

Frustration Induction Alters Reward Prediction Error Encoding in Pediatric Irritability

Marta Korom, PhD¹, Diana Burk, PhD², Tahjane Givens, BA¹, Lili Massac, BA¹, Caroline M. Miller, BA³, Yannie D. Lee, BA⁴, Jennifer Meigs, MA⁵, Ellie P. Xu, MA⁶, Hong N.T. Bui, MS⁷, Sofia I. Cárdenas, PhD⁸, Gretchen R. Perhamus, PhD⁹, Michal Clayton, PhD¹⁰, Daniel S. Pine, MD¹, Melissa A. Brotman, PhD¹, Simone P. Haller, PhD¹, Bruno B. Averbeck, PhD², Ellen Leibenluft, PhD¹, & Katharina Kircanski, PhD^{1,6}

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, Ohio University, Athens, OH

⁴ Department of Psychology, University of California at Los Angeles, CA

⁵ Department of Psychology, Catholic University of America, Washington DC

⁶ Department of Psychology, University of Southern California, Los Angeles, CA

⁷ Department of Psychology, University of Maryland, College Park, MD

⁸ School of Medicine, University of California, Irvine, CA

⁹ Department of Psychiatry, The National Crime Victims Research and Treatment Center, Medical University of South Carolina, Charleston, SC

¹⁰ McLean Hospital, Massachusetts General Brigham Department of Psychiatry, Harvard Medical School, Belmont, MA

This work was supported by the National Institute of Mental Health Intramural Research Program projects ZIA-MH002782 and ZIA-MH002928(BA). 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, Massac, Miller, Lee, Meigs, Xu, Bui, Cárdenas, Perhamus, Clayton, 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 families who have continued to dedicate their time for 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.

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Abstract

Background

Irritability is a prevalent, impairing, transdiagnostic affective symptom domain in youth associated with elevated risk for adverse outcomes. Theoretical models of irritability implicate disruptions in reward learning and frustrative nonreward. This is the first study to investigate reward prediction error (RPE) signaling before and after frustration as a neural correlate of irritability.

Methods

We used reinforcement learning modeling to derive individualized trial-by-trial RPEs from behavior during a child-friendly, fMRI-based probabilistic reward learning task in 88 youth ($M_{\text{age}}=14.43$ years, 48.86% female). Irritability severity varied across participants and was characterized dimensionally. Neural encoding of RPEs was examined across prefrontal, ventral striatal/pallidal, and insular regions of interest (ROIs). To test the impact of frustration, RPE-modulated BOLD activity was assessed before and after an acute frustration induction.

Results

Prior to frustration induction, higher irritability was associated with *enhanced* positive coupling between trial-by-trial RPEs and BOLD activity in the ventral striatal/pallidal ROI. Following frustration, this pattern reversed, such that higher irritability was associated with *attenuated* positive coupling between RPEs and BOLD activity in ventral striatal/pallidal and right insular areas, indicating decreased neural encoding of reward outcomes. Irritability was not associated with RPE-modulated orbitofrontal or left insular activity.

Conclusion

Findings provide computationally-grounded evidence linking youth irritability to altered neural RPE encoding during learning and a reversal following task-induced frustration. This emotional state-dependent change in RPE-BOLD coupling suggests that frustration disrupts how reward outcomes shape learning in youth with elevated irritability and highlights a candidate target for interventions addressing frustration reactivity.

Keywords

irritability, reward-prediction error, frustrative non-reward, computational psychiatry, probabilistic reward learning

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Irritability is a transdiagnostic affective symptom domain characterized by elevated proneness to anger in response to frustration (1,2). Clinically impairing irritability affects ~3–5% of community pediatric samples and frequently co-occurs with attention and hyperactivity deficits, as well as anxiety and depressive symptoms (1,3–5). In youth, clinically significant irritability predicts increased risk for suicidality and poorer educational and economic outcomes in adulthood (6,7). Despite the ability to induce frustration translationally (8) and well-characterized behavioral responses (1,2), its neural mechanisms remain poorly understood. Here, we examined RPE signaling before and after frustration as a neural correlate of irritability-related negative affect.

Irritability has been postulated to involve disruptions in reward processing via dysfunction in reward learning, altered RPE signaling, and deficits in threat processing and top-down behavioral control (2). RPE is central to learning from outcomes (9–12), encoding the discrepancy between expected and received outcomes: positive RPEs reflect better-than-expected results, negative RPEs signal worse-than-expected outcomes, and zero RPEs indicate accurate predictions. These signals reinforce adaptive behaviors and discourage those leading to nonreward or aversive outcomes. Behaviorally, they arise from neural computations within reward circuitry, including midbrain dopamine neurons coding RPEs and projecting to the ventral and dorsal striatum, insula, anterior cingulate cortex, and medial prefrontal cortex (10,13,14). In healthy individuals, RPE signaling supports flexible learning through incremental updating of expectations based on feedback (11). Disruptions in reinforcement learning are a transdiagnostic risk factor across ADHD (15), depression (16), and addictive behaviors (17), reflecting impairments in reward processing and decision-making.

Consistent with RPE-based learning mechanisms, youth with high irritability show impairments in flexible learning from feedback when contingencies change (18,19), shifting attention under frustration (18), and suppressing dominant responses, including aggression and temper outbursts (19). These deficits disrupt behavior regulation and goal attainment (20), creating a misalignment between experiences and expectations that heightens negative affect (e.g., anger, frustration) and fuels temper outbursts and verbal and physical aggression (2,18). When maladaptive behaviors such as aggression or tantrums yield desired outcomes (e.g., parents giving in to avoid escalation), they are reinforced, perpetuating maladaptive coping with anger and frustration.

Although behavioral and theoretical models implicate altered reward learning and frustrative nonreward in irritability, RPE encoding has not been directly examined. Prior work has relied on non-learning paradigms (e.g., MID, Go/No-Go), linking irritability to phase-specific alterations in neural activation and connectivity during reward processing, with inconsistencies across studies (18,21–30). During reward anticipation and receipt, studies report enhanced prefrontal and reward-network responses (22,25,27,31) (but see (21)) and heightened prefrontal and parietal connectivity (21,26). During frustrative states or reward omission, irritability has been associated with blunted activation (18,24,26–28) (but see (22,29)) and connectivity in prefrontal and reward networks (22). Post-frustration recovery is less understood: fNIRS data suggest recovery may be particularly challenging (30), and Linke and colleagues (23) found that post-frustration (but not pre-frustration) resting-state global efficiency—a graph theory index of parallel information transfer and network integration—was associated with irritability in three network modules, including the ventral striatum. Together, prior work links irritability to altered neural activation and connectivity during reward processing but has not directly tested RPE encoding as learning unfolds.

Advances in computational modeling offer a principled way to quantify trial-by-trial learning signals not directly observable from behavior (32–34). By fitting choice behavior, these models estimate

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

each participant's evolving expected value and corresponding RPE on every trial—the difference between obtained and expected outcomes (35,36). This is critical for clinical neuroscience because identical feedback can yield different expectancy violations depending on updating, and standard reward-versus-non-reward contrasts can obscure this heterogeneity (37). Model-derived RPEs thus test whether neural activity tracks the magnitude and sign of expectancy violations rather than feedback categories. Modeling also assesses asymmetric learning for positive versus negative outcomes and whether non-reinforcing events (e.g., missed response, “Try Again” trials) shape updating (2), yielding interpretable parameters comparable across contexts such as before versus after frustration (38).

The goal of this study was to use computational psychiatry methods to examine trial-by-trial, person-specific RPE encoding during probabilistic reward learning tasks before and after frustration. Using a transdiagnostic sample and dimensional symptom measures, we first aimed to characterize whole-brain RPE-modulated activation, hypothesizing positive associations between RPEs and BOLD magnitude in ventral striatal/pallidal areas and prefrontal cortices (39,40), and negative associations in the insular cortex (41). We also hypothesized that before frustration, higher irritability would positively correlate with RPE-modulated BOLD activity in ventral striatal/pallidal, and orbitofrontal cortices (OFC) (22,25,27,31), reflecting enhanced positive RPE encoding when outcomes exceed expectations. After frustration, we expected reward encoding to be disrupted in youth with higher irritability, evidenced by blunted RPE modulation in these regions relative to pre-frustration trials, reflecting poor recovery and reduced neural reward signaling during learning (18,23,30,42).

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Methods

Participants and Procedures

Neuroimaging data were collected from 106 participants (51 females; age 8.55–18.57 years, $M=14.10$, $SD=2.63$) at the National Institute of Mental Health (NIMH), recruited via community advertisements. Eighteen participants were excluded ($n=5$ insufficient responding, >25 omitted responses; $n=11$ excessive motion, $>15\%$ censored TRs; $n=1$ incidental finding; $n=1$ poor fieldmap quality). The final sample captured irritability transdiagnostically and dimensionally, including youth with disruptive mood dysregulation disorder ($n=18$), attention deficit hyperactivity disorder ($n=24$), anxiety disorders (generalized, social, separation anxiety disorder; $n=12$), major depressive disorder ($n=20$), and no psychiatric diagnosis ($n=14$). Diagnoses were determined using the Kiddie Schedule for Affective Disorders and Schizophrenia interview (43). Exclusion criteria included estimated $IQ<70$ (matrix reasoning and vocabulary subtests of the Wechsler Abbreviated Scale of Intelligence assessed by a trained research assistant) (44); diagnosis of pervasive developmental, posttraumatic stress, or psychotic disorders; recent substance use (past 3 months); cardinal bipolar symptoms; neurological disorders; or unstable medical conditions. Demographics are summarized in Table 1.

Parent-report clinical measures were completed within three months of imaging. Participants first completed a mock scan and training on each fMRI task phase (10 trials each of “Bell Ring,” “Balloon Pop,” and “Mole Catch,” with 1 rigged trial in the “Mole Catch” training; see “Carnival Paradigm” below). Due to frustration-inducing elements, participants were debriefed after scanning. Parents provided written informed consent, and youth provided assent. Procedures were approved by the NIMH Institutional Review Board, with data collected between 2018 and 2025.

Measures

Irritability

Affective Reactivity Index (ARI): Parent-reported ARI scores (45) provide reliable and clinically informative assessments of observable irritability and functional impairment (45,46). Given documented parent–youth discrepancies—more pronounced in younger children and with elevated irritability (46)—parent-report was used to capture irritability severity. The 6-item ARI assesses temper outbursts and irritable mood over the past six months, yielding a total score of 0–12. In this sample, the ARI showed excellent internal consistency (Cronbach $\alpha=0.9$) and strong test–retest reliability (parent: $r=0.85$) (45,46).

fMRI 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 three 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 (23,29) 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

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

(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 Figure 1 for a depiction of the Carnival Task.

Figure 1

Overview of the Carnival instrumental learning task.

[FIGURE_CARNIVAL_1.jpeg]

Note. Regressors are displayed at the top. The frustration induction run is shown but was not included in analyses. Affect ratings were collected three times, at the end of each task phase.

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 (36,47). 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 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):

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

$$\begin{aligned} v_i(t_{TA}) &= \gamma * v_i(t_{TA}) \\ v_i(t+1) &= v_i(t) + \alpha_{pos} * \delta(t) \quad \text{if } R = 1 \\ v_i(t+1) &= v_i(t) + \alpha_{neg} * \delta(t) \quad \text{if } R = 0 \end{aligned}$$

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 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.

Parameter Inference. Parameters were estimated using hierarchical Bayesian inference implemented in Julia with [BAT.jl](#) (48). 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 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 $p(D, \theta) = p(D|\theta)p(\theta)$ denote the joint likelihood and prior. We first identified the maximum a posteriori (MAP) estimate $\hat{\theta}$ using density maximization over posterior samples. The log model evidence was then computed as:

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

where d is the number of free parameters, and $|Q|$ is the determinant of the posterior covariance matrix. We then chose the model with the highest evidence value (see Supplementary Section 1, Table 1). We also examined model fit via posterior predictive checks to ensure that simulated behavior resembled the observed data.

Parameter Recovery. To validate the identifiability of the selected model, we conducted a parameter recovery analysis. Synthetic pre- and post-frustration datasets were generated using parameter values from the original fits. Because the γ parameter showed low variance in both pre- and post-model fits, its values were sampled across both parameter sets. Simulation data included 20–50 blocks of trials and Try Again trials at the same proportion as the clinical data. Each model (pre and post) was then fit to the simulated data using the same Bayesian framework. Correlations between recovered parameters and the original fitted parameters were computed to assess recovery fidelity (see Supplementary Section 1, Table 2, and Figure 1).

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Neuroimaging Data analysis

Detailed description of the imaging data acquisition is available in Supplementary Section 2.

Person-level imaging data preprocessing

Amplitude-modulation analyses in AFNI (v25.1.15) modeled trial-by-trial neural RPE encoding, allowing examination of variation in RPE alongside BOLD response. Amplitude modulation coefficients reflect percentage BOLD signal change per unit increase in RPE, indexing the strength of this association. Event duration varied across trials and was modeled using duration modulation. As hypotheses focused on RPE-modulated activation during pre- and post-frustration reward-learning blocks, the frustration induction run was not modeled. See Supplementary Section 3 for details.

Group-level analysis

Preliminary whole brain analyses. Three separate whole-brain voxel-wise analyses were performed. In the first analysis, one-sample t-test (AFNI's 3dttest++) was conducted to assess whether mean RPE-modulated BOLD activation, averaged across pre- and post-frustration trials, differed from zero. In the second analysis, a paired subjects t-test (AFNI's 3dttest++) was used to evaluate within-person differences in RPE-modulated activation between pre- and post-frustration trials. In a third exploratory analysis, AFNI's 3dMVM was used to examine associations between irritability and RPE-modulated activation, with Phase (pre vs. post frustration) modeled as a two-level within-person factor and irritability included as a continuous between-person predictor. Multiple-comparison correction was carried out using Monte Carlo simulations with AFNI's 3dClustSim. More details available in Supplementary Section 4.

Region-of-interest (ROI) analyses. Primary analyses evaluated whether RPE-modulated activation in the post-minus-pre frustration contrast in three bilateral ROIs predicted irritability (ARI). Mean beta estimates for this contrast were extracted from participants' first-level glt models and analyzed via regression in R (version 2025.05.1+513). We used the Julich-Brain probabilistic atlas (50) to create separate left and right hemisphere striatum/pallidum, insula, and OFC masks, by adding together the maximum probability versions of 7 areas for the bilateral OFC (areas Fo1-7), 16 regions for the insula (areas Id1 - Id10, Ia1 - Ia3, Ig1 - Ig3), and 4 areas for the ventral striatum (the fundus of the caudate nucleus and the putamen, medial and lateral accumbens), combined with the ventral pallidum, and binarizing the results. All masks were eroded (1 voxel) and intersected with a group mask that retained the voxels that were present for at least 90% of all participants. Figure 2 details the extent of the masks.

Figure 2

Masks used in the primary ROI analyses.

[FIGURE_MASK_2.jpeg]

Note. Left (blue) and right (red) hemisphere ventral striatal/pallidal region, insula and orbitofrontal cortex (OFC) masks used in the primary ROI analyses. These masks were created using the Julich-Brain 3D probabilistic atlas and are overlaid on the MNI 2009 atlas.

Results

Demographic information

The analytic sample included 88 participants. The average age was 14.43 years (SD=2.52), average IQ was 115.84 (SD=14.66), and 43 participants (48.86%) were female. For demographics, see Table 1; for diagnostic group differences, Supplementary Section 5; for included vs. excluded comparisons, Supplementary Section 6; and for zero-order correlations among demographics, IQ, modeling parameters, and brain outcomes, Supplementary Section 8.

Behavior during the Carnival task

Participants successfully learned the reward contingencies in both pre- and post-frustration phases, as indicated by a progressive increase in choosing the higher-value option (Figure 3). Group-averaged performance did not differ significantly between phases, confirmed by Bonferroni-corrected one-sample t-tests on pre–post choice differences ($p > .05$).

The reinforcement learning model tracked observed behavior in both phases (Figure 3). Model comparison showed that the Rescorla–Wagner model with separate positive and negative learning rates and a Try-Again trial discounting parameter ($TA\gamma + 2LR$; Model 4) best fit the data. Parameters were estimated separately for pre- and post-frustration trials. Paired comparisons of posterior mean estimates revealed significant reductions in the negative learning rate (α_{neg}) and Try-Again discounting (γ) after frustration (Table 2), whereas the positive learning rate (α_{pos}) and the softmax inverse temperature (β) showed no significant pre-to-post changes. These results indicate preserved average task performance across phases, alongside phase-dependent changes in latent learning parameters.

At the group level, there were no statistically significant changes in behavioral performance from pre- to post-task (one-sample t-tests on pre–post differences, Bonferroni-corrected). In contrast, model-based analyses revealed parameter-specific changes following the frustrative reward task. Paired t-tests comparing pre- and post-task parameter estimates demonstrated significant reductions in the negative learning rate ($t(87)=-2.90$, $p<.01$) and TA gamma ($t(87)=-94.39$, $p<.0001$). No significant pre–post differences were observed for the positive learning rate or softmax inverse temperature ($p>.05$). Details on missing responses during the task are available in Supplementary Section 7.

Figure 3

Mean probability of choosing the higher-value option across trials during the pre-frustration (left) and post-frustration (right) phases of the Carnival task.

[FIGURE_MPROB_3.jpeg]

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Note. Solid lines depict observed group-averaged behavioral data, and dashed lines depict predictions from the best-fitting RW-RL model (model 4). $\pm$ SEM across participants. Behavioral performance did not differ significantly between task phases. Model-based parameter estimates were derived separately for pre- and post-frustration trials and are summarized in Table 2.

Self-reported affect across task phases

Self-reported frustration, unhappiness, and excitement did not differ between pre- and post-frustration phases, but frustration and unhappiness increased and excitement decreased significantly during the frustration phase (see Table 3). Parent-reported ARI was not significantly associated with self-reported frustration ratings during the frustration task run ($t(86)=0.352$, $p=.726$).

Preliminary whole brain analyses

The initial group-level analysis, consistent with prior studies of human RPE encoding (39–41), revealed significant modulation of regional activity by RPE in 34 positive and 10 negative clusters: positive correlations in the ventral striatum, posterior cingulate, medial and orbital prefrontal cortex, and negative correlations in the bilateral insula and anterior cingulate (Figure 4). Cluster details are in Supplementary Section 9; results of the second and third exploratory whole-brain analyses are in Supplementary Section 10.

Figure 4

Whole-brain RPE modulated average BOLD activation results across all participants and across pre- and post-frustration trials.

[FIGURE_WBRPE_4.jpeg]

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

Ventral striatal/pallidal region

A significant main effect of the RPE-modulated BOLD activation difference between the post- and pre-frustration trials on irritability emerged: elevated irritability was associated with a greater decrease in RPE-modulated BOLD activation from pre- to post-frustration in both the right and left ventral striatal/pallidal region (left hemisphere: $\beta=-5.87$, 95% CI [-9.21, -2.53], $t(86)=-3.50$, $p<.001$; right hemisphere: $\beta=-5.79$, 95% CI [-9.17, -2.41], $t(86)=-3.41$, $p<.001$). Follow-up analyses examined the pre- and post-frustration runs separately. During pre-frustration trials, higher irritability was associated with higher RPE-modulated activation in the ventral striatal/pallidal region (left hemisphere: $\beta=3.72$, 95% CI [-1.28, 8.72], $t(86)=1.48$, $p=.143$; right hemisphere: $\beta=5.17$, 95% CI [0.20, 10.14], $t(86)=2.07$, $p=.042$). In contrast, during post-frustration trials, higher irritability was associated with lower RPE-modulated activation (right hemisphere: $\beta=-6.40$, 95% CI [-11.29, -1.51], $t(86)=-2.60$, $p=.01$). Including age and sex as covariates did not change the significance or direction of the effects. Figure 5 details these effects.

Exploratory analyses examined separate bilateral ROIs for the ventral striatum, dorsal striatum, and the ventral pallidum. The results were consistent with the combined ROI findings and are available in Supplementary Section 11.

Figure 5

[FIGURE_STRIATUMSCATTER_5.jpeg]

RPE-modulated BOLD activation associations with irritability from pre- to post-frustration runs in bilateral ventral striatal/pallidal ROIs.

Insula and the OFC

No significant association emerged between pre- to post-frustration RPE-modulated BOLD magnitude difference and irritability in the insula and the OFC. Follow-up analyses revealed that a significant negative association emerged between RPE-modulated brain activation in the right insula and irritability during the post-frustration trials. See Supplementary Section 12 for detailed results and visualization of these effects.

Discussion

This transdiagnostic study investigated RPE signaling before and after frustration induction as a correlate of pediatric irritability. Frustration was induced in the scanner through a separate task involving rigged feedback. This manipulation built on prior experimental work in irritability (23,29,30) and successfully increased youth's subjective frustration. As hypothesized, parent-reported irritability interacted with RPE-modulated BOLD activation in the bilateral ventral striatal/pallidal regions, with higher irritability associated with enhanced RPE-modulated activation during pre-frustration trials and attenuated RPE-modulated activation during post-frustration trials. Notably, this negative post-frustration association was stronger than the positive pre-frustration association, indicating that frustration profoundly alters reward prediction processes in youth with elevated irritability. Expected findings generally were not shown in the left insula or bilateral OFC, but a significant negative post-frustration effect was observed in the right insula.

The use of computational modeling to derive trial-by-trial RPEs represents a key aspect of this study. By estimating person-specific RPEs derived from hierarchical Bayesian modeling and using these values to parametrically modulate the BOLD response, we were able to directly link computational learning signals to their neural correlates in the reward network, insular, and prefrontal cortices, and associate it with irritability symptoms pre- and post-frustration in key regions of the reward network. This approach moves beyond canonical task-based contrasts by leveraging within-person expectancy violations (both positive and negative) to characterize how the magnitude of neural responses scales with deviations from expected outcomes. To our knowledge, this study represents the first application of a computational psychiatry framework combining trial-by-trial RPE estimation and neural amplitude modulation to the study of youth irritability, providing a mechanistic account of altered reward learning processes that may underlie transdiagnostic behavioral dysfunction.

Our results are broadly consistent with prior work demonstrating increased neural reward sensitivity and fronto-striatal activation and connectivity during reward receipt during non-frustrated affective states (21,22,25,26). Our findings further support prior evidence linking irritability to blunted reward signaling post-frustration (18,22,24,26,42) and poor recovery from frustration (23,30). Some studies, however, have reported null associations with irritability during pre-frustration trials (18,29), or successful recovery during post-frustration (27). While inconsistent with the current results, these differences may reflect distinct analytic approaches (e.g., examining average BOLD activation), nature of the tasks (e.g., non-learning), or neuroimaging modality used. In particular, no prior work has captured the computational signals that drive value updates, and their associations with variations in BOLD signal magnitude. Here, we showed that irritability is associated with altered encoding of expectancy violations within the reward circuitry, revealing a state-dependent shift from heightened RPE signaling in non-frustrated states to blunted signaling following frustrative nonreward that is likely masked in average-activation analyses.

The ventral striatum is a core hub of RPE encoding, making it a coherent locus for irritability-related modulation of RPE signaling (40). In contrast, the OFC is more implicated in value representation and salience processing than direct RPE encoding, which may explain the absence of robust OFC effects. This difference may reflect coding patterns: single neurons in the ventral striatum correlate positively with RPEs, whereas OFC populations show mixed positive and negative encoding (51). The isolated post-frustration association in the right insula requires replication. Irritability was not associated with behavioral learning parameters, possibly due to task characteristics. The Carnival task's fixed 80:20

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

reward ratio produces rapid learning and limited variability, consistent with a ceiling effect, whereas prior studies often used more cognitively demanding paradigms such as reversal learning tasks. Heightened RPE-modulated activation in the ventral striatal/pallidal region during pre-frustration trials may reflect increased neural reward sensitivity and approach motivation in irritable youth. In contrast, blunted post-frustration RPE signaling in these regions may indicate reduced response to learning signals, limiting recovery from frustration. Thus, rewarding and non-rewarding outcomes may be less differentiated after frustration despite intact behavioral performance. These findings suggest that in non-frustrated states, youth with higher irritability show amplified responses to positive expectancy violations, which may contribute to extreme behaviors and emotional volatility when rewards fail to materialize. Following frustrating events, they exhibit blunted reward expectancy updating and reduced learning from positive outcomes, hindering recovery. Although speculative, increased network efficiency through parallel information flow post-frustration, as shown by Linke and colleagues (23), may also contribute to these frustration-state dependent effects. Together, this state-dependent dampening of adaptive value updating may reflect a neurobiological vulnerability for affective dysregulation, increasing persistent negative affect and externalizing problems.

The present study has important implications for future designs. Our findings underscore the need to consider affective context in fMRI studies and behavioral interventions for youth with irritability, as reward processing differs between non-frustrated and post-frustration states. Future work should replicate and extend these findings in larger samples by testing pre- to post-intervention changes in reward processing. Examining associations between depressive and anxious symptoms and RPE-modulated BOLD activation will further clarify how internalizing symptoms are shaped by reward updating and expectancy violations.

There are several limitations to our study. First, its cross-sectional design provides initial insights into brain function but cannot establish causality or developmental trajectories in the associations between irritability and RPE encoding, for which longitudinal studies are needed. Second, we relied on a single parent-reported assessment of irritability, which may introduce rater bias or limited observation. Strengths include the computational modeling approach to the behavioral data, the transdiagnostic focus, and the wide developmental age range of the sample, enhancing the generalizability of these findings.

Conclusions

We demonstrated that frustrative nonreward altered neural encoding of RPEs within canonical reward circuitry in youths with varying irritability. Irritability was associated with heightened RPE-modulated BOLD activation at baseline and marked attenuation of RPE updating following frustration in the ventral striatal/pallidal region and right insula, despite intact behavioral performance. These effects were not observed in the orbitofrontal cortex or left insula. Modeling trial-by-trial RPEs revealed a state-dependent shift in neural reward sensitivity across affective contexts that may underlie frustration intolerance in irritable youth. Repeated blocked goals or omitted rewards may reduce the reinforcing value of subsequent rewards, contributing to persistent negative affect and maladaptive behavioral responses. These findings underscore the importance of considering affective states when evaluating reward learning in irritability.

References

1. Leibenluft E. Severe Mood Dysregulation, Irritability, and the Diagnostic Boundaries of Bipolar Disorder in Youths. *Am J Psychiatry*. 2011 Feb;168(2):129–42.
2. Brotman MA, Kircanski K, Stringaris A, Pine DS, Leibenluft E. Irritability in Youths: A Translational Model. *Am J Psychiatry*. 2017 Jun 1;174(6):520–32.
3. Brotman MA, Schmajuk M, Rich BA, Dickstein DP, Guyer AE, Costello EJ, et al. Prevalence, Clinical Correlates, and Longitudinal Course of Severe Mood Dysregulation in Children. *Biol Psychiatry*. 2006 Nov;60(9):991–7.
4. Stoddard J, Stringaris A, Brotman MA, Montville D, Pine DS, Leibenluft E. Irritability in Child and Adolescent Anxiety Disorders. *Depress Anxiety*. 2014 Jul;31(7):566–73.
5. Stringaris A, Maughan B, Goodman R. What's in a Disruptive Disorder? Temperamental Antecedents of Oppositional Defiant Disorder: Findings from the Avon Longitudinal Study. *J Am Acad Child Adolesc Psychiatry*. 2010 May;49(5):474–83.
6. Stringaris A, Cohen P, Pine DS, Leibenluft E. Adult Outcomes of Youth Irritability: A 20-Year Prospective Community-Based Study. *Am J Psychiatry*. 2009 Sep;166(9):1048–54.
7. Pickles A, Aglan A, Collishaw S, Messer J, Rutter M, Maughan B. Predictors of suicidality across the life span: The Isle of Wight study. *Psychol Med*. 2010 Sep;40(9):1453–66.
8. Naik AA, Ma X, Munyeshyaka M, Leibenluft E, Li Z. A New Behavioral Paradigm for Frustrative Non-reward Reveals a Global Change in Brain Networks by Frustration. *bioRxiv*. 2023 Jun 6;2023.02.28.530477.
9. Lerner TN, Holloway AL, Seiler JL. Dopamine, Updated: Reward Prediction Error and Beyond. *Curr Opin Neurobiol*. 2021 Apr;67:123–30.
10. Schultz W. Dopamine reward prediction error coding. *Dialogues Clin Neurosci*. 2016 Mar 31;18(1):23–32.
11. Schultz W. The Dopamine Reward Prediction Error Signal. *SSRN Electron J* [Internet]. 2024 [cited 2026 Jan 22]; Available from: <https://www.ssrn.com/abstract=4816666>
12. Tobler PN, Fiorillo CD, Schultz W. Adaptive Coding of Reward Value by Dopamine Neurons. *Science*. 2005 Mar 11;307(5715):1642–5.
13. Garrison J, Erdeniz B, Done J. Prediction error in reinforcement learning: A meta-analysis of neuroimaging studies. *Neurosci Biobehav Rev*. 2013 Aug;37(7):1297–310.
14. D'Ardenne K, McClure SM, Nystrom LE, Cohen JD. BOLD Responses Reflecting Dopaminergic Signals in the Human Ventral Tegmental Area. *Science*. 2008 Feb 29;319(5867):1264–7.
15. Aster HC, Waltmann M, Busch A, Romanos M, Gamer M, Maria van Noort B, et al. Impaired flexible reward learning in ADHD patients is associated with blunted reinforcement

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

sensitivity and neural signals in ventral striatum and parietal cortex. *NeuroImage Clin.* 2024 Mar 1;42:103588.

16. 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.

17. Liebenow B, Jones R, DiMarco E, Trattner JD, Humphries J, Sands LP, et al. Computational reinforcement learning, reward (and punishment), and dopamine in psychiatric disorders. *Front Psychiatry* [Internet]. 2022 Oct 20 [cited 2026 Jan 22];13. Available from: <https://www.frontiersin.org/journals/psychiatry/articles/10.3389/fpsyt.2022.886297/full>

18. Deveney CM, Connolly ME, Haring CT, Bones BL, Reynolds RC, Kim P, et al. Neural Mechanisms of Frustration in Chronically Irritable Children. *Am J Psychiatry.* 2013 Oct 1;170(10):1186–94.

19. Haller SP, Khosravi P, Kim KY, Brotman MA. Targeting inhibitory control in youth irritability: challenges and opportunities. *Neuropsychopharmacology.* 2026 Jan;51(1):230–3.

20. Cardinale EM, Bezek J, Siegal O, Freitag GF, Subar A, Khosravi P, et al. Multivariate Assessment of Inhibitory Control in Youth: Links With Psychopathology and Brain Function. *Psychol Sci.* 2024 Apr 1;35(4):376–89.

21. Parker AJ, Walker JC, Takarae Y, Dougherty LR, Wiggins JL. Neural mechanisms of reward processing in preadolescent irritability: Insights from the ABCD study. *J Affect Disord.* 2025 Feb;370:286–98.

22. Kryza-Lacombe M, Palumbo D, Wakschlag LS, Dougherty LR, Wiggins JL. Executive functioning moderates neural mechanisms of irritability during reward processing in youth. *Psychiatry Res Neuroimaging.* 2022 Jul;323:111483.

23. Linke JO, Haller SP, Xu EP, Nguyen LT, Chue AE, Botz-Zapp C, et al. Persistent Frustration-Induced Reconfigurations of Brain Networks Predict Individual Differences in Irritability. *J Am Acad Child Adolesc Psychiatry.* 2023 Jun;62(6):684–95.

24. Perlman SB, Jones BM, Wakschlag LS, Axelson D, Birmaher B, Phillips ML. Neural substrates of child irritability in typically developing and psychiatric populations. *Dev Cogn Neurosci.* 2015 Jul 20;14:71–80.

25. Kessel EM, Dougherty LR, Kujawa A, Hajcak G, Carlson GA, Klein DN. Longitudinal Associations Between Preschool Disruptive Mood Dysregulation Disorder Symptoms and Neural Reactivity to Monetary Reward During Preadolescence. *J Child Adolesc Psychopharmacol.* 2016 Mar;26(2):131–7.

26. Dougherty LR, Schwartz KTG, Kryza-Lacombe M, Weisberg J, Spechler PA, Wiggins JL. Preschool- and School-Age Irritability Predict Reward-Related Brain Function. *J Am Acad Child Adolesc Psychiatry.* 2018 Jun;57(6):407–417.e2.

27. Gatzke-Kopp LM, Willner CJ, Jetha MK, Abenavoli RM, DuPuis D, Segalowitz SJ. How does reactivity to frustrative non-reward increase risk for externalizing symptoms? *Int J Psychophysiol.* 2015 Nov 1;98(2, Part 2):300–9.

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

28. Adleman NE, Kayser R, Dickstein D, Blair RJR, Pine D, Leibenluft E. Neural correlates of reversal learning in severe mood dysregulation and pediatric bipolar disorder. *J Am Acad Child Adolesc Psychiatry*. 2011 Nov;50(11):1173-1185.e2.
29. Tseng WL, Deveney CM, Stoddard J, Kircanski K, Frackman AE, Yi JY, et al. Brain mechanisms of attention orienting following frustration: Associations with irritability and age in youth. *Am J Psychiatry*. 2019 Jan 1;176(1):67-76.
30. Quiñones-Camacho LE, Fishburn FA, Camacho MC, Hlutkowsky CO, Huppert TJ, Wakschlag LS, et al. Parent-child neural synchrony: a novel approach to elucidating dyadic correlates of preschool irritability. *J Child Psychol Psychiatry*. 2020 Nov;61(11):1213-23.
31. Hawes SW, Waller R, Byrd AL, Bjork JM, Dick AS, Sutherland MT, et al. Reward Processing in Children With Disruptive Behavior Disorders and Callous-Unemotional Traits in the ABCD Study. *Am J Psychiatry*. 2021 Apr 1;178(4):333-42.
32. Daw ND. Trial-by-trial data analysis using computational models.
33. Montague PR, Dolan RJ, Friston KJ, Dayan P. Computational psychiatry. *Trends Cogn Sci*. 2012 Jan 1;16(1):72-80.
34. Sutton RS, Barto AG. Reinforcement Learning: An Introduction [Internet]. Cambridge, MA, USA: MIT Press; 1998 [cited 2026 Feb 8]. 344 p. (Bach F, editor. Adaptive Computation and Machine Learning series). Available from: <https://mitpress.mit.edu/9780262193986/reinforcement-learning/>
35. O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. Temporal difference models and reward-related learning in the human brain. *Neuron*. 2003 Apr 24;38(2):329-37.
36. Rescola RA, Wagner AR. A Theory of Pavlovian Conditioning Variations in the Effectiveness of Reinforcement and Nonreinforcement. In A. H. Black, & W. F. Prokasy (Eds.), *Classical Conditioning II Current Research and Theory* (pp. 64-99). New York: Appleton-Century-Crofts.;
37. Huys QJM, Maia TV, Frank MJ. Computational psychiatry as a bridge from neuroscience to clinical applications. *Nat Neurosci*. 2016 Mar;19(3):404-13.
38. Leibenluft E. Pediatric Irritability: A Systems Neuroscience Approach. *Trends Cogn Sci*. 2017 Apr;21(4):277-89.
39. O'Doherty J, Kringelbach ML, Rolls ET, Hornak J, Andrews C. Abstract reward and punishment representations in the human orbitofrontal cortex. *Nat Neurosci*. 2001 Jan;4(1):95-102.
40. Seymour B, Daw N, Dayan P, Singer T, Dolan R. Differential Encoding of Losses and Gains in the Human Striatum. *J Neurosci*. 2007 May 2;27(18):4826-31.
41. Preuschoff K, Quartz SR, Bossaerts P. Human Insula Activation Reflects Risk Prediction Errors As Well As Risk. *J Neurosci*. 2008 Mar 12;28(11):2745-52.

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

42. Adleman NE, Kayser R, Dickstein D, Blair RJR, Pine D, Leibenluft E. Neural correlates of reversal learning in severe mood dysregulation and pediatric bipolar disorder. *J Am Acad Child Adolesc Psychiatry*. 2011 Nov;50(11):1173-1185.e2.
43. Kaufman J, Birmaher B, Brent D, Rao U, Flynn C, Moreci P, et al. Schedule for Affective Disorders and Schizophrenia for School-Age Children-Present and Lifetime Version. 1997;
44. Wechsler D. Wechsler Abbreviated Scale of Intelligence--Second Edition. 2011;
45. Stringaris A, Goodman R, Ferdinando S, Razdan V, Muhrer E, Leibenluft E, et al. The Affective Reactivity Index: a concise irritability scale for clinical and research settings. *J Child Psychol Psychiatry*. 2012;53(11):1109–17.
46. Mallidi A, Meza-Cervera T, Kircanski K, Stringaris A, Brotman MA, Pine DS, et al. Robust caregiver-youth discrepancies in irritability ratings on the affective reactivity index: An investigation of its origins. *J Affect Disord*. 2023 Jul;332:185–93.
47. Hull CL. Principles of behavior: an introduction to behavior theory. Oxford, England: Appleton-Century; 1943. x, 422 p. (Principles of behavior: an introduction to behavior theory).
48. Schulz O, Beaujean F, Caldwell A, Grunwald C, Hafych V, Kröninger K, et al. BAT.jl: A Julia-Based Tool for Bayesian Inference. *SN Comput Sci*. 2021 Apr 12;2(3):210.
49. Tierney L, Kadane JB. Accurate Approximations for Posterior Moments and Marginal Densities. *J Am Stat Assoc*. 1986 Mar;81(393):82–6. doi:10.1080/01621459.1986.10478240
50. 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.
51. Tang JCY, Paixao V, Carvalho F, Silva A, Klaus A, da Silva JA, et al. Dynamic behaviour restructuring mediates dopamine-dependent credit assignment. *Nature*. 2024 Feb;626(7999):583–92.

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Tables

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%)
Parent-reported ARI score	3.432 (3.32)
IQ (n = 70)	115.843 (14.66)
Motion in the scanner ¹	0.12 (0.07)

¹Calculated as the mean Euclidean distance of framewise volume shift after censoring.

Table 2

RW-RL Model Parameters

Best Model Variable	Mean Value (Pre)	S.E.M. (Pre)	Mean Value (Post)	S.E.M. (Post)	T-test result (N=88)
Positive learning rate: α_{pos}	0.39	0.022	0.39	0.022	t(87) = -0.03, p = .975
Negative learning rate: α_{neg}	0.19	0.016	0.14	0.014	t(87) = -2.90, p < .01
Softmax beta: β	8.90	0.374	9.61	0.477	t(87) = 1.56, p = .123
TA gamma: γ	0.88	0.0006	0.79	0.0007	t(87) = -94.39, p < .0001

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Table 3

Self-reported affective ratings across task phases.

Task Phase	Excited (M, SD)	Frustrated (M, SD)	Unhappy (M, SD)
Pre-frustration	3.67 (1.70)	2.32 (1.37)	2.19 (1.28)
Frustration	2.50 (1.53)	4.51 (2.33)	3.34 (2.01)
Post-frustration	3.59 (1.78)	2.28 (1.4)	2.30 (1.44)
Statistical comparisons (Paired-sample t-tests)			
Pre-frustration vs. Frustration	t(87)= 6.561 , CI 95% [0.815, 1.523], $p < .001$	t(87)= -10.176, CI 95% [-2.618, -1.762], $p < .001$	t(87)= -6.621, CI 95% [-1.498, -0.807], $p < .001$
Frustration vs. Post-frustration	t(87)= -6.347, CI 95% [-1.433, -0.749], $p < .001$	t(87)= 9.71, CI 95% [1.771, 2.683], $p < .001$	t(87)= 6.313, CI 95% [0.714, 1.37], $p < .001$
Pre-frustration vs. Post-frustration	t(87)= 0.558, CI 95% [-0.199, 0.354], $p = .578$	t(87)= 0.294, CI 95% [-0.212, 0.286], $p = .769$	t(87)= -1.211, CI 95% [-0.293, 0.071], $p = .229$

Section 1

Computational Model Comparison

Supplementary Table 1

Model evidence values computed at the group level

Model	Model evidence logZ (Pre)	Model evidence logZ (Post)
(1) Vanilla RL (RW)	-2325.6	-2335.2
(2) Two Learning Rates (2LR)	-2164.2	-2348.5
(3) Try Again Trial Gamma (TA γ)	-2277.2	-1987.4
(4) TA γ + 2LR	-2024.9	-1921.5

Parameter Recovery Results

Supplementary Table 2

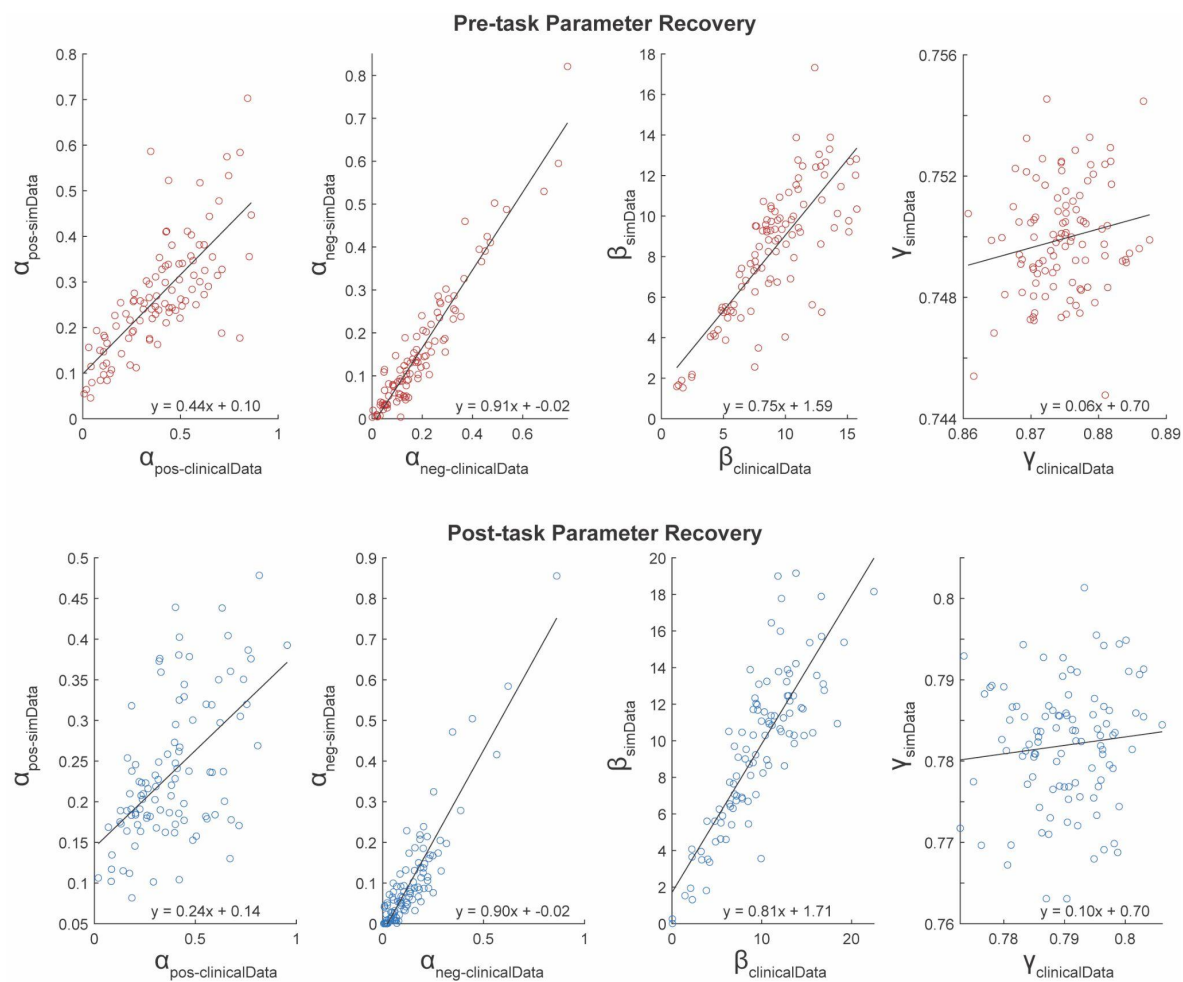
Correlations between recovered parameters and original parameters

Parameter	Correlation (Pre)	Correlation (Post)
Positive learning rate:	0.74	0.55
Negative learning rate:	0.95	0.91
Softmax beta: b	0.79	0.83
TA gamma:	0.18	0.10

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Supplementary Figure 1

Parameter values for simulated data (20-50 blocks, n=88) vs. clinical data (3 blocks, n=88).



Supplementary Section 2

Imaging data acquisition

MRI data were acquired using a 3-Tesla GE DISCOVERY MR750 MRI scanner with a 32-channel headcoil. A sagittal localizer scan preceded an automated higher-order shim to calibrate the magnetic field to reduce signal dropout, particularly in the prefrontal cortices. Seven functional runs (3 pre-frustration, 1 frustration induction, and 3 post-frustration runs) were acquired using 40 interleaved axial slices that provided whole-brain coverage except for the cerebellum. Seventy-three volumes for pre- and post-frustration runs and 183 volumes for the frustration induction run were collected. Scans used a 96×96 matrix with echo planar single shot gradient echo T2* weighing with the following specifications: repetition time, TR = 2300ms; echo time, TE = 27ms; flip angle, FA = 70°; field of view, FOV = 240×240 mm, yielding 2.5mm isotropic voxels. The first four initial images from each run were discarded to ensure the scanner had reached longitudinal magnetization equilibrium. For co-registration with the functional data, a high resolution structural MPRAGE scan was acquired with the following specifications: 1 mm interleaved sagittal slices, repetition time, TR = 7.7ms; echo time, TE = 3.48ms; flip angle, FA = 7 degree; field of view, FOV = 176×256 mm; matrix size = 176×256 mm.

Supplementary Section 3***Person-level imaging data preprocessing***

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 of the pre- and post-frustration runs: 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.

One generalized linear test (GLT), contrasting post-frustration with pre-frustration feedback events modulated by reward prediction error, was specified at the person level to facilitate subsequent ROI analyses.

Supplementary Section 4

Group-level analysis multiple comparisons correction

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 6.46 mm (ACF parameters: $a=0.68$, $b=2.53$, $c=6.87$). 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 11 or more voxels, which served as the correction for the whole-brain models.

Section 5

Diagnostic Group Differences in Demographic Characteristics, IQ, Irritability, 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, $F(4, 65) = 0.51, p = .73$.

Irritability: As expected, based on the nature of diagnoses, there was a significant main effect of diagnosis on parent-reported ARI scores, $F(4, 83) = 16.24, p < .001$. Post hoc comparisons indicated that participants in the DMDD group had significantly higher irritability scores than those in the ADHD, anxiety, MDD, and healthy volunteers (all $p < .001$) groups. Additionally, participants in the ADHD group had higher irritability scores than healthy volunteers ($p = .015$), and those in the MDD group had higher irritability scores than healthy volunteers ($p = .041$). No other significant pairwise comparisons emerged (all $ps > .15$). Mean irritability scores ($\pm$ SD) were 3.12 ± 2.85 for ADHD, 7.33 ± 2.70 for DMDD, 2.90 ± 2.59 for MDD, 2.67 ± 2.77 for the anxiety group, and 0.36 ± 1.08 for healthy volunteers.

Model Parameters: No significant diagnostic group differences emerged for the model parameters during the pre- (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: $F(4, 83) = 1.087, p = .368$) and post-frustration trials (positive learning rate: $F(4, 83) = 0.381, p = .822$; negative learning rate: $F(4, 83) = 1.284, p = .283$; TA gamma: $F(4, 83) = .385, p = .833$; softmax beta: $F(4, 83) = 1.643, p = .171$). 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: $t(86) = 1.15, p = .253$), and parent-reported ARI scores (positive learning rate: $F(1, 86) = 1.122, p = .292$; negative learning rate: $F(1, 86) = 0.289, p = .592$; TA gamma: $F(1, 86) = 0.147, p = .902$; softmax beta: $F(1, 86) = 1.283, p = .261$) were not significantly associated with any of the model parameters. IQ, however, was 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 6

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, $t(18.514) = 0.09$, $p = .93$), or ARI scores (mean in: excluded group: 3.722, included group: 3.432, $t(23.842) = 0.327$, $p = .747$). 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 7

Missing responses during the task

The number of total missed responses were not significantly associated with parent-reported ARI scores ($\beta = 0.238$, 95% CI [-0.05, 0.53], $t(86) = 1.642$, $p = .104$), 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).

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Section 8

Supplementary Table 3

Zero-order Pearson correlations between pairs of demographic variables, irritability scores, brain measures, and RW-RL model parameters.

	Age (years)	ARI ¹	IQ	Avg. motion per TR	L ² VSP ³ , pre	L VSP, post	R ⁴ VSP, pre	R VSP, post	L Insula, pre	L Insula, post	R Insula, pre	R Insula, post	L OFC ⁵ , pre	L OFC, post	R OFC, pre	R OFC, post	Beta, pre	Beta, post	Pos. LR ⁶ , pre	Pos. LR, post	Neg. LR ⁷ , pre	Neg. LR, post	TA gamma , pre	TA gamma , post
Age (years)	1																							
ARI	-.37 ***	1																						
IQ	-.003	-.137	1																					
Avg. motion per TR	-.39 ***	.368 ***	-.135	1																				
L VSP, pre	.107	.157	-.323 **	.179	1																			
L VSP, post	.084	-.353 ***	-.075	-.166	-.034	1																		
R VSP, pre	.001	.217 *	-.136	.151	.776 ***	-.016	1																	
R VSP, post	.136	-.27 *	-.064	-.142	-.133	.795 ***	.001	1																
L Insula, pre	.148	.075	-.257 *	.000	.473 ***	.044	.409 ***	.08	1															
L Insula, post	.183	-.13	.002	-.073	-.152	.554 ***	-.056	.563 ***	.057	1														
R Insula, pre	.088	.038	-.213	.045	.458 ***	.045	.517 ***	.078	.781 ***	.079	1													
R Insula, post	.121	-.218 *	.027	-.012	-.237 *	.624 ***	-.152	.653 ***	.025	.754 ***	.103	1												
L OFC, pre	-.039	.010	-.162	-.052	.496 ***	.098	.526 ***	.024	.432 ***	-.146	.382 ***	-.099	1											
L OFC, post	.154	-.171	-.062	-.147	-.154	.467 ***	-.068	.533 ***	.137	.453 ***	.16	.493 ***	.036	1										
R OFC, pre	-.071	.018	-.058	.211 *	.336 **	-.065	.434 ***	-.055	.207	-.176	.288 **	-.058	.635 ***	-.007	1									
R OFC, post	.05	-.121	.125	-.017	-.324 **	.457 ***	-.128	.567 ***	.01	.313 **	.039	.515 ***	.032	.735 ***	.095	1								
Beta, pre	.121	-.121	.266*	-.287**	0	.064	-.032	.081	.068	.008	-.076	.152	.103	.117	.047	.143	1							
Beta, post	.165	-.045	-.018	-.186	.058	.096	.068	.051	.064	.117	.010	.217*	.177	.228*	.052	.071	.418***	1						
LR, pre	.102	-.114	-.051	-.064	.094	-.021	.029	-.086	.113	-.019	.098	-.007	.049	.074	-.038	-.039	-.232*	.232*	1					
LR, post	.145	-.033	.160	-.018	.076	-.097	.090	-.070	.116	-.044	.139	-.076	.026	-.022	.029	-.074	.054	-.040	.021	1				
Neg. LR, pre	-.027	-.058	-.239*	.011	.213*	.025	.236*	.007	.023	-.075	.065	-.030	.220*	-.106	.092	-.004	-.151	.025	-.074	-.136	1			
Neg. LR, post	.044	-.015	-.030	.023	.023	-.004	.144	.061	-.060	.267*	.121	.199	-.028	.076	.091	.030	-.115	.146	-.004	.110	.263*	1		
TA Gamma, pre	-.067	-.041	.208	-.109	-.125	.029	.016	.128	-.070	-.037	-.065	.058	.048	-.113	.156	.015	.018	-.163	-.051	.038	-.105	-.158	1	
TA Gamma, post	-.091	.097	-.005	.073	.285**	-.076	.250*	-.096	.191	-.028	.194	-.139	0	-.111	-.061	-.181	.012	-.229*	-.216*	.029	.043	-.092	.063	1

¹ARI - Affective Reactivity Index, parent report; ²L - left hemisphere; ³VSP - ventral striatal/pallidal area; ⁴R - right hemisphere; ⁵OFC - orbitofrontal cortex; ⁶Pos. LR - positive learning rate; ⁷Neg. LR - negative learning rate; *** p < .001; ** p < .01; * p < .05

Section 9

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Supplementary Table 4

Summary of clusters showing a significant RPE modulated average BOLD activation across all participants and across pre- and post-frustration trials.¹

ROI overlap and location	Cluster size	Coordinates (center of mass)			Coordinates (at peak)			Mean F	SEM	Max Int
	k	CM RL	CM AP	CM IS	Peak x	Peak y	Peak z			
R occipital cortex, calcarine sulcus (area hOc1, 28.7%; 974 h0Oc2, 26.8%);		-22.8	92.9	0.1	-27.5	94.2	8	4.9217	0.0369	9.3
R lingual gyrus, 14.9%										
L frontal I (GapMap), 22.3%, L frontal pole (13.5%)	951	9.1	-49.2	14	0	-55.8	-7	4.2014	0.0204	7.277
L frontal to occipital (GapMap), 68.2%; R frontal to occipital (GapMap), 25.8%	809	4.2	52.5	23.6	7.5	56.8	13	4.7007	0.0385	8.4705
L lingual gyrus, 23.6%; L V2, 20.2%; L occipital lobe (Area h0c4lp), 16.6%; L calcarine sulcus, 13.7%	635	20.7	94.6	-4.8	12.5	91.8	-12	4.5901	0.0353	7.8101
R frontal I (GapMap), 59.3%; R frontal to occipital (GapMap), 22.8%	364	-3.4	-28.2	38.6	-2.5	-23.2	40.5	-4.2567	0.0364	-6.4243
R insula, 19.2%; R frontal operculum 16.5%; R inferior frontal gyrus 13.3%	321	-39.7	-21.6	-0.9	-35	-25.8	-7	-4.5683	0.0473	-6.9797
L inferior parietal lobule, 41.8%; (area PF); L inferior parietal lobule, 41% (area PFt)	231	61.2	30.8	35.3	65	31.8	35.5	4.1525	0.0398	6.4201
L inferior parietal lobule (area PGp), 77.4%	220	44.8	72.2	32.9	45	74.2	33	4.2567	0.0415	6.4917
L insula (area Id8, 23.3%; Id7, 14%; Id6 13.4%); L frontal to temporal I (GapMap), 21%; L frontal operculum, 15.8%;	219	35.4	-21.6	-1	32.5	-28.2	0.5	-4.5364	0.0579	-7.4832
R frontal to occipital (GapMap), 82.7%; R superior parietal lobule (area 7P, 9.8%; 7M 6.3%)	181	-10.9	70	39.8	-17.5	64.2	33	-4.1955	0.0397	-6.0542
R postcentral sulcus, 26.9%; R superior parietal lobule 16.6%	165	-30.7	32.4	51.3	-37.5	29.2	43	3.9534	0.0359	6.021

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

L frontal I (GapMap), 47.4%; L inferior frontal gurus, 46%	161	49.6	-34.5	11	50	-38.2	5.5	4.2455	0.0521	6.7816
R frontal to occipital (GapMap), 42.1%; R mesial superior frontal gyrus, 38.4%	120	-4.4	13.1	49.1	-5	16.8	55.5	3.8545	0.0316	5.0459
R superior temporal sulcus (area STS2,, 47.9%; STS1, 42.5%)	106	-53.1	24.6	-7.2	-50	26.8	-4.5	-4.686	0.0899	-7.0552
L temporal to parietal (GapMap), 83.1%; L fusiform gyrus (15.7%)	101	55.7	53.4	-9	57.5	54.2	-7	4.1387	0.059	5.7749
L parahippocampal gyrus, 48%; L hippocampus (CA1), 13.3%	95	29.3	37.3	-12.6	32.5	36.8	-14.5	4.1824	0.0685	6.5546
R inferior parietal lobule, 73.1%; R postcentral sulcus, 16.2%	92	-55.8	21.5	43.9	-50	21.8	45.5	4.0667	0.0492	5.2986
R inferior parietal lobule, 53.6%; R superior temporal gyrus, 22.2%	90	-55.6	26.6	19	-47.5	29.2	20.5	3.9157	0.0387	5.1174
R intraparietal sulcus, (area hIP1 47.5%; hIP3 38.9%)	90	-37	52.9	44.2	-40	49.2	43	-4.1425	0.0592	-5.6117
L basal forebrain, 17.9%; L amygdala, 17.5%;	89	18.7	0.3	-12.8	12.5	-8.2	-9.5	4.2279	0.0597	5.7828
L frontal to temporal II (GapMap), 16. 1%; L basal ganglia (fundus of putamen, 14.8%; lateral accumbens, 8.3%)										
L orbitofrontal cortex, 80.2%; L frontal to temporal 14.8%	62	37.1	-35.8	-13.5	37.5	-35.8	-14.5	4.3911	0.1008	6.3537
R parietal operculum, 37.3%; R frontal operculum (area Op5, 27.1%; Op6, 18.4%)	60	-54.1	0.8	7.3	-52.5	1.8	8	3.9584	0.06	5.3249
R inferior parietal lobule, 70.8%; R temporal to parietal (GapMap), 26.3%	59	-56.3	49.2	20.7	-55	49.2	20.5	-3.9838	0.0645	-5.4791
L inferior frontal gyrus, 95.3%; R superior parietal lobule, 81.1%;	48	51.4	-8.8	22.7	55	-10.8	20.5	3.8913	0.0584	5.3307
R postcentral gyrus, 56.6%; R frontal II (GapMap), 25.9%	37	-13.3	30.9	46.4	-12.5	29.2	45.5	3.8713	0.0633	5.1602
R frontal operculum, 46.6%; R insula (area Id4 44%; Id6, 8%)	37	-33.8	29.7	69.4	-30	34.2	73	3.8094	0.0575	4.7297
R inferior parietal lobule, 84.3%; R temporal to parietal (GapMap), 15.7%	34	-39	-0.7	12.8	-40	1.8	15.5	3.9734	0.0815	4.9902
	31	-49.7	69.2	29.9	-47.5	71.8	35.5	3.7135	0.0441	4.3404

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

R basal ganglia (ventral pallidum, 24.3%; lateral accumbens, 8.3%; fundus of putamen, 9.3%) R basal forebrain, 22.7%; R amygdala, 12%	30	-16.9	-3	-10.8	-12.5	-8.2	-9.5	3.9621	0.0887	5.576
R superior parietal lobule (area 5L, 47.3%; 7A 33.1%; 7PC, 19.3%)	25	-22.5	50.9	70.7	-20	51.8	75.5	3.7837	0.0731	5.1253
R postcentral gyrus (area 3b, 50.9%; 4a, 30.6%; 4p, 16.6%)	22	-39.5	12.8	47.1	-40	11.8	48	3.8211	0.0859	4.7898
R orbitofrontal cortex, 69.6%; R frontal to temporal I (GapMap), 27.9%	21	-37.4	-35.1	-13.8	-35	-35.8	-14.5	3.8801	0.0818	4.9157
R putamen, 82% ²	21	-28.4	10	10.2	-27.5	14.2	10.5	3.9675	0.0777	4.8505
L insula (Id4, 48%; Id6, 18.2%); L frontal operculum (Op5, 27.1%; Op6, 6.7%)	21	37.9	1.3	13.6	42.5	-0.8	10.5	3.9807	0.1045	5.0953
L temporal to parietal (GapMap), 54.9%; L superior temporal sulcus, 45.1%	20	64.4	10.9	-17.3	65	11.8	-17	3.8961	0.0915	4.5493
L frontal I (GapMap), 78.1%; L middle frontal gyrus (area MFG2, 12%; MFG1, 9.6%)	19	32.7	-54.7	12.5	32.5	-55.8	10.5	-3.9417	0.0936	-4.7467
R superior frontal sulcus (area SFS1, 79.6%; SFS2, 19%)	19	-26.9	-55.7	19.1	-22.5	-50.8	18	-3.8392	0.09	-4.8944
R superior parietal lobule, 90.4%; R intraparietal sulcus, 9%	15	-33.5	46	64.6	-35	44.2	63	3.8605	0.0819	4.5236
R anterior cingulate cortex, 27.1%; L basal forebrain, 20%; R basal forebrain, 18.8%; R subgenual anterior cingulate cortex, 9.4%; L anterior cingulate cortex, 9.4%	14	0	-8.8	-8.3	0	-8.2	-9.5	3.8015	0.1115	5.031
R hippocampus (CA3, 50.6%; DG, 33.5%; CA2, 12%)	13	-27.4	16.9	-18.5	-30	16.8	-19.5	4.262	0.1413	4.9774
R frontal to occipital (GapMap), 92.5%	13	-2.5	-1.7	38.3	0	-3.2	35.5	3.7146	0.0607	4.0806
L inferior parietal sulcus (area hIP3, 95.4%; hIP1, 4.6%)	13	29.4	53.1	41.1	30	54.2	40.5	-3.8648	0.1424	-4.9888
R inferior parietal lobule (area PFcm, 62.8%; PF, 35.2%; PFm, 1.4%)	12	-57.5	29.5	32.8	-57.5	31.8	33	3.6247	0.0417	3.8754
R lingual gyrus, 71.9%; R fusiform gyrus, 23.1%	11	-25.6	68.2	-5.8	-27.5	69.2	-4.5	3.9967	0.115	4.8091

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

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 11 voxels, and are defined using NN = 2 connectivity.

² 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.

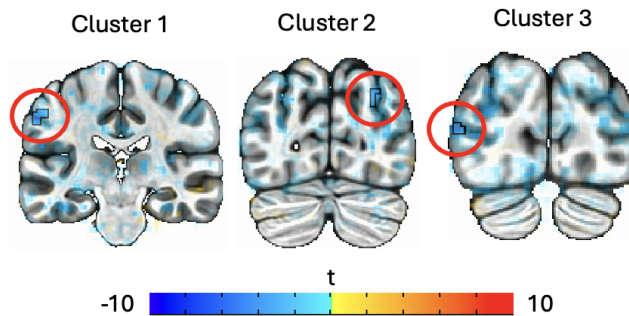
FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Section 10

Paired t-test of RPE-modulated activation differences between pre- and post-frustration trials

The three clusters in which significant RPE-modulated activation differences emerged between pre- and post-frustration trials are shown in Supplementary Figure 5. The results indicate that, across all three clusters, the magnitude of the modulated BOLD signal was significantly smaller during post-frustration trials compared to pre-frustration trials.

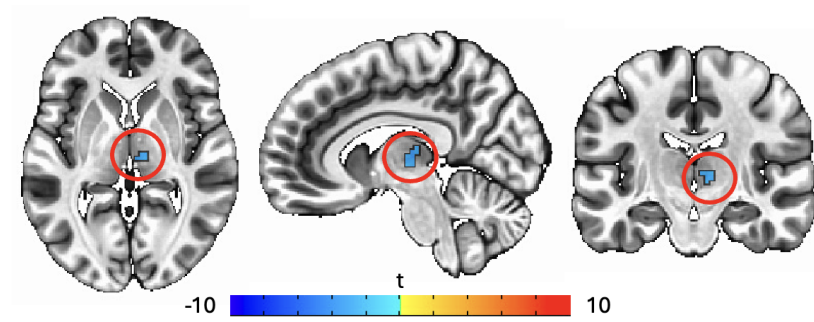
Cluster #	ROI Overlap and Location	Cluster size, k	CM RL	CM AP	CM IS	Peak x	Peak y	Peak z	Mean	SEM	Max Int
1	L inferior parietal lobule, 78.5%; L postcentral sulcus, 21.5%	18	57.2	25	42	57.5	24.2	40.5	4.020	0.141	5.413
2	R parieto-occipital sulcus, 37.9%; R inferior parietal lobule (34.8% hIP7; 24.2% hIP8; 3% hIP5)	17	-27.7	77.2	36.8	-30	76.8	40.5	3.859	0.077	4.476
3	L occipital cortex lateral posterior area (95.3% 41p; 1.7% 41a); L lingual gyrus, 2.9%	16	38.4	90.9	7.2	45	86.8	8	3.729	0.084	4.562



Exploratory whole-brain analysis of irritability predicting within-person changes in RPE-modulated activation (pre- vs. post-frustration)

A significant negative association was observed between irritability and the change in RPE-modulated activation from pre- to post-frustration in a single cluster in the right medial thalamus proper (cluster size, $k = 19$; CM RL = -6.9, CM AP = 18.9, CM IS = 5; Peak $x = -7.5$, Peak $y = 19.2$, Peak $z = 5.5$; Mean $F = -3.971$, SEM = 0.101; Max Int = -4.895). These results suggest that higher irritability was associated with a greater decrease in RPE-modulated activation change from pre- to post-frustration in the right thalamus.

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY



Supplementary Section 11

Detailed results of the exploratory regression analyses examining the association between RPE-modulated brain activation difference from post to pre (post minus pre) and parent-reported irritability scores at subregions of the combined ventral striatal/pallidal ROI. The results were consistent with the combined ROI findings, with the exception of left ventral pallidum, wherein the post-minus-pre RPE-modulated BOLD activation was not significantly associated with irritability ($p = .115$), but became statistically significant when age was included in the model as a co-variate of no-interest ($p = .045$).

Hemisphere / Region	Intercept	Unstandardized Coefficient (β)	Std. Error	t-value	p-value
Left					
Ventral striatum ¹	3.236	-3.312	1.574	-2.105	.038
Dorsal striatum ²	3.171	-5.734	1.690	-3.393	.001
Ventral pallidum	3.378	-1.951	1.225	-1.593	.115
Ventral pallidum + age as covariate ³	3.368	-2.314	1.1375	-2.034	.045
Right					
Ventral striatum	3.361	-4.586	1.358	-3.376	.001
Dorsal striatum	3.264	-5.184	1.675	-3.095	.003
Ventral pallidum	3.367	-5.258	1.424	-3.692	<.001

Note. The dependent variable was parent-reported ARI across all models. ¹The ventral striatum included the lateral and medial accumbens; ²The dorsal striatum included the fundus of the putamen and fundus of the caudate nucleus; ³Age was mean centered.

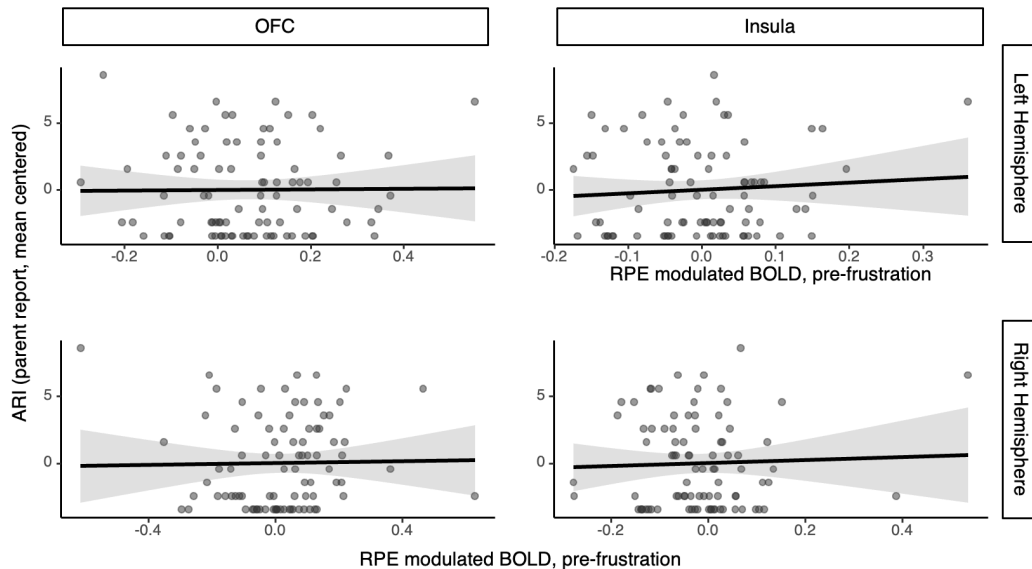
FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

Section 12

Insula

No significant association emerged between pre- to post-frustration RPE-modulated BOLD magnitude difference and irritability (ARI) in the insula (left hemisphere: $\beta = -2.88$, 95% CI [-6.84, 1.08], $t(86) = -1.44$, $p = .15$; right hemisphere: $\beta = -4.08$, 95% CI [-8.39, 0.22], $t(86) = -1.89$, $p = .06$). Follow-up analyses revealed that the association between RPE-modulated brain activation during the pre-frustration trials was not significantly associated with irritability (right hemisphere: $\beta = 2.71$, 95% CI [-5.04, 10.47], $t(86) = 0.70$, $p = .49$; left hemisphere: $\beta = 1.12$, 95% CI [-5.15, 7.40], $t(86) = 0.36$, $p = .72$). During the post-frustration runs, a significant negative association emerged for the right hemisphere ($\beta = -5.66$, 95% CI [-11.09, -0.23], $t(86) = -2.07$, $p = .041$) but not the left ($\beta = -2.74$, 95% CI [-7.22, 1.74], $t(86) = -1.22$, $p = .23$). See Supplementary Section 9 for visualization of these effects.

RPE-modulated BOLD activation associations with irritability during pre-frustration runs in bilateral OFC and Insula.

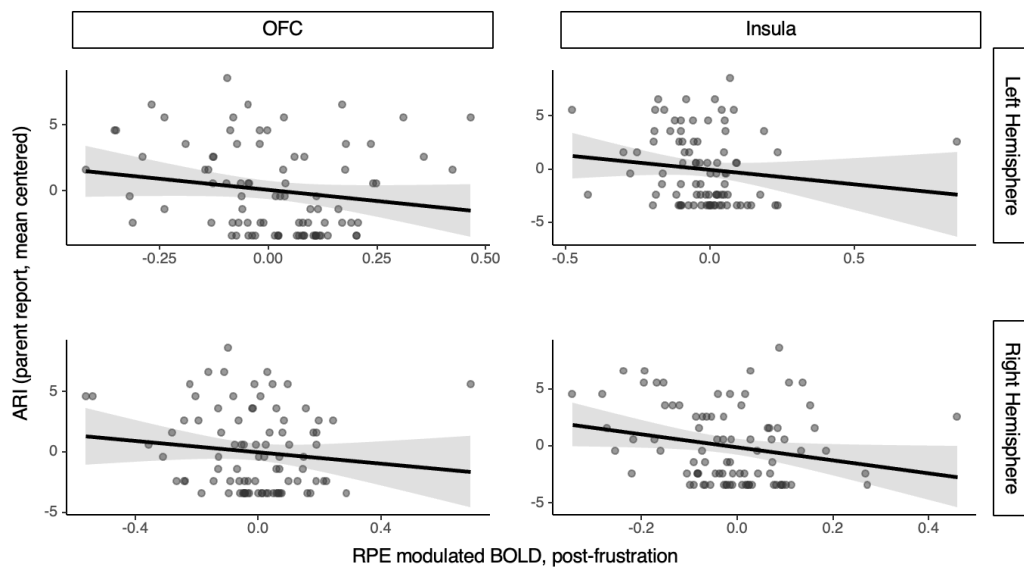


OFC

No significant association emerged between irritability and pre- to post-frustration RPE-modulated BOLD magnitude difference in the OFC (post-minus-pre contrast: left hemisphere: $\beta = -2.08$, 95% CI [-5.31, 1.14], $t(86) = -1.28$, $p = .20$; right hemisphere: $\beta = -1.49$, 95% CI [-4.54, 1.56], $t(86) = -0.97$, $p = .34$; pre-frustration: left hemisphere: $\beta = 0.22$, 95% CI [-4.62, 5.06], $t(86) = 0.09$, $p = .93$; right hemisphere: $\beta = 0.35$, 95% CI [-3.81, 4.51], $t(86) = 0.17$, $p = .87$; post-frustration: left hemisphere: $\beta = -3.39$, 95% CI [-7.59, 0.80], $t(86) = -1.61$, $p = .11$; right hemisphere: $\beta = -2.32$, 95% CI [-6.39, 1.75], $t(86) = -1.13$, $p = .26$). See Supplementary Section 9 for visualization of these effects.

FRUSTRATION ALTERS RPE ENCODING IN IRRITABILITY

RPE-modulated BOLD activation associations with irritability during post-frustration runs in bilateral OFC and Insula.



Supplementary Figure 6

The difference between post- and pre-frustration runs in RPE-modulated BOLD activation associations with irritability in bilateral OFC and Insula.

