined as the difference between the outcome and the expected value:

$$RPE: \delta(t) = r(t) - v_i(t)$$

where $r(t)$ is the reward received on trial t , and $v_i(t)$ is the expected value for the chosen option i . The four models included:

1. *Vanilla RL (RW)*. A single positive learning rate α updated value estimates based on $\delta(t)$:

$$v_i(t + 1) = v_i(t) + \alpha * \delta(t)$$

where i is the chosen option on trial t . Try Again (no-response) trials were ignored.

2. *Two Learning Rates (2LR)*. Separate learning rates for positive and negative prediction errors (α_{pos} , α_{neg}):

$$v_i(t + 1) = v_i(t) + \alpha_{pos} * \delta(t) \quad \text{if } R = 1$$

$$v_i(t + 1) = v_i(t) + \alpha_{neg} * \delta(t) \quad \text{if } R = 0$$

Try Again (missing response) trials were ignored.

3. *Try Again Trial Gamma (TA γ)*. “Try Again” trials are missed-response trials. A trial-dependent discounting factor γ scaled the value updating term on Try Again trials:

$$v_i(t_{TA}) = \gamma * v_i(t_{TA})$$

4. *TA γ + 2LR*. Combined models (2) and (3):

$$v_i(t_{TA}) = \gamma * v_i(t_{TA})$$

$$v_i(t + 1) = v_i(t) + \alpha_{pos} * \delta(t) \quad \text{if } R = 1$$

$$v_i(t + 1) = v_i(t) + \alpha_{neg} * \delta(t) \quad \text{if } R = 0$$

To model the choice probability, we used the softmax equation:

$$P(a = i) = \frac{\exp(\beta v_i)}{\sum_{j=1:k} \exp(\beta v_j)}$$

where v_j were the values of the unchosen option were, and β was a free parameter for choice consistency.

Using model comparison (see Model Comparison), model 4 was determined to be the best model to fit the data across subjects.

Model-based RW-RL model parameter estimates were as follows: positive learning rate (α_{pos}): M = 0.39, S.E.M = 0.022; negative learning rate (α_{pos}): M = 0.19, S.E.M = 0.016; Softmax beta (β): M = 8.90, S.E.M = 0.374; TA gamma (γ): M = 0.88, S.E.M = 0.0006.

Parameter Inference. Parameters were estimated using hierarchical Bayesian inference implemented in Julia with BAT.jl. For each model, posterior distributions over parameters were obtained using adaptive Markov Chain Monte Carlo (MCMC), specifically the Metropolis-Hastings sampling algorithm. We defined population-level priors over each parameter and estimated person-level values conditioned on these priors. Prior distributions were given a parameter for the mean

($\mu_{\beta prior} = U(0.05, 30)$, $\mu_{other prior} = N(0, 1)$). and had fixed standard deviations

($\sigma_{\beta prior} = 3$, $\sigma_{other prior} = 2$). Parameters with constrained domains were defined on the real line and transformed to the [0,1] interval within the likelihood function using an approximation to the inverse probit function:

$$iprobit(x) = \frac{1}{1 + \exp(-1.5976x - 0.07056x^3)}$$

Sampling was carried out with four independent chains using an initial adaptation period followed by the collection of posterior samples.

Model Comparison. To determine the best-fitting model, we approximated the log model evidence at the group level using a Gaussian Laplace approximation (49). Let $p(\mathbf{D}, \boldsymbol{\theta}) = p(\boldsymbol{\theta})p(\boldsymbol{\theta})$ denote the joint likelihood and prior. We first identified the maximum a posteriori (MAP) estimate $\hat{\boldsymbol{\theta}}$ using density maximization over posterior samples. The log model evidence was then computed as:

$$\log Z = \log p(\mathbf{D}|\hat{\boldsymbol{\theta}}) + \log p(\hat{\boldsymbol{\theta}}) + \frac{1}{2}(d \log(2\pi) + \log \log |\mathbf{Q}|)$$

where d is the number of free parameters, and $|\mathbf{Q}|$ is the determinant of the posterior covariance matrix. We then chose the model with the highest evidence value.

Neuroimaging Data Analysis

Imaging data acquisition

Imaging data were collected on a 3-Tesla GE DISCOVERY MR750 MRI scanner equipped with a 32-channel head coil. A sagittal localizer scan was followed by an automated higher-order shim to adjust the magnetic field and minimize signal dropout, with particular attention to the prefrontal cortices. Seven functional runs were obtained (3 baseline runs included in the present study, 1 frustration induction run not included, and 3 post-frustration runs not included), using 40 interleaved axial slices that yielded whole-brain coverage except for the cerebellum. For the pre-frustration runs, 73 volumes were acquired. Functional scans used a 96×96 matrix and echo-planar single-shot gradient-echo T2*-weighted imaging with the following parameters: repetition time TR = 2300 ms; echo time TE = 27 ms; flip angle FA = 70°; field of view FOV = 240×240 mm, producing 2.5 mm isotropic voxels. The first four images from each run were discarded to allow longitudinal magnetization to reach equilibrium. For co-registration with the functional images, a high-resolution structural MPAGE scan was also obtained with these parameters: 1 mm interleaved sagittal slices; repetition time TR = 7.7 ms; echo time TE = 3.48 ms; flip angle FA = 7°; field of view FOV = 176×256 mm; matrix size = 176×256 mm.

Person-level imaging data preprocessing

We modeled trial-by-trial neural encoding of RPE using amplitude modulation analyses in AFNI (v25.1.15). This approach allowed us to examine how fluctuations in RPE covary with the BOLD signal. The resulting modulation parameters quantify the percent change in BOLD signal per unit increase in RPE, serving as an index of the coupling strength between these variables. Because trial durations differed, varying trial length was also modeled using duration modulation.

Preprocessing steps included standard despiking, slice-timing correction, volume alignment to a base volume, non-linear registration to the MNI template, spatial smoothing to 5mm full-width at half maximum to balance noise reduction with the need to preserve signal from small subcortical structures, masking, and intensity scaling. Pairs of volumes were censored when the framewise displacement—calculated as the Euclidean norm of the derivatives of the translation and rotation parameters—exceeded 1 mm between consecutive TRs. Additional censoring was applied to TRs in which more than 10% of voxels were identified as time-series outliers.

Person-level linear time-series models were estimated using generalized least-squares with restricted maximum likelihood estimation of the temporal autocorrelation structure. Each person-level regression model included six motion parameters and third-order Legendre polynomials to model baseline drift, along with five task-related regressors for each run: stimulus presentation, responses, missed responses (empty file if no responses were missed), feedback, and affect ratings. Task regressors were specified as follows: Stimulus events were modeled with the canonical hemodynamic response function (HRF) as 1.6-second fixed-duration blocks. Responses with variable timings were modeled using a duration-modulated block, which requires amplitude modulation (AM1 parameter) to capture trial-varying response characteristics. Missed responses were modeled analogously to valid responses, using the same duration-modulated HRF and amplitude-modulation structure (AM1). Feedback events were modeled with a gamma function given their fixed, brief (1-second) duration. The feedback regressor incorporated amplitude modulation by RPE values (AM2). Finally, affect ratings at the end of each run were modeled as 6-second fixed-duration HRF-convolved blocks.

Group-level analysis

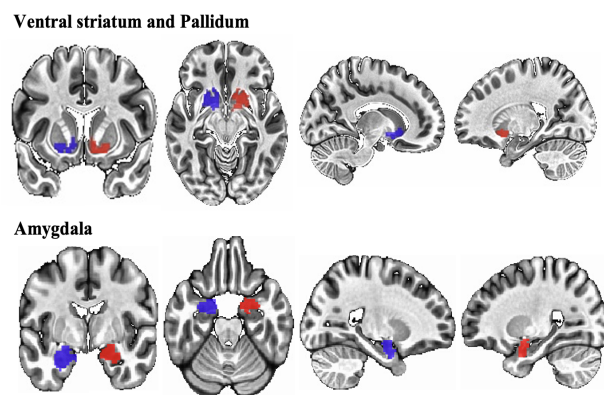
Preliminary whole-brain analyses. We conducted a whole-brain, voxel-wise one-sample t-test in AFNI's 3dttest++ to evaluate whether average RPE-related BOLD responses were different from zero. In an exploratory analysis, AFNI's 3dMVM was used to examine the main effects of anxiety and depression on RPE-modulated brain activation, and whether anxiety interacted with depressive symptoms to predict RPE-modulated activation, with symptoms included as continuous between-person predictors. Multiple-comparison correction was carried out using Monte Carlo simulations with AFNI's 3dClustSim. Analyses were constrained to a whole-brain mask comprising 90,943 voxels. To estimate spatial smoothness, residual time series from the first-level models were submitted to 3dFWHMx with the ACF option, using a Gaussian-plus-mono-exponential model. Smoothness estimates were averaged across participants and produced an effective FWHM of 6.83 mm (ACF parameters: $a=0.7$, $b=2.57$, $c=9.63$). Cluster thresholds were derived using two-sided whole-brain F-tests with second-nearest-neighbor clustering (NN=2, face and edge). A family-wise error rate of $p<.05$ was achieved by applying a voxel-wise threshold of $p<.001$ combined with a minimum cluster size of 13 or more voxels, which served as the correction for the whole-brain models.

Region-of-interest (ROI) analyses. Contrast-specific mean beta values were obtained from subject-level general linear tests (GLTs) and entered into regression analyses conducted in R (version 2025.05.1+513). Anatomical masks were defined using the Julich-Brain Atlas (33), generating separate regions for each hemisphere encompassing the ventral striatal/pallidum area and the amygdala. These masks were constructed by binarizing the combined maximum-probability maps from constituent subregions, including ventral striatal components (fundus of the caudate, putamen, and medial and lateral accumbens), along with the ventral pallidum, and six amygdala areas (centromedial, superficial, laterobasal, ventromedial, intermediate fiber, medial fiber). Each mask was eroded by 1 voxel and intersected with a group-level inclusion mask that retained voxels present in at least 90% of participants. The spatial maps of the ROIs are illustrated in Figure 2.

Primary models tested whether depressive symptoms moderated the association between anxiety symptoms and RPE-modulated BOLD activation in the bilateral ROIs. Additional analyses examined the main effects of anxiety and depressive symptoms on RPE-modulated brain activation for each bilateral ROI, and the results are available in Supplementary Section 8. Age and sex were entered as covariates, and all continuous predictors were mean-centered. Irritability was also explored as a confounding variable. We report separate models for child- and parent-reported CADS without multiple comparison correction, as these measures capture complementary aspects of anxiety and depression, with parent report better indexing observable features and child report indexing subjective processes.

Figure 2

Masks used in the primary ROI analyses.



Note. Left (blue) and right (red) hemisphere ventral striatal/pallidal region and amygdala masks used in the primary ROI analyses. These masks were created using the Juelich-Brain 3D probabilistic atlas and overlaid onto the MNI 2009 atlas.

Results

Demographic information

Eighty-eight participants were included in the analysis sample ($M_{\text{age}} = 14.43$ years, $SD = 2.52$ years; $N = 43$ females, 48.86%). The average IQ was 115.84 ($SD = 14.66$). See Table 1 for demographic information; Supplementary Section 2 for diagnostic group differences; Supplementary Section 3 for

included vs. excluded comparisons; Supplementary Section 5 for zero-order correlations among demographic information, IQ, modeling parameters, internalizing symptoms, affect ratings, average motion in the scanner, RPE-modulated brain activation in the bilateral ROIs, and computational parameters.

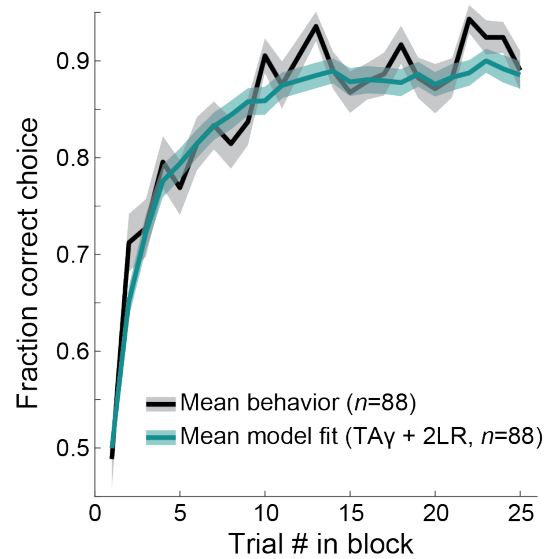
Behavior during the Carnival task

Participants demonstrated successful acquisition of reward contingencies during the baseline trials, as evidenced by a gradual increase in the probability of selecting the higher-value option (Figure 3). The reinforcement learning model closely captured observed behavior. Model comparison indicated that a Rescorla–Wagner specification including separate learning rates for gains and non-gains, along with a Try-Again discounting parameter ($TA\gamma + 2LR$; Model 4), provided the best fit (Vanilla RL (RW): $\log Z = -2325.6$; Two Learning Rates (2LR): $\log Z = -2164.2$; Try Again Trial Gamma ($TA\gamma$): $\log Z = -2277.2$; $TA\gamma + 2LR$: $\log Z = -2024.9$). Additional information regarding missing responses is provided in Supplementary Section 4.

In general, computational parameters (positive and negative learning rates, TA gamma, and softmax beta) were not associated with anxiety and depressive symptoms, except for a significant association between parent-reported MFQ scores and baseline TA gamma ($F(1, 86) = 4.367, p = .040$). See Section 2 of the supplementary for the detailed results of the non-significant results.

Figure 3

Mean probability of choosing the higher-value option across trials during the three baseline runs of the Carnival task.



Note. The black line depicts observed group-averaged behavioral data, and the teal line depicts predictions from the best-fitting RW-RL model (model 4). $\pm$ SEM across participants. Model-based parameter estimates and methods are summarized in the *Reinforcement Learning Models* section.

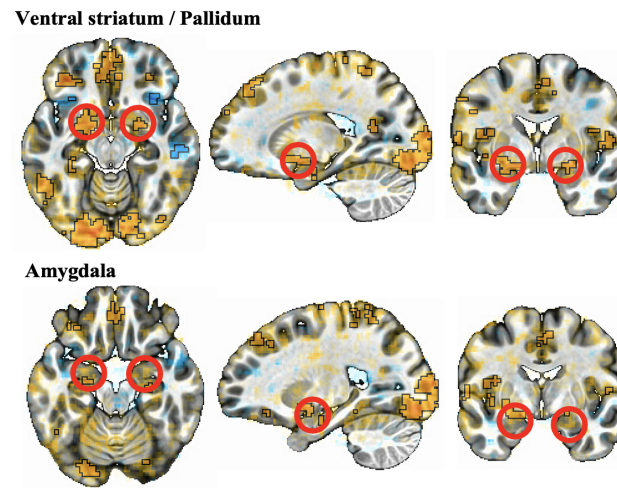
Preliminary whole-brain analyses

The group-level analysis, aligned with prior work on human RPE-related signaling, identified significant RPE-modulated neural activity across 28 clusters, including 23 positive and 5 negative effects. Positive associations were observed in the ventral striatum, posterior cingulate cortex, and medial and orbitofrontal regions, whereas negative associations emerged in bilateral insula and anterior cingulate areas (Figure 4). Detailed cluster information is in Supplementary Section 6.

No clusters survived whole-brain multiple-comparison correction for the anxiety*depression interaction on RPE-modulated BOLD. Significant main effects emerged with parent-reported symptoms that survived correction for multiple comparisons: anxiety was positively related to RPE-modulated BOLD in the left superior frontal cortex, whereas depressive symptoms were negatively related to two left-hemisphere clusters in the superior frontal and middle cingulate cortices. The detailed results and visualization of these findings are available in Supplementary Section 7.

Figure 4

Whole-brain RPE modulated average BOLD activation results across all participants during the baseline trials.



Note. The figure highlights positive RPE encoding in coronal, axial, and sagittal slices of the ventral striatal/pallidal regions, with the region further highlighted in red in the middle images only.

Primary ROI analyses

See Figure 5 for a visualization of the significant interaction effects, described below.

Ventral striatum/pallidum region

Child-reported symptoms: Depressive and anxiety symptoms showed a significant interaction in predicting RPE-modulated brain activation in the right ventral striatal/pallidal ROI. Specifically, as self-reported depressive symptoms increased, the association between anxiety symptoms and RPE-modulated brain activation became increasingly blunted ($b = -0.00039$, 95% CI $[-.000745, -.000031]$, $SE = 0.00018$, $t(84) = -2.16$, $p = .034$). No significant interaction effect emerged in the left hemisphere ($b = -0.00028$, $SE = 0.00018$, 95% CI $[-0.00065, 0.00008]$, $t(84) = -1.53$, $p = .130$). The significance and direction of effects did not change when controlling for age, sex, and irritability scores.

Parent-reported symptoms: No significant interaction effects emerged.

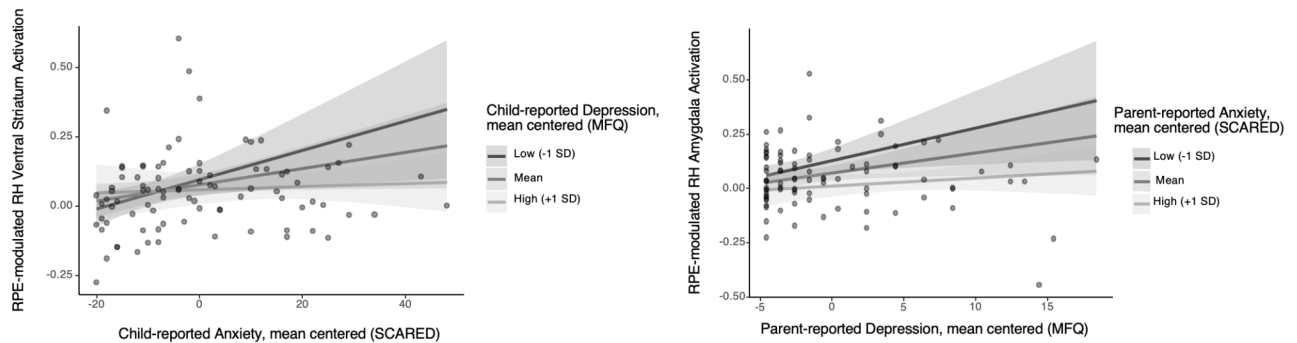
Amygdala

Child-reported symptoms: No significant interaction effects emerged.

Parent-reported symptoms: Anxiety and depression symptoms showed a significant interaction in predicting RPE-modulated brain activation in the right amygdala ROI. Specifically, as parent-reported anxiety symptoms increased, the association between depression symptoms and RPE-modulated brain activation became increasingly blunted ($b = -0.00043$, 95% CI $[-0.00083, -0.00004]$, $SE = 0.0002$, $t(84) = -2.174$, $p = .032$). No significant interaction effect emerged in the left hemisphere ($b = -0.00032$, $SE = 0.00022$, 95% CI $[-0.00075, 0.00011]$, $t(84) = -1.469$, $p = .145$). The significance and direction of effects did not change when controlling for age, sex, and irritability scores.

Figure 5

The interactive effects of anxiety and depression symptoms on RPE-modulated BOLD in the ventral striatal/pallidal area and amygdala.



Note. RH - right hemisphere; MFQ - Mood and Feelings Questionnaire; SCARED - Screen for Child Anxiety Related Emotional Disorders.

Discussion

This study examined whether co-occurring dimensional symptoms of anxiety and depression interact to predict RPE signaling in the ventral striatal/pallidal region and amygdala during reinforcement learning in a transdiagnostic sample of youth aged 8–18 years. By incorporating trial-by-trial, model-derived RPEs as parametric modulators of the BOLD signal, we tested interactive effects of symptom dimensions on neural reward signaling, rather than relying on task-based contrasts or categorical diagnoses. We found preliminary evidence supporting the interactive model of CADS: depressive symptoms moderated the association between anxiety and RPE-modulated activation in the right ventral striatal/pallidal area (child-reported symptoms) and right amygdala (parent-reported symptoms), such that increasing co-occurring symptom severity was associated with progressively blunted RPE encoding. ROI analyses showed no main effects of depressive symptoms on RPE-modulated activation, but a negative main effect of parent-reported anxiety emerged on the right amygdala. Exploratory whole-brain analyses yielded significant associations with parent-reported symptoms that survived multiple comparison correction: anxiety showed a positive association with RPE-modulated BOLD in the left superior frontal cortex, and depressive symptoms were negatively associated with two left hemisphere clusters in the superior frontal and middle cingulate cortices. No effects survived whole-brain correction for the interaction. Together, these results suggest that modeling anxiety and depressive symptoms as independent dimensions may not capture the neurocomputational mechanisms associated with CADS, which instead appear to be shaped by their interaction, with CADS-related alterations in RPE signaling most robustly and specifically detectable within reward-circuit ROIs and only more limited, primarily main-effect associations in broader cortical networks.

A key strength of this study is the use of computational modeling that moves beyond standard task contrasts and instead captures learning signals as they unfold on a trial-by-trial basis. By estimating person-specific RPEs, we were able to examine how individual differences in updating expectations are reflected in reward-related circuitry in youth with CADS, rather than inferring learning processes indirectly from group-level averages. This approach foregrounds person-specific expectancy violations as

a key feature of emotional dysfunction, allowing us to ask how neural responses scale with deviations from anticipated outcomes in a way that is not accessible with conventional contrasts. More broadly, embedding interactive symptom dimensions within this modeling framework illustrates how co-occurring anxiety and depression jointly shape reward learning signals, and provides preliminary evidence for future studies with mechanistically precise hypotheses in larger, fully powered samples.

Both the ventral striatal/pallidal area and the amygdala play central roles in RPE encoding within cortico-striatal-limbic reward circuits and are reliably implicated in mood and anxiety disorders (12,26). The interaction effects in the present study indicate that the lowest activation in these regions is most pronounced when anxiety and depressive symptoms co-occur at elevated levels, rather than reflecting the influence of either domain alone. The significant main effect of parent-reported anxiety on amygdala RPE signaling is consistent with previous reports of altered amygdala RPE encoding in samples with elevated threat sensitivity (34). Furthermore, the corrected whole-brain main effects in left superior frontal and middle cingulate cortices suggest that linear associations between parent-reported anxiety and depression and RPE-related activity extend into higher-order control and monitoring regions, consistent with broader frontal–cingulate alterations in prediction error and reward processing reported in depression and related psychopathology (35). In this context, the ROI interactions may capture a more specific, interaction-sensitive component of CADS within subcortical reward/threat circuits, whereas the whole-brain findings highlight co-occurring but primarily additive cortical correlates of symptom severity. These interpretations remain preliminary, given the statistical considerations and the negative whole-brain analyses, and future work is needed to evaluate them in adequately powered, pre-registered designs.

Child- and parent-reported CADS were differentially associated with RPE-modulated activation in ventral striatal/pallidal and amygdala regions, respectively. Such informant discrepancies between child and parent reports are well-documented in internalizing psychopathology and highlight the importance of modeling symptom associations reported by multiple informants. If this pattern replicates, it could reflect distinct facets of symptom expression, with child reports capturing subjective motivational

and emotional processes, and parent reports indexing observable distress, avoidance, and reactivity. Given the relatively small sample and the early phase of the research, this pattern may also reflect sampling variability or insufficient power to detect consistent interaction effects at the whole-brain level. Our results may be consistent with this interpretation. Depression, which is less behaviorally apparent, was associated with RPE-modulated striatal activation. In contrast, anxiety-related risk, which is more behaviorally apparent, was associated with RPE-modulation in the amygdala, a region central to threat and salience processing. However, as noted above, these dissociations should be interpreted as preliminary, as the relatively lenient statistical thresholds, ROI-based analyses, and uncorrected multiple comparisons increase the risk that some or all of the effects reflect Type I error. Future work with stricter whole-brain correction and preregistered analytic plans will be needed to confirm this pattern.

This study has several strengths, including the use of computational modeling within a transdiagnostic framework and a broad developmental age range. The findings highlight the importance of modeling co-occurring anxiety and depressive symptoms as interacting dimensions rather than examining them in isolation. Several limitations also warrant consideration. The sample size (N=88) was reasonable, but greater trial density per participant is needed to improve estimation of trial-by-trial RPE signaling. The cross-sectional design precluded the assessment of reliability and developmental trajectories, as longitudinal studies are needed to evaluate the stability and predictive utility of these neural markers for identifying youth at elevated risk for CADS. The sample was oversampled for irritable and inattentive/hyperactive presentations, and although controlling for irritability did not meaningfully change the conclusions, the results may have differed had the sample been recruited solely for CADS presentations. Finally, incorporating multi-informant and multi-method clinical assessments would strengthen confidence in these findings, which currently rely on single child- and parent-report measures.

Conclusion

Taken together, these findings demonstrate that co-occurring anxiety and depressive symptoms synergistically alter how the developing brain learns about reward contingencies, with youth exhibiting elevated CADS showing the greatest disruption in reward-related neural signaling. This pattern has

important implications for how risk is conceptualized in depression and anxiety, given that up to 70% of affected youth present with co-occurring symptoms, yet most mechanistic work continues to examine each dimension in isolation or rely on categorical diagnoses. Our results suggest that when simultaneously elevated anxiety and depression symptoms are observed, the clinical symptom profile is associated with blunted reward signaling, which may contribute to the elevated chronicity, impairment, treatment resistance, and suicide risk in CADS. Future computational and longitudinal research is needed to track person-specific, trial-by-trial changes in reward signaling in youth with CADS, to develop neurocognitive markers to inform earlier identification and targeted intervention.

References

1. ter Meulen WG, Draisma S, van Hemert AM, Schoevers RA, Kupka RW, Beekman ATF, et al. Depressive and anxiety disorders in concert—A synthesis of findings on comorbidity in the NESDA study. *Journal of Affective Disorders*. 2021 Apr 1;284:85–97. doi:10.1016/j.jad.2021.02.004
2. Kircanski K, LeMoult J, Ordaz S, Gotlib IH. Investigating the nature of co-occurring depression and anxiety: Comparing diagnostic and dimensional research approaches. *J Affect Disord*. 2017 Jul;216:123–35. doi:10.1016/j.jad.2016.08.006 PubMed PMID: 27554605; PubMed Central PMCID: PMC5308464.
3. Cuijpers P, Miguel C, Ciharova M, Ebert D, Harrer M, Karyotaki E. Transdiagnostic treatment of depression and anxiety: a meta-analysis. *Psychol Med*. 2023;53(14):6535–46. doi:10.1017/S0033291722003841 PubMed PMID: 36606456; PubMed Central PMCID: PMC10600931.
4. Balázs J, Miklósi M, Keresztény Á, Hoven CW, Carli V, Wasserman C, et al. Adolescent subthreshold-depression and anxiety: psychopathology, functional impairment and increased suicide risk. *Journal of Child Psychology and Psychiatry*. 2013;54(6):670–7. doi:10.1111/jcpp.12016
5. Altamura AC, Montresor C, Salvadori D, Mundo E. Does comorbid subthreshold anxiety affect clinical presentation and treatment response in depression? A preliminary 12-month naturalistic study. *Int J Neuropsychopharmacol*. 2004 Dec 1;7(4):481–7. doi:10.1017/S1461145704004626
6. Cummings CM, Caporino NE, Kendall PC. Comorbidity of anxiety and depression in children and adolescents: 20 years after. *Psychological Bulletin*. 2014;140(3):816–45. doi:10.1037/a0034733
7. Cowden RG, Chapman I, Houghtaling A. Additive, Curvilinear, and Interactive Relations of Anxiety and Depression With Indicators of Psychosocial Functioning. *Psychol Rep*. 2021 Apr 1;124(2):627–50. doi:10.1177/0033294120918809
8. Rivas-Vazquez RA, Saffa-Biller D, Ruiz I, Blais MA, Rivas-Vazquez A. Current Issues in Anxiety and Depression: Comorbid, Mixed, and Subthreshold Disorders. *Professional Psychology: Research and Practice*. 2004;35(1):74–83. doi:10.1037/0735-7028.35.1.74

9. Hankin BL, Snyder HR, Gulley LD, Schweizer TH, Bijttebier P, Nelis S, et al. Understanding comorbidity among internalizing problems: Integrating latent structural models of psychopathology and risk mechanisms. *Development and Psychopathology*. 2016 Nov;28(4pt1):987–1012. doi:10.1017/S0954579416000663
10. Brown TA, Barlow DH. A proposal for a dimensional classification system based on the shared features of the DSM-IV anxiety and mood disorders: Implications for assessment and treatment. *Psychological Assessment*. 2009;21(3):256–71. doi:10.1037/a0016608
11. Goldberg D. Plato versus Aristotle: Categorical and dimensional models for common mental disorders. *Comprehensive Psychiatry*. 2000 Mar 1;Subthreshold Disorders in Psychiatry41(2, Supplement 1):8–13. doi:10.1016/S0010-440X(00)80002-4
12. Kumar P, Goer F, Murray L, Dillon DG, Beltzer ML, Cohen AL, et al. Impaired reward prediction error encoding and striatal-midbrain connectivity in depression. *Neuropsychopharmacology*. 2018 Jun;43(7):1581–8. doi:10.1038/s41386-018-0032-x PubMed PMID: 29540863; PubMed Central PMCID: PMC5983542.
13. Pike AC, Robinson OJ. Reinforcement Learning in Patients With Mood and Anxiety Disorders vs Control Individuals: A Systematic Review and Meta-analysis. *JAMA Psychiatry*. 2022 Apr 1;79(4):313–22. doi:10.1001/jamapsychiatry.2022.0051
14. Mukherjee D, Filipowicz ALS, Vo KD, Satterthwaite T, Kable J. Reward and punishment reversal learning in major depressive disorder [Internet]. *PsyArXiv*; 2020 [cited 2026 Apr 26]. Available from: https://osf.io/preprints/psyarxiv/aqgx3_v1/ doi:10.31234/osf.io/aqgx3
15. Brown VM, Zhu L, Solway A, Wang JM, McCurry KL, King-Casas B, et al. Reinforcement Learning Disruptions in Individuals With Depression and Sensitivity to Symptom Change Following Cognitive Behavioral Therapy. *JAMA Psychiatry*. 2021 Oct 1;78(10):1113. doi:10.1001/jamapsychiatry.2021.1844

16. Katz BA, Matanky K, Aviram G, Yovel I. Reinforcement sensitivity, depression and anxiety: A meta-analysis and meta-analytic structural equation model. *Clinical Psychology Review*. 2020 Apr 1;77:101842. doi:10.1016/j.cpr.2020.101842
17. Garrison J, Erdeniz B, Done J. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neuroscience & Biobehavioral Reviews*. 2013 Aug;37(7):1297–310. doi:10.1016/j.neubiorev.2013.03.023
18. Chase HW, Kumar P, Eickhoff SB, Dombrowski AY. Reinforcement learning models and their neural correlates: An activation likelihood estimation meta-analysis. *Cogn Affect Behav Neurosci*. 2015 Jun 1;15(2):435–59. doi:10.3758/s13415-015-0338-7
19. Santo-Angles A, Fuentes-Claramonte P, Argila-Plaza I, Guardiola-Ripoll M, Almodóvar-Payá C, Munuera J, et al. Reward and fictive prediction error signals in ventral striatum: asymmetry between factual and counterfactual processing. *Brain Struct Funct*. 2021 Jun 1;226(5):1553–69. doi:10.1007/s00429-021-02270-3
20. Calabro FJ, Montez DF, Larsen B, Laymon CM, Foran W, Hallquist MN, et al. Striatal dopamine supports reward expectation and learning: A simultaneous PET/fMRI study. *NeuroImage*. 2023 Feb 15;267:119831. doi:10.1016/j.neuroimage.2022.119831
21. Cromwell HC, Schultz W. Effects of expectations for different reward magnitudes on neuronal activity in primate striatum. *J Neurophysiol*. 2003 May;89(5):2823–38. doi:10.1152/jn.01014.2002 PubMed PMID: 12611937.
22. Schultz W, Dayan P, Montague PR. A neural substrate of prediction and reward. *Science*. 1997 Mar 14;275(5306):1593–9. doi:10.1126/science.275.5306.1593 PubMed PMID: 9054347.
23. Schultz W. Dopamine reward prediction error coding. *Dialogues in Clinical Neuroscience*. 2016 Mar 31;18(1):23–32. doi:10.31887/DCNS.2016.18.1/wschultz PubMed PMID: 27069377.
24. Schultz W. The Dopamine Reward Prediction Error Signal. *SSRN Journal*. 2024. doi:10.2139/ssrn.4816666

25. Queirazza F, Fouragnan E, Steele JD, Cavanagh J, Philiastides MG. Neural correlates of weighted reward prediction error during reinforcement learning classify response to cognitive behavioral therapy in depression. *Sci Adv.* 2019 Jul;5(7):eaav4962. doi:10.1126/sciadv.aav4962 PubMed PMID: 31392266; PubMed Central PMCID: PMC6669013.
26. Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T, Hrynkiewicz L, et al. Association of Neural and Emotional Impacts of Reward Prediction Errors With Major Depression. *JAMA Psychiatry.* 2017 Aug 1;74(8):790–7. doi:10.1001/jamapsychiatry.2017.1713
27. Cavanagh JF, Bismark AW, Frank MJ, Allen JJB. Multiple Dissociations Between Comorbid Depression and Anxiety on Reward and Punishment Processing: Evidence From Computationally Informed EEG. *Comput Psychiatr.* 2019 Jan;3:1–17. doi:10.1162/cpsy_a_00024 PubMed PMID: 31149639; PubMed Central PMCID: PMC6515849.
28. Grabowska A, Zabielski J, Senderecka M. Machine learning reveals differential effects of depression and anxiety on reward and punishment processing. *Sci Rep.* 2024 Apr 10;14(1):8422. doi:10.1038/s41598-024-58031-9
29. Wiehler A, Peters J. Decomposition of Reinforcement Learning Deficits in Disordered Gambling via Drift Diffusion Modeling and Functional Magnetic Resonance Imaging. *Computational Psychiatry.* 2024 Mar 20;8(1). doi:10.5334/cpsy.104
30. Tang H, Bartolo R, Averbach BB. Ventral frontostriatal circuitry mediates the computation of reinforcement from symbolic gains and losses. *Neuron.* 2024 Nov;112(22):3782-3795.e5. doi:10.1016/j.neuron.2024.08.018
31. Yacubian J, Gläscher J, Schroeder K, Sommer T, Braus DF, Büchel C. Dissociable Systems for Gain- and Loss-Related Value Predictions and Errors of Prediction in the Human Brain. *J Neurosci.* 2006 Sep 13;26(37):9530–7. doi:10.1523/JNEUROSCI.2915-06.2006 PubMed PMID: 16971537.
32. Angold A, Costello E, Messer S, Pickles A, Winder F, Silver D. The Development of a Questionnaire for Use in Epidemiological Studies of Depression in Children and Adolescents. *International Journal of Methods in Psychiatric Research.* 1995 Jan 1;5:237–49.

33. Amunts K, Mohlberg H, Bludau S, Zilles K. Julich-Brain: A 3D probabilistic atlas of the human brain's cytoarchitecture. *Science*. 2020 Aug 21;369(6506):988–92. doi:10.1126/science.abb4588
34. Enten LK, DeMarco A, Kline R, Globisch FM, Dunsmoor JE, Cisler JM, et al. Altered amygdala and striatal responsivity and prediction error encoding of temporal reward dynamics in post-trauma psychopathology. *Neuropsychopharmacology*. 2026 Apr;51(5):904–15. doi:10.1038/s41386-025-02276-z PubMed PMID: 41249466; PubMed Central PMCID: PMC13013824.
35. Zhang S, Zhang Y, Ma W, Qi Z, Wang Y, Tao Q. Neural correlates of negative emotion processing in subthreshold depression. *Soc Cogn Affect Neurosci*. 2022 Jul 1;17(7):655–61. doi:10.1093/scan/nsac003

Section 1

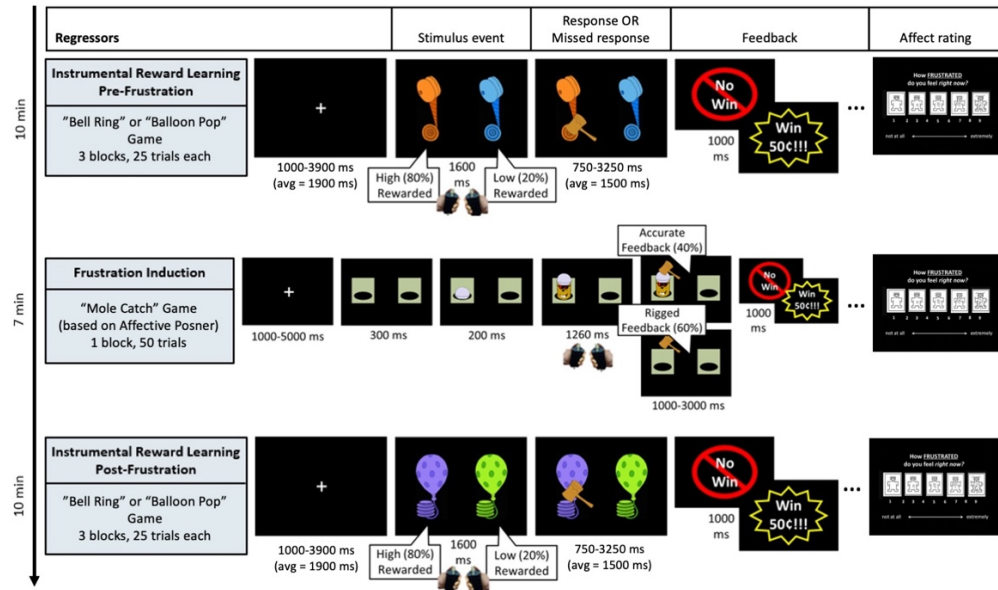
The Carnival Task

The “Carnival” paradigm is a child-friendly, task-based fMRI paradigm designed to interrogate RPE encoding during reinforcement learning pre- and post-frustration. It consists of a series of two-armed bandit reinforcement learning tasks integrated with an acute frustration induction. In simulated Carnival games (“Bell Ring” and “Balloon Pop”), participants learn to choose between stimulus pairs to maximize monetary rewards, conveyed via trial-by-trial feedback (“Win 50¢,” “No Win”). The first 3 runs (“Bell Ring” or “Balloon Pop,” counterbalanced) form the pre-frustration phase (3 runs $\times$ 25 trials = 75 trials; 10 minutes), with an 80:20 reward:nonreward ratio for the better option and 20:80 for the worse. Participants respond using grip-force dynamometers in their left and right hands. Next, a second game (“Mole Catch” for all participants) comprises one run (50 trials; 7 minutes), modeled on the Affective Posner task and rigged to induce frustration with a 40:60 reward:nonreward ratio on correct trials and 60:40 on rigged trials. Incorrect trials are always nonrewarded, while rigged trials retract the target (mole) immediately after a correct response, creating the impression of narrowly missing or being denied an expected reward. The final three runs (“Bell Ring” or “Balloon Pop,” whichever was not used initially) form the post-frustration phase (3 runs $\times$ 25 trials = 75 trials; 10 minutes; 80:20 reward:nonreward ratio). After each phase, participants rate frustration, excitement, and unhappiness on a nine-point Likert scale, anchored by five manikin figures illustrating increasing emotional intensity. See Supplementary Figure 1 for a depiction of the Carnival Task.

Before the fMRI scan, participants completed 10 trials each of the “Bell Ring,” “Balloon Pop,” and “Mole Catch,” phases of the task, with 1 rigged trial in the “Mole Catch” phase.

Supplementary Figure 1

Overview of the Carnival instrumental learning task.



Section 2

Diagnostic Group Differences in Demographic Characteristics, IQ, Internalizing Symptoms, and Model Parameters

The primary diagnoses included 24 participants with ADHD (27.27%), 18 with disruptive mood dysregulation disorder (20.45%), 12 with anxiety (13.64%), 20 with major depressive disorder (22.73%), and 14 healthy volunteers (15.91%).

Demographic information: A one-way ANOVA revealed significant age differences among diagnostic groups, $F(4, 83) = 5.86, p < .001$. Post hoc comparisons indicated that participants in the MDD group were significantly older than those in the ADHD ($p = .03$) and DMDD ($p < .001$) groups. No other pairwise differences were significant (all $ps > .07$). Mean ages ($\pm$ SD) were 14.19 ± 2.09 years for ADHD, 16.27 ± 1.35 years for MDD, and 12.76 ± 2.79 years for DMDD, 14.08 ± 2.72 years for anxiety, and 14.66 ± 2.51 for the healthy volunteers.

IQ: No significant IQ differences emerged across diagnostic groups, $F(4, 65) = 0.51, p = .73$.

Anxiety symptoms:

1. *Child-report:* There was a significant main effect of diagnosis on child-reported SCARED scores, $F(4, 83) = 10.34, p < .001$. Post hoc comparisons indicated that participants with an ADHD ($p = .027$) diagnosis and those in the healthy comparison group ($p = .002$) had significantly lower anxiety scores than those who had an anxiety diagnosis. Additionally, participants with an ADHD ($p < .001$) and DMDD ($p < .006$) diagnosis, as well as healthy volunteers ($p < .001$), had lower child-reported anxiety scores than those with an MDD diagnosis. No other significant pairwise comparisons emerged (all $ps > .22$). Mean child-reported anxiety scores ($\pm$ SD) were 13.958 ± 12.551 for ADHD, 17.778 ± 16.969 for DMDD, 33.050 ± 14.303 for MDD, 28.083 ± 11.221 for the anxiety group, and 7.786 ± 7.607 for healthy volunteers.
2. *Parent-report:* There was a significant main effect of diagnosis on parent-reported SCARED scores, $F(4, 83) = 13.59, p < .001$. Post hoc comparisons indicated that participants with an ADHD ($p = .003$) diagnosis and those in the healthy comparison group ($p < .001$) had significantly lower anxiety scores than those who had an anxiety diagnosis. Additionally, participants with an ADHD ($p < .001$) and DMDD ($p = .015$) diagnosis, as well as healthy volunteers ($p < .001$), had lower parent-reported anxiety scores than those with an MDD diagnosis. Finally, participants with a DMDD diagnosis ($p < .001$) had significantly higher anxiety scores than those in the healthy volunteer group. No other significant pairwise comparisons emerged (all $ps > .12$). Mean parent-reported anxiety scores ($\pm$ SD) were 10.5 ± 8.361 for ADHD, 20.944 ± 13.956 for DMDD, 23.600 ± 13.280 for MDD, 24.333 ± 7.303 for the anxiety group, and 2.071 ± 2.868 for healthy volunteers.

Depression Symptoms:

1. *Child-report:* There was a significant main effect of diagnosis on MFQ scores, $F(4, 83) = 7.805, p < .001$. Post hoc comparisons indicated that participants with an ADHD ($p < .001$) diagnosis and those in the healthy comparison group ($p < .001$) had significantly lower depression symptom scores than those who had a depression

diagnosis. No other significant pairwise comparisons emerged (all $ps > .069$). Mean child-reported depression scores ($\pm$ SD) were 3.958 ± 5.729 for ADHD, 6.333 ± 6.808 for DMDD, 10.950 ± 4.548 for MDD, 6.667 ± 5.549 for the anxiety group, and 1.429 ± 2.652 for healthy volunteers.

2. *Parent-report:* There was a significant main effect of diagnosis on MFQ scores, $F(4, 83) = 6.275, p < .001$. Post hoc comparisons indicated that participants with an ADHD ($p = .015$) diagnosis and those in the healthy comparison group ($p < .001$) had significantly lower depression symptom scores than those who had a depression diagnosis. Furthermore, participants with a DMDD ($p = .004$) diagnosis had a significantly higher depression symptom score than those in the healthy volunteer group. No other significant pairwise comparisons emerged (all $ps > .102$). Mean child-reported depression scores ($\pm$ SD) were 2.875 ± 3.893 for ADHD, 6.722 ± 5.969 for DMDD, 7.700 ± 6.814 for MDD, 4.750 ± 4.181 for the anxiety group, and 0.214 ± 0.426 for healthy volunteers.

Model Parameters: No significant diagnostic group differences emerged for the model parameters during the baseline trials (positive learning rate: $F(4, 83) = 0.431, p = .786$; negative learning rate: $F(4, 83) = 0.716, p = .586$; TA gamma: $F(4, 83) = 1.213, p = .586$; softmax beta: $F(4, 83) = 1.087, p = .368$).

Similarly, age (positive learning rate: $F(1, 86) = 0.905, p = .344$; negative learning rate: $F(1, 86) = 0.062, p = .805$; TA gamma: $F(1, 86) = 0.394, p = .532$; softmax beta: $F(1, 86) = 1.285, p = .26$), sex (positive learning rate: $t(86) = 1.087, p = .28$; negative learning rate: $t(86) = 1.888, p = .062$; TA gamma: $t(86) = -1.046, p = .299$; softmax beta: $t(86) = 1.15, p = .253$), child-reported SCARED scores (positive learning rate: $F(1, 86) = 1.581, p = .212$; negative learning rate: $F(1, 86) = 0.049, p = .826$; TA gamma: $F(1, 86) = 0.127, p = .723$; softmax beta: $F(1, 86) = 0.215, p = .644$), parent-reported SCARED scores (positive learning rate: $F(1, 86) = .281, p = .598$; negative learning rate: $F(1, 86) = 0.002, p = .968$; TA gamma: $F(1, 86) = 2.254, p = .137$; softmax beta: $F(1, 86) = 0.791, p = .376$), child-reported MFQ (positive learning rate: $F(1, 86) = 1.39, p = .126$; negative learning rate: $F(1, 86) = 0.395, p = .531$; TA gamma: $F(1, 86) = 0.624, p = .432$; softmax beta: $F(1, 86) = 0.033, p = .856$), and parent-reported MFQ (positive learning rate: $F(1, 86) = 0.040, p = .841$; negative learning rate: $F(1, 86) = 0.094, p = .759$; softmax beta: $F(1, 86) = 3.391, p = .069$), were not significantly associated with any of the model parameters. The parent-reported MFQ score was significantly associated with the baseline TA gamma ($F(1, 86) = 4.367, p = .040$) parameter. IQ was also significantly associated with negative learning rate ($F(1, 68) = 4.11, p = .047$), TA gamma ($F(1, 68) = 3.071, p = .084$), and softmax beta ($F(1, 68) = 5.167, p = .026$), but not with positive learning rate ($F(1, 68) = 0.176, p = .676$).

Section 3

Comparison of Included vs. Excluded Participants on Demographics, IQ, and Irritability

Included and excluded participants did not differ significantly on sex ($\chi^2 = 0.474$, $df = 1$, $p = .491$) or diagnostic distribution ($\chi^2 = 4.0599$, $df = 5$, $p = .541$), IQ (mean in: excluded group: 116.167, included group = 115.843, $t(18.514) = 0.09$, $p = .93$), child- (mean in: excluded group: 22.611, included group: 20.023, $t(25.446) = 0.663$, $p = .513$) and parent-reported SCARED (mean in: excluded group: 17.444, included group: 16.159, $t(25.423) = 0.416$, $p = .681$), or child- (mean in: excluded group: 5.235, included group: 6.00, $t(24.324) = -0.515$, $p = .611$) and parent reported MFQ scores (mean in: excluded group: 4.833, included group: 4.591, $t(25.64) = 0.18$, $p = .859$). A significant average age difference emerged between the included and the excluded groups, such that the excluded group was on average younger than the included group (mean age in the excluded group = 12.518 years; included group = 14.429 years; $t(23.818) = -2.828$, $p = .009$).

Section 4

Missing responses during the task

The number of total missed responses were not significantly associated with child- ($\beta = -0.035$, 95% CI [-0.096, 0.026], $t(86) = -1.145$, $p = .255$) and parent-reported SCARED scores ($\beta = -0.013$, 95% CI [-0.088, 0.061], $t(86) = -0.358$, $p = .721$), child- ($\beta = -0.105$, 95% CI [-0.262, 0.052], $t(86) = -1.327$, $p = .188$) and parent-reported MFQ scores ($\beta = 0.024$, 95% CI [-0.152, 0.201], $t(86) = -0.275$, $p = .784$), diagnostic group (all pairwise comparisons $p > .684$), age ($\beta = -0.121$, 95% CI [-0.506, 0.264], $t(86) = -0.627$, $p = .533$) or sex ($\beta = -0.023$, 95% CI [-1.956, 1.910], $t(86) = -0.023$, $p = .981$), but they were with IQ ($\beta = 0.096$, 95% CI [-0.17, -0.019], $t(68) = -2.5$, $p = .015$), such that higher IQ points were associated with fewer missing responses (“Try Again”s).

Section 5

Supplementary Table 3

Zero-order Pearson correlations between pairs of demographic variables, anxiety, depression scores, brain measures, RW-RL model parameters, and affect ratings after the baseline runs.

	Age (years)	IQ	SCARED (child- report)	SCARED (parent- report)	MFQ (child- report)	MFQ (parent- report)	Avg. motion per TR	Left Striatal/ Pallidal Area	Right Striatal/ Pallidal Area	Left Amygdala	Right Amygdala	Beta	Positive LR	Negative LR	TAgamma	Unhappy rating	Excited rating	Frustrated rating
Age (years)	1																	
IQ	-.003	1																
SCARED ¹ (child-report)	.178	.085	1															
SCARED (parent-report)	.082	-.085	.532***	1														
MFQ ² (child-report)	.231*	0.135	.708***	.427***	1													
MFQ (parent-report)	.124	-.093	.381***	.728***	.456***	1												
Avg. motion per TR ³	-.390***	-.135	-.162	.155	-.096	.054	1											
Left Striatal / Pallidal Area	.107	-.323**	.091	.046	.102	-.03	.179	1										
Right Striatal / Pallidal Area	.001	-.136	.101	.101	.019	.005	.151	.776***	1									
Left Amygdala	.089	.036	-.098	-.177	.043	-.031	-.012	.041	.06	1								
Right Amygdala	-.012	.057	-.068	-.271*	-.028	-.127	.058	.094	.084	.697***	1							
Beta	.121	.266*	.05	-.095	.02	-.195	-.287**	0	-.032	.085	.025	1						
Positive LR ⁴	.102	-.051	.134	-.057	.164	.022	-.064	.094	.029	-.193	-.107	-.232*	1					
Negative LR	-.027	-.239*	-.024	-.004	-.068	-.033	.011	.213*	.236*	-.206	-.181	-.151	-.074	1				
TAgamma	-.067	.208	-.038	-.16	-.085	-.220*	-.109	-.125	.016	.055	.093	.018	-.051	-.105	1			
Unhappy rating	.116	-.026	.476***	.078	.355***	.144	-.132	-.059	-.091	.059	-.02	-.038	.196	-.061	-.018	1		
Excited rating	-.099	.103	.038	-.094	.041	-.056	.005	-.128	-.08	.054	.087	-.203	.132	.09	-.039	.101	1	
Frustrated rating	.078	-.003	.477***	.069	.367***	.048	-.116	.067	-.028	.052	.058	-.128	.194	.033	-.008	.614***	.256*	1

¹ SCARED - Screen for Child Anxiety Related Emotional Disorders; ² MFQ – Mood and Feelings Questionnaire; ³ TR – Repetition Time; ⁴ LR - learning rate; *** p < .001; ** p < .01; * p < .05

Section 6

Supplementary Table 4

Summary of clusters showing a significant RPE modulated average BOLD activation across all participants and across pre- and during the three baseline trials.

ROI overlap and location	Cluster size		Coordinates (center of mass)		Coordinates (at peak)			Mean	SEM	Max Int
	k	CM RL	CM AP	CM IS	Peak x	Peak y	Peak z			
R V1 (area hOc1, 16.7%) R V2 (hOc2, 14.4%); L Lingual Gyrus (hOc3v, 10.1%); L V2 (hOc2, 9.5%)	1512	-4.2	94.7	-0.6	-27.5	94.2	8	0.1556	0.0016	0.358
L Frontal to Occipital (GapMap, 52.7%); R Frontal to Occipital (GapMap, 27.6%)	1053	1.9	45.8	27.6	7.5	56.8	13	0.1238	0.0013	0.3403
L Frontal I (GapMap, 26.1%); L Frontal Pole (15.2%); R Frontal Pole (8.2%); L Pregenua Anterior Cingulate cortex (8.1%); L Superior Frontal Gyrus (6.6%)	760	6.9	-52.2	9.9	5	-50.8	-7	0.1547	0.0019	0.3519
R Superior Parietal Lobule (7PC, 15.1%; 5L, 8.5%; 7A, 8.1%); R Postcentral Sulcus (area 2, 15%; area 1, 6%); Right Inferior Parietal Lobule (PFt, 14.9%; PFcm, 9%);	739	-42.5	31.5	47	-57.5	21.8	40.5	0.1041	0.0012	0.3375
L Inferior Parietal Lobule (PF, 23.8%; PFt, 23.1%; PFop, 7.4%; PFcm, 4.9%); L Superior Parietal Lobule (7PC, 7.8%; 7A, 5.3%); L Postcentral Gyrus (7.2%)	543	52.4	33.6	41.8	60	26.8	40.5	0.0933	0.0012	0.233
R Insula (Id8, 25%; Id7, 12%; Id6, 10%; Id10, 8.2%); R Frontal Operculum (15.1%); R Frontal to Temporal I (GapMap, 10.9%); R Inferior Frontal Gyrus (8.9%);	210	-38.3	-21.9	-1.5	-32.5	-20.8	-9.5	-0.1462	0.0025	-0.2837
R Frontal I (GapMap, 67.7%); R Frontal to Occipital (GapMap, 28.8%);	183	-5	-28.2	37.8	-5	-33.2	33	-0.1176	0.002	-0.2203
L Inferior Parietal Lobule (85.8%); L Temporal to Parietal (GapMap, 9.5%)	159	44.7	74.8	32	47.5	69.2	28	0.1031	0.0023	0.218
L Temporal to Parietal (GapMap, 83.9%); L Fusiform Gyrus (12.5%)	156	55.2	55.8	-8.5	55	51.8	-12	0.0877	0.0018	0.1814
L Frontal I (GapMap, 48.9%); L Inferior Frontal Gyurs (42.7%)	152	49.4	-35.3	11.3	52.5	-30.8	10.5	0.1268	0.0022	0.216
L Insula (Id8, 33.3%; Id7, 14.2%; Id6, 12%); L Frontal to Temporal I (GapMap, 18.2%); L Frontal Operculum (15.6%)	140	34.9	-21.9	-0.4	32.5	-25.8	-4.5	-0.1194	0.0022	-0.1928

L Fundus of Putamen, Ventral Striatum (26.8%); Left Basal Forebrain (15.2%); Left Amygdala (14.4%); Left Frontal to Temporal II (GapMap, 11.7%); Left Lateral Accumbens (5.2%)	110	20.9	0.6	-10.9	20	-0.8	-12	0.1494	0.0059	0.2935
R Superior Temporal Sulcus (STS2, 49%; STS1, 43.1%); Left Orbitofrontal Cortex (Fo7, 66%; Fo3, 14.5%); L Frontal to Temporal I (GapMap, 18.3%)	95	-53.5	24.9	-7.3	-52.5	26.8	-9.5	-0.1234	0.0025	-0.1951
L Inferior Frontal Gyrus (86.9%); L Frontal II (GapMap, 3.7%)	72	35.6	-35	-14.3	35	-35.8	-12	0.1251	0.0035	0.1941
L Insula (Id4, 34.1%; Id6, 15.7%; Id5, 8.5%) L Frontal Operculum (32.1%)	67	50.5	-8	23.2	50	-8.2	23	0.1004	0.002	0.1447
R Precentral Gyrus (41.1%); R Postcentral Gyrus (26.9%); R Frontal II (Gapmap, 20.6%)	63	39.5	1.9	10.9	37.5	4.2	15.5	0.0939	0.0017	0.1196
R Precentral Gyrus (75.7%); R Frontal II (Gapmap, 24.3%)	63	-35.3	21.3	51.4	-37.5	21.8	48	0.1019	0.0023	0.1434
R Temporal to parietal (GapMap, 77.5%); Right Lateral Occipital Cortex (17.6%)	56	-28	11.1	66.1	-27.5	9.2	70.5	0.1011	0.0043	0.2428
R Inferior Parietal Lobule (64.2%); R Temporal to Parietal (GapMap, 17.3%)	53	-52.3	63.6	-4	-52.5	64.2	0.5	0.0939	0.0036	0.1883
L Parahippocampal Gyrus (53.2%); L Hippocampus (CA1, 22%; DG, 6.8%); L Fusiform Gyrus (11.6%)	48	-46	70.4	29.6	-37.5	69.2	30.5	0.084	0.0025	0.1342
R Frontal to Occipital (GapMap, 86.8%); R Superior Parietal Lobule (11.5%)	43	31.1	37.7	-13.4	32.5	34.2	-14.5	0.0976	0.0029	0.1499
R Frontal Operculum (61%); R Insula (Id6, 18.9%; Id4, 18.5%)	32	-9.2	69.7	39.4	-5	71.8	40.5	-0.0922	0.0028	-0.126
R Basal Forebrain (30.2%); R Frontal to Temporal II (GapMap, 22.8%); R Fundus of the Putamen, Ventral Striatum (22%); R Amygdala (19.8%)	21	-39.1	-2.2	13.5	-45	-0.8	13	0.1015	0.0029	0.1275
R Frontal II (GapMap, 55.7%); R Inferior Frontal Gyrus (40.9%)	18	-23.4	0.1	-11.4	-25	1.8	-12	0.147	0.0083	0.2163
L Frontal II (GapMap, 94.4%); L Inferior Frontal Gyrus (5.1%)	17	-63	-6.8	15	-65	-3.2	13	0.1263	0.0103	0.2102
L Frontal Operculum (51.7%); L Superior Temporal Gyrus (19%); Left Opercular Area 4 (10.9%); L Heschl Gyrus (9.5%); L Postcentral Gyrus (8.2%)	16	56.7	-4.8	38.2	57.5	1.8	40.5	0.0911	0.0044	0.1242
R Postcentral Gyrus (38.4%); R Frontal II (GapMap, 30.8%); R Precentral Gyrus (4a, 22.6; 4p, 2.5%)	13	60.7	2.3	6.9	60	1.8	5.5	0.1079	0.0082	0.153
	13	-18.9	26.5	62	-17.5	24.2	63	0.0924	0.0032	0.1116

¹ Cluster-corrected voxel-wise t-test results are presented, summarizing regions showing a main effect of RPE modulation. Anatomical locations are based on JulichBrain v3.1 atlas for MNI 2009c asymmetric space. For t tests, df=87. k=number of voxels in cluster; CM=center of mass of cluster; LR=left-right (x); AP=anterior-posterior axis (y); IS=inferior-superior axis (z); SEM=standard error of the mean; GapMap - uncharted parts of the region. Coordinates are reported in RAI orientation (Right-Anterior-Inferior). Clusters are two-sided ($t = \pm 3.4059$, $p = .001$), meet a minimum size of 13 voxels, and are defined using NN = 2 connectivity.

Section 7

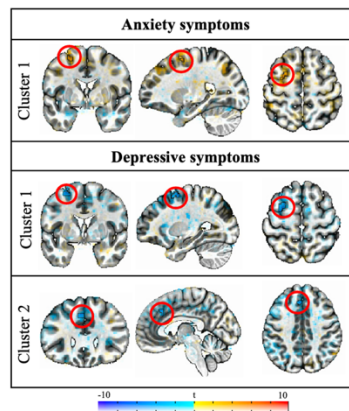
Whole-brain analyses examining the main effects of anxiety and depression symptoms on RPE-modulated brain activation.

A) Anxiety symptoms

- **Child-report:** No clusters survived multiple comparisons correction.
- **Parent-report:** One cluster survived multiple comparisons correction. The results indicate that there was a significant positive association between the magnitude of the modulated BOLD signal and the parent-reported anxiety symptoms in the cluster.

B) Depressive symptoms

- **Child-report:** No clusters survived multiple comparisons correction.
- **Parent-report:** Two clusters survived multiple comparisons correction. The results indicate that in both clusters, there was a significant negative association between the magnitude of the modulated BOLD signal and the parent-reported depressive symptoms.



Symptom / Cluster #	ROI Overlap and Location	Cluster size, k	CM RL	CM AP	CM IS	Peak x	Peak y	Peak z	Mean	SEM	Max Int
Anxiety / Cluster 1	L Frontal II (GapMap, 67.8%); L superior frontal sulcus (25.1%); L precentral gyrus (4.7%)	17	26.4	1.4	57.6	27.5	1.8	58	3.7465	0.0707	4.378
Depression / Cluster 1	L Frontal II (GapMap, 61.9%); L superior frontal sulcus (28%); L precentral gyrus (4.8%)	30	27.4	-1.7	59.1	27.5	-0.8	60.5	-3.9573	0.0798	-5.3385
Depression / Cluster 2	L superior middle gyrus (82.3%); L middle cingulate cortex (17.7%) ¹	15	5.0	-27.7	38.6	5	-28.2	40.5	-3.9344	0.1068	-5.0693

¹ The Julich Atlas template space did not report any overlapping location with this atlas-defined region. For this cluster, Eickhoff–Zilles macro labels from the N27 (CA_ML_18_MNI) atlas were used.

Section 8

Detailed statistical results of the main effects of anxiety and depressive symptoms on RPE-modulated BOLD in the selected ROIs.

1. Ventral striatal/pallidal region

The significance and direction of effects did not change when controlling for age and sex.

Child-reported symptoms:

1. **Anxiety symptom:** No significant main effect emerged for anxiety symptoms:
 - a. right hemisphere: $b = 0.00089$, $SE = 0.00095$, 95% CI [-0.00099, 0.00278], $t(86) = 0.944$, $p = .348$.
 - b. left hemisphere: $b = 0.00081$, $SE = 0.00095$, 95% CI [-0.00109, 0.00270], $t(86) = 0.847$, $p = .399$.
2. **Depressive symptoms:** No significant main effect emerged for depression symptoms
 - a. right hemisphere: $b = -0.00042$, $SE = 0.00246$, 95% CI [-0.00447, 0.00532], $t(86) = 0.172$, $p = .864$.
 - b. left hemisphere: $b = 0.00235$, $SE = 0.00246$, 95% CI [-0.00255, 0.00725], $t(86) = 0.952$, $p = .344$;

Parent-reported symptoms: No significant main effects emerged.

2. Amygdala

Child-reported symptoms: No significant main effects emerged.

Parent-reported symptoms:

1. **Anxiety symptoms:**
 - a. A significant main effect emerged for anxiety symptoms, such that as parent-reported anxiety symptoms increased, RPE-modulated BOLD activation decreased in the right amygdala ($b = -0.00286$, $SE = 0.00110$, 95% CI [-0.00505, -0.000685], $t(86) = -2.612$, $p = .011$).
 - b. The main effect of anxiety was marginally significant in the left amygdala ($b = -0.00197$, $SE = 0.00118$, 95% CI [-0.00432, 0.00038], $t(86) = -1.67$, $p = .099$).
2. **Depressive symptoms:** No significant main effect of depression symptoms emerged.
 - a. right hemisphere: $b = -0.00318$, $SE = 0.00268$, 95% CI [-0.00851, 0.00214], $t(86) = -1.190$, $p = .237$.
 - b. left hemisphere: $b = -0.00081$, $SE = 0.00284$, 95% CI [-0.00645, 0.00484], $t(86) = -.284$, $p = .777$).