

Research article

Developmental differences in reward-learning and its connection to resting-state functional connectivity modeled using a hierarchical Bayesian model

Zsofia Karlocai^{a,b,1,*}, Johan Vegelius^{a,c,1}, Ebba Widegren^a, Johan Lundin Kleberg^d, David Fällmar^e, Nils B. Kroemer^{b,f,g,h}, Malin Gingnell^{a,2}, Andreas Frick^{a,c,2}

^a Department of Medical Sciences, Experimental Cognitive and Affective Neuroscience Lab, Uppsala University, Uppsala, Sweden

^b Department of Psychiatry & Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Germany

^c Department of Psychology, Uppsala University, Uppsala, Sweden

^d Department of Psychology, Stockholm University, Stockholm, Sweden

^e Department of Surgical Sciences, Neuroradiology, Uppsala University Hospital, Uppsala, Sweden

^f Section of Medical Psychology, Department of Psychiatry & Psychotherapy, Faculty of Medicine, University of Bonn, Germany

^g German Center for Mental Health (DZPG), partner site Tübingen, Germany

^h German Center for Diabetes Research (DZD), Neuherberg, Germany



ARTICLE INFO

Keywords:

Cognitive development
Resting-state functional connectivity
Bayesian model
fMRI
WLSL strategy

ABSTRACT

Adolescence is a period of heightened sensation-seeking, risk-taking, and reward sensitivity, characterized by structural and functional changes in the brain. Developmental changes in functional connectivity between cortical and subcortical regions may refine communication within reward-related circuitry, influencing learning and decision-making. Here, we compared reinforcement learning behavior and its relationship to resting-state functional connectivity in reward-related circuits in adolescents and adults. Fifty-eight healthy participants (32 adolescents aged 13–16; 26 adults aged 30–40) completed a probabilistic two-armed bandit task and resting-state functional magnetic resonance imaging (fMRI). The learning-related parameters learning rate (α) and inverse temperature (β , an index of the randomness of choices) and their relationship to functional connectivity were modeled from behavioral data using Q learning in a hierarchical Bayesian framework. In the whole sample, learning rate was associated with functional connectivity in several cortico-subcortical pathways, particularly involving the anterior cingulate cortex. Adolescents exhibited lower learning rate and inverse temperature values than adults and had a stronger association between learning rate and fronto-striatal connectivity. Adolescents also showed less tendency to stay with winning options in the task, defined as the proportion of trials where participants repeated the previous choice after a reward. These findings highlight the involvement of the ACC in reward learning and indicate that behavior in a reinforcement learning context is characterized by reduced feedback-driven learning and more variable choice behavior or greater exploration in adolescents compared to adults, and suggest that adolescents rely more on fronto-striatal connectivity during learning.

1. Introduction

Adolescence is a sensitive developmental period, characterized by marked neural and behavioral changes [1,2]. These changes drive independence, exploration, and the refinement of preferences and

behaviors through trial and error. As adolescents evaluate the outcomes of their actions, they develop decision-making strategies that shape future behavior. Like many other functions, learning-related processes undergo substantial changes during this period, leading to distinct decision-making patterns compared to adults [3]. While adolescence is

* Correspondence to: Department of Psychiatry & Psychotherapy, Tübingen Center for Mental Health, University of Tübingen, Calwerstr. 14, Tübingen 72076, Germany.

E-mail address: zsofia.karlocai@uni-tuebingen.de (Z. Karlocai).

¹ Shared first authorship

² Joint shared author

often linked to increased risk-taking and impulsivity, these tendencies are also adaptive and contribute to cognitive and social maturation [4]. Traits like sensation-seeking, peer influence, and novelty exploration are typical and evolutionarily beneficial, facilitating autonomy and experiential learning [5]. To investigate these processes, researchers have applied reinforcement learning models to study how adolescents process and integrate feedback and adjust their decisions [6].

From a reinforcement learning perspective, decision-making can be understood through the exploration-exploitation framework, which examines the cognitive processes underlying the choice to either explore unfamiliar options or exploit familiar ones. These processes are essential for learning and reward acquisition, enabling individuals to navigate dynamic environments [7]. In reinforcement learning terms, exploration can be divided into random and directed forms. Random exploration reflects stochastic variability in decision making, whereas directed exploration is purposeful, driven by uncertainty, and aimed at acquiring information about unknown options. The present study examines developmental differences in value-guided choice behavior, within a probabilistic reinforcement context, which can include variability often associated with random exploration [8]. Developmental studies suggest that adolescents exhibit a stronger tendency toward exploration compared to adults [9,10]. This preference aligns with normative developmental trajectories that emphasize autonomy through experimentation and adaptive risk-taking [11]. While exploration may not always lead to immediate optimal outcomes, it enhances long-term adaptability, particularly in uncertain or volatile environments [12, 13]. However, increased exploration can be counterproductive in very stable environments, such as a two-option task where one choice consistently yields a higher reward probability than the other, focusing on exploiting known rewarding options proves more advantageous for optimizing outcomes [14].

Reinforcement learning behavior is often modeled using computational models such as Q learning. These models allow researchers to estimate latent variables that influence behavior but cannot be directly observed [15]. One key estimate is the learning rate, which reflects how individuals integrate feedback to guide decision-making [6]. A higher learning rate means that decisions are updated quickly based on recent feedback. In contrast, a lower learning rate means that people adjust their expectations more gradually, taking into account a longer history of feedback. Another important estimate is inverse temperature, which represents randomness of the choice. High values of inverse temperature indicate deterministic decisions exploiting the most optimal learned choice, whereas low inverse temperature indicates that choices are more stochastic and the individual explores more. Thus, inverse temperature is often interpreted as reflecting the trade-off between exploration and exploitation [16], although it may also capture broader influences on choice variability [6]. Developmental studies indicate that adolescents exhibit lower learning rates and are slower in updating expected values than adults [10,17,18]. Similarly, inverse temperature tends to rise with age, indicating more variable choice behavior or exploratory behavior in younger individuals [3,18–20].

Repeated decision-making tasks often rely on reward behavior strategies, such as the win-stay, lose-switch strategy, where a win increases the likelihood of repeating a choice, while a loss encourages switching [21]. Some studies suggest that children switch more frequently after wins compared to adults, with adolescents performing at an intermediate level [22]. Similarly, other research has identified a positive correlation between age and the tendency to stay with a winning option [23]. Together, these findings suggest that feedback-based decision-making processes remain underdeveloped in children and adolescents but improve with age. These developmental advancements may reflect the asynchronous maturation of brain systems involved in reward processing, with subcortical regions such as the striatum showing heightened reactivity to rewarding outcomes during adolescence, while prefrontal regions that support cognitive control and regulatory processes mature more gradually [24,25].

Recent neurodevelopmental studies have demonstrated that resting-state functional connectivity within and between key reward-related networks undergo significant reorganization during adolescence [26]. These connectivity changes, particularly within the fronto-striatal and salience networks, are thought to support the maturation of reward processing, behavioral adaptability, and reinforcement learning [27, 28]. These networks provide the functional architecture for adaptive decision making: the fronto-striatal network integrates dopaminergic reward signals and cognitive control, while the salience network, anchored in the anterior cingulate cortex (ACC), regulates the dynamic allocation of attention to motivationally relevant stimuli [29,30].

Together, these systems converge within the broader meso-corticolimbic reward circuit, which includes structures such as the striatum, comprised of the nucleus accumbens (NAcc), putamen and caudate, key parts of the striatum, and the ventral tegmental area (VTA), a primary dopaminergic midbrain region, together form the core of the mesolimbic reward pathway that mediates motivation, reinforcement learning, and reward-related signaling across the lifespan is central to reward processing [31]. Cortical components, including the ventromedial prefrontal cortex (vmPFC), orbitofrontal cortex (OFC), and anterior cingulate cortex (ACC), contribute to reasoning, regulatory processes, and adaptive decision-making. The prefrontal cortex undergoes significant changes during adolescence, reflecting age-related (intrinsic) maturation rather than experience-based learning-specific processes [32].

Functional connectivity between brain regions develops substantially during adolescence [33,34] and can be studied using resting-state functional magnetic resonance imaging (rs-fMRI). The strength of this resting-state functional connectivity (rsFC) is considered a stable predictor of behavioral indices, such as measures of decision-making, risk-taking and impulsivity, and serves as a neural correlate of reward-guided decision-making [35]. Moreover, connectivity strength may reveal mechanisms underlying developmental changes in reward learning. Choi and colleagues [36] investigated developmental changes in cortico-striatal connectivity and found that early in life, there is strong connectivity within subcortical regions, and as individuals age, connectivity between the caudate and putamen (dorsal striatum) and cortical regions increase. Adolescents exhibit heightened sensitivity to rewards a phenomenon that is thought to stem from an imbalance between the developmental trajectories of the striatum and the prefrontal cortex [24]. This imbalance results in an overactive reward-related circuitry (bottom-up system) and an immature prefrontal cortex, limiting top-down regulatory control over motivational stimuli. Increasing connectivity between cortical and subcortical regions during adolescent development into adulthood enhances top-down control of behavior [37]. This gradual maturation of connectivity underscores the interplay between cortical and subcortical systems in shaping reward processing and decision-making strategies during adolescence.

1.1. Aims

Despite extensive research on the neural correlates of reward-learning, there is a dearth of studies examining the association between reward learning parameters and functional connectivity development. In this study, we aim to bridge this gap by investigating the relationship between learning rate, inverse temperature, and rsFC using data from adolescents and adults who underwent rs-fMRI and a two-armed bandit task.

We implement a hierarchical Bayesian model where we investigate behavioral differences as well as associations between behavioral parameters and rsFC. First, we test the hypothesis that connectivity between cortical and subcortical regions will be associated with learning parameters (i.e., learning rate and inverse temperature). Second, we use maximum a posteriori estimates of learning rates and inverse temperature from the Bayesian model in a multivariate regression analysis to test the hypothesis that adolescents differ from adults in their relationship

between learning parameters and rsFC. Third, we investigate age-related differences in behavior using both the hierarchical Bayesian model and the win-stay lose-switch strategies, which are quantified using choices from the two-armed bandit task. The win-stay strategy is defined as the proportion of trials in which participants repeat their previous choice after a reward, while the lose-switch strategy reflects the proportion of trials in which participants switch choices following no reward. Specifically, we expect adolescents to show a more variable choice behavior in decision-making, reflected in lower inverse temperature values, and a reduced capacity to integrate feedback, indicated by lower learning rates and lower proportions of win-stay and lose-shift.

By integrating computational modeling and neuroimaging techniques, this study seeks to provide novel insights into the developmental changes in reward learning and its neural underpinnings from adolescence to adulthood. Understanding these mechanisms has important implications for identifying factors that contribute to adaptive and maladaptive behaviors in youth and may inform strategies for promoting positive developmental outcomes.

2. Methods

2.1. Participants

Participants were recruited between September 2019 and June 2022 through existing longitudinal studies at the Department of Psychology, Uppsala University, as well as public advertisements. Recruitment has been described in detail in a previous publication [38] and results from a partly overlapping sample of the adolescent group including behavioral data from the same reward learning task has been published [39]. Adolescents (13–17 years) were included to capture the period of active structural and functional maturation of the prefrontal cortex and its connectivity with subcortical reward regions. Adults (30–40 years) were chosen to represent a group in which prefrontal development is fully complete and neural systems are stable well beyond the late-20s plateau in cortical maturation [24]. Selecting adults above 30 therefore, ensured a clear developmental contrast between individuals undergoing neuro-cognitive maturation and those with mature circuitry. Exclusion criteria comprised hearing or uncorrected visual impairments, severe psychiatric or neurological conditions, somatic illnesses that could affect results, atypical development, ongoing medication use, pregnancy, menopause, or contraindications for MRI. Adult participants provided written informed consent, while adolescents provided written assent alongside parental written consent. The Swedish Ethical Review Authority approved the study. Participation was reimbursed with a gift card worth \$85.

2.2. Procedure

Potential participants, and parents of adolescent participants, received a link to a webpage with detailed study information. Upon agreeing to participate, they completed an online consent form and eligibility screening. Eligible participants were contacted by phone and underwent the Mini International Neuropsychiatric Interview (MINI; [40]) or MINI-KID with a trained clinician, the instrument (MINI or MINI-KID) was selected depending on participant age.

Participants then completed online questionnaires and attended two in-person sessions. The first session, held in a lab setting, included a series of cognitive and emotional tasks not included here. The second session, conducted at Uppsala University Hospital, involved MRI scanning and a subsequent 2-Bandit reward learning task outside the scanner. The total MRI scan time was approximately 45 min.

2.3. Acquisition of imaging data

During the MRI session, magnetic resonance imaging (MRI) was performed using a Philips Achieva 3.0 T whole body MR-scanner

(Philips Medical Systems, Best, The Netherlands) equipped with a 32-channel head-coil. An anatomical T1-weighted image (echo time (TE) = 3.8 ms; repetition time (TR) = 8.2 ms; inversion time = 685.5 ms; flip angle = 8°; field of view = 240 × 240 mm; isotropic 1 mm voxels; 220 contiguous slices) was used for anatomical referencing. Resting-state fMRI was conducted for 7 min (210 volumes) using a blood-oxygenation-level dependent T2*-weighed echo-planar imaging (EPI) sequence (TR = 2000 ms; TE = 30 ms; flip angle: 90°; slice thickness = 3 mm; slice spacing = 3.9 mm; slices number = 32, interleaved acquisition). Participants were reminded not to focus their thoughts on anything and to keep their eyes open. Morphological images from all subjects were reviewed by a senior consultant in neuroradiology (DF) to exclude malformations and significant parenchymal defects. Here, no such images were excluded.

2.4. Resting-state functional connectivity analysis

2.4.1. Preprocessing

Functional and anatomical data were preprocessed in CONN version 22a [41] using a modular preprocessing pipeline [42] including realignment with correction of susceptibility distortion interactions, slice timing correction, outlier detection, direct segmentation and Fig. 3-space normalization, and smoothing. Functional data were realigned using SPM realign & unwarp procedure [43], where all scans were co-registered to a reference image (first scan of the first session) using a least squares approach and a 6 parameter (rigid body) transformation [44], and resampled using b-spline interpolation to correct for motion and magnetic susceptibility interactions. Temporal misalignment between different slices of the functional data (acquired in interleaved Philips order) was corrected following SPM slice-timing correction (STC) procedure [45,46] using sinc temporal interpolation to resample each slice BOLD timeseries to a common mid-acquisition time. Potential outlier scans were identified using ART [47] as acquisitions with framewise displacement above 0.9 mm or global BOLD signal changes above 5 standard deviations [48,49]. A reference BOLD image was computed for each participant by averaging all non-outlier BOLD volumes across the time series. Participants were excluded if maximum framewise displacement exceeded 2.4 mm, mean framewise displacement exceeded 0.3 mm, or if fewer than 80 % of scans remained valid after outlier removal. Functional and anatomical data were normalized into standard MNI space, segmented into grey matter, white matter, and CSF tissue classes, and resampled to 2 mm isotropic voxels following a direct normalization procedure [48,50] using SPM unified segmentation and normalization algorithm [51,52] with the default Ixi-549 tissue probability map template. Last, functional data were smoothed using spatial convolution with a Gaussian kernel of 8 mm full width half maximum (FWHM).

2.4.2. Denoising

In addition, functional data were denoised using a standard denoising pipeline [53] including the regression of potential confounding effects characterized by white matter timeseries (5 CompCor noise components), CSF timeseries (5 CompCor noise components), motion parameters and their first order derivatives (12 regressors) [54], outlier volumes (below 135 regressors) [40], session effects and their first order derivatives (2 regressors), and linear trends (2 regressors) within each functional run, followed by bandpass frequency filtering of the BOLD timeseries [55] between 0.008 Hz and 0.09 Hz. CompCor [56,57] noise components within white matter and CSF were estimated by computing the average BOLD signal as well as the largest principal components orthogonal to the BOLD average, motion parameters, and outlier volumes within each subject's eroded segmentation masks.

2.4.3. First-level analysis

We focused on 7 regions of interest (ROI; [Supplementary Material Figure S3](#)): medial prefrontal cortex (mPFC), cingulate gyrus anterior

division (ACC), frontal orbital cortex (OFC), caudate (Cau), putamen (Put), nucleus accumbens (NAcc), ventral tegmental area (VTA). All cortical and subcortical ROIs except for the VTA were defined using the Harvard-Oxford atlas provided with the CONN toolbox. The VTA was defined using the probabilistic mask from Trutti et al. [58], thresholded at 50 % to ensure anatomical specificity. ROI-to-ROI connectivity matrices were estimated characterizing the functional connectivity between each pair of ROIs. Functional connectivity strength was represented by Fisher-transformed bivariate correlation coefficients from a general linear model (weighted-GLM [59]), estimated separately for each pair of ROIs, characterizing the association between their BOLD signal timeseries. In order to compensate for possible transient magnetization effects at the beginning of each run, individual scans were weighted by a step function convolved with an SPM canonical hemodynamic response function and rectified.

2.5. Reward-learning experiment

Behavioral reward-learning was assessed using a two-armed Bandit task where participants performed 60 trials where they chose between two stimuli, triangles with different orientations presented with a uniform gray background (Fig. 1). One of the stimuli had an 85 % probability of winning (+1 point) and 15 % probability of losing (-1 point), while the other stimulus had the reversed probability of winning (15 % and losing (85 %). The contingencies were switched between the stimuli every 12th trial.

The win-stay metric was defined as the proportion of trials where participants chose the same stimulus on the next trial following a win, and lose-shift as the proportion of trials where participants chose the other stimulus on the next trial following a loss.

2.6. Statistical analysis and modeling

All statistical analyses and modeling were performed in R [60]. Group differences in win-stay and lose-shift were tested using the Wilcoxon test for non-parametric data. Age group differences in sex distribution were tested using the chi-square test.

To test group differences in learning rate and inverse temperature and the associations between these parameters and rsFC, we developed a hierarchical Bayesian model in R. Behavior on the two-arm Bandit task was modeled using Q-learning and we entered rsFC values from the conn analyses. By hierarchical we mean that parameters were estimated at

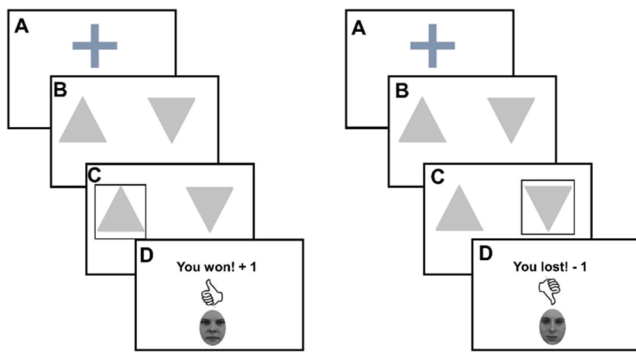


Fig. 1. The reward-learning task. Participants completed 60 trials of the two-arm Bandit task where one stimulus had an 85 % probability of winning and 15 % of losing, and the other stimulus had an 85 % probability of losing and 15 % of winning. Contingencies were reversed after every 12 trials. Each trial consisted of four phases. First a fixation cross presented for 1 s (A), then the two stimuli were presented (B) and remained on screen until participants made a choice (C). Feedback consisting of written notification of win or loss together with a stylized hand with a thumb pointing upwards or downwards and a face with a neutral expression. For each participant, the same face was always associated with wins, and the other with losses.

both the individual and group levels. Each individual is assumed to have its own learning rate and inverse temperature (individual level). On the other hand, parameters such as associations between learning rate and inverse temperature, and rsFC are group-level parameters. We refer to the model as joint since the distribution of all parameters (on both individual and group levels) were estimated simultaneously within a single model, rather than in separate regressions. This joint formulation allows us to account for correlations between connections, share statistical strength across parameters, and reduce the need for extensive multiple-comparison corrections. Posterior inference was carried out using a custom Metropolis-Hastings within Gibbs sampling scheme. The joint Bayesian model was used to estimate the posterior distributions of each model parameter simultaneously, along with their 98 % (and 80 %) credible intervals. If a 98 % credible interval excluded 0, we interpret this as evidence of a group difference or association in the corresponding parameter.

2.6.1. Resting-state functional connectivity

The resting-state functional connectivity (rsFC) data were modeled using a latent factor model to capture underlying dimensions of network covariance structure across subjects, adjusting for potential confounds such as head movement. The full mathematical specification of the factor model, including definitions of the observed and latent variables, factor loadings, and assumed error distributions, is provided in the [Supplementary Material S4](#).

2.6.2. Reward-learning model of behavior

To model reward-learning behavior we used the Q-learning model [61]. The value function (Q-value) $Q_{j,a}^t \in (-1, +1)$ of subject j at trial t , represents one of the choices $b_{j,t} \in \{0, 1\}$. $Q_{j,a}^t$ represents the subjective belief in the respective choices. $Q_{j,a}^t = 0$ represents complete uncertainty. The value function is updated according to the rule

$$Q_{i,a}^{t+1} = Q_{i,a}^t + \alpha_j (R_j^t - Q_{i,a}^t),$$

where α_j is the learning rate and $R_j^t \in \{-1, +1\}$ is the reward (getting one point). Let $b_j^t \in \{0, 1\}$ be the actual choice of subject j at choice t . $b_j^t = 1$ represents choosing one symbol, and $b_j^t = 0$ represents the opposite choice. Let the choice probabilities be $p_j^t = P(b_j^t = 1 | \alpha_j, \beta_j) = \text{logit}^{-1}(\beta_j(Q_{j,1}^t - Q_{j,0}^t))$, where β_j is the inverse temperature. This corresponds to the simplest special case of the Q-learning model, as applied to a two-choice bandit task, which is parameterized by a learning rate, α_j , and inverse temperature, β_j . Other Q-learning models which could be considered, include Q-learning with asymmetric learning rates, i.e., separate learning rates for losses and gains [62], and choice bias models which also includes a “stickiness parameter” to bias choices towards what was chosen previously, independent of Q-values [63]. We adopt the simplest model with learning rate and Q-learning because the relatively small number of trials [60] can limit the reliability of parameter estimation for α_j and β_j .

2.6.3. Joint model

In the joint model for neural data and behavior we identify the factors in the factor model $\theta_{j,1} \in (-\infty, +\infty)$ and $\theta_{j,2} \in (-\infty, +\infty)$ with the behavioral parameters α_j and β_j in a one-to-one correspondence as model for α_j and β_j :

$$\alpha_j = \text{logit}^{-1}(\theta_{j,1})$$

$$\beta_j = \theta_{j,2}$$

The logistic formulation of the relationship between $\theta_{j,1}$ and α_j result in the constraint $\alpha_j \in (0, 1)$ and $\beta_j \in (-\infty, +\infty)$. Although

negative values of β_j are rather unrealistic, our model does not restrict inverse temperatures to be strictly positive. Let $Y = \{n_1, \dots, n_N, b_1, \dots, b_N\}$ be the observed data and N the sample size. A schematic representation of the hierarchical model is shown in Fig. 2. Further, assuming $\theta_j \sim N(\kappa + Bx_j, \Phi)$, where Φ is a diagonal covariance matrix, and multivariate or matrix normal priors $p(\kappa)$, $p(B)$, $p(\Lambda)$ then the joint distribution of Y , the latent quantities, and all model parameters is known and we can construct an efficient MCMC estimator as described in the following subsection.

2.6.4. MCMC estimator

We fit the hierarchical Bayesian model using a hybrid Markov chain Monte Carlo (MCMC) algorithm combining Gibbs sampling and Metropolis–Hastings updates [64]. Gibbs sampling was applied to parameters with tractable full conditionals, while Metropolis–Hastings steps were used for more complex parameters. Convergence was assessed via standard diagnostics, and model performance was validated through extensive simulation studies (see Supplementary Material, Sections S1–S3, for convergence diagnostics, parameter recovery, and prior choices).

3. Results

3.1. Participants

A total of 62 participants completed the reward-learning paradigm and the MRI measurements. Three participants were excluded due to technical and behavioral issues during the reward-learning paradigm, specifically completing less than 85 % of trials ($n = 2$) and shifting after winning being more likely than shifting after a loss indicating that the participant did not understand or follow instructions ($n = 1$). From the neuroimaging data, one participant was excluded due to excessive head movement (max framewise displacement >2.4 mm, mean framewise displacement >0.3 mm, valid scans <80 %). After exclusion, 58 participants remained for behavioral and functional connectivity analyses, 32 healthy adolescents (age 13.6 ± 1.10 years, range 13–16; 16 male, 16 female) and 26 adults (age 34.3 ± 2.65 years, range 30–40, 16 male, 10 female). The groups did not differ on sex distribution ($\chi^2(1) = 0.376$, $p = .540$).

All adolescent participants were actively enrolled in school at the time of the study. For adults, educational levels varied but all participants had completed at least secondary education. Adult participant SES was calculated as the sum of self-reported educational level and yearly

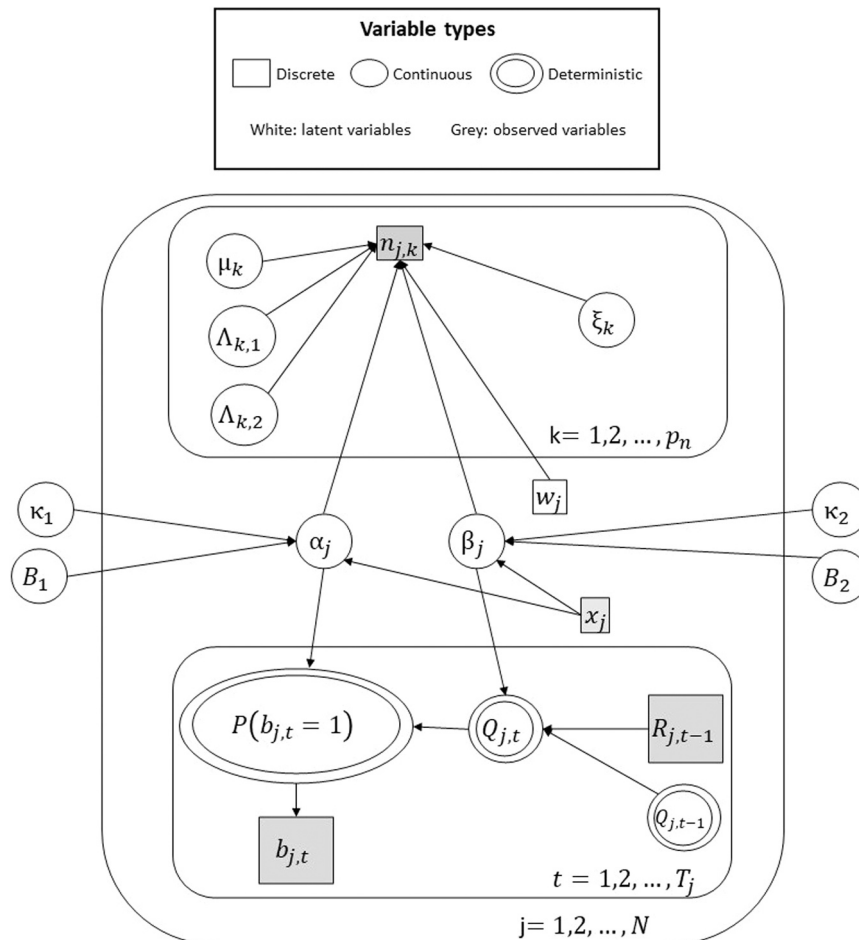


Fig. 2. Graphical illustration of the hierarchical Bayesian model. Subject and resting-state functional connectivity (rsFC) pair indices are j and k , respectively, there are T_j trials in the reward-learning task for subject j , and the sample size is N . μ_k is the mean connectivity of the k^{th} rsFC pair, $\Lambda_{k,1}$ is the association between the learning rate and the k^{th} rsFC pairs. $\Lambda_{k,2}$ is the association between the inverse temperature and the k^{th} rsFC pair, w_j is the head movement of subject j , ξ_k is the effect of head movement on the k^{th} rsFC pairs. $P(b_{j,t} = 1)$ is the probability of choosing one of the symbols in the reward-learning task ($b_{j,t}$), $R_{j,t-1}$ is the previous reward and $Q_{j,t}$ and $Q_{j,t-1}$ are the current and previous value functions. Here, we were specifically interested in α_j (learning rate), β_j (inverse temperature), κ_1 (the group average of the transformed learning rate of adolescents), κ_2 , (the group average inverse temperature of adolescents), B_1 (the average difference in transformed learning rate between adults and adolescents), B_2 (the average difference in inverse temperature between adults and adolescents), $\Lambda_{k,1}$ (the association between resting-state functional connectivity and transformed learning rate), and $\Lambda_{k,2}$ (the association between resting-state functional connectivity and inverse temperature).

income scores. For adolescents, SES was first calculated separately for both parents (using parental educational level and income), with the average of these values then used as an indicator of household SES [38]. SES did not differ significantly between groups, $t(55) = 0.40$, $p = .69$ (adolescents: $M = 7.69$, $SD = 1.04$; adults: $M = 7.58$, $SD = 1.17$).

To examine possible SES-related confounds, we included SES as a covariate in an exploratory hierarchical Bayesian model ($N = 57$ after excluding one participant with missing SES). However, the model failed to converge ($R = 1.62$), indicating unstable parameter estimates. Post-hoc regression of neural (21 variables) and behavioral residuals on the additional model complexity revealed only a few nominal associations, and after FDR correction (FDR, 22 tests) only one borderline effect remained (adjusted $p = 0.0459$). This isolated effect had a small magnitude and did not form any coherent pattern. The behavioral residuals showed no association (adjusted $p = 0.99$). Overall, the extended components do not explain meaningful residual variance, indicating that the simpler model is adequate.

3.2. Associations between reward-learning parameters and functional connectivity

To investigate how learning rate (α) and inverse temperature (β) parameters are related to functional connectivity strength in the reward circuit ($n_{j,k}$), we employed the proposed hierarchical model. The posterior distributions of the association between learning rate and rsFC ($\Lambda_{k,1}$) has most of its density (>99 %) to the left of zero for five connectivity pairs, ACC-NACC, OFC-Caudate, ACC-Caudate, mPFC-NACC, and mPFC-ACC (see Fig. 3 for density plots). This provides evidence for negative associations between learning rate and rsFC for those connectivity pairs. The posterior densities of association between inverse temperature and rsFC overlapped with zero for all of the connectivity pairs, suggesting no evidence for non-zero associations between inverse temperature and rsFC (see Fig. 4 for density plots).

3.3. Age-group differences in reward learning parameters and their association with functional connectivity

We then investigated the developmental effect on the behavioral parameters learning rate, $\alpha_j = \text{logit}^{-1}(-\theta_{j,2})$ and inverse temperature, $\beta_j = \theta_{j,1}$ and the relationship between those and the functional brain connectivity. Here, we have

$$B = \begin{pmatrix} B_1 \\ B_2 \end{pmatrix},$$

where B_1 and B_2 are the linear effects of age group on $\theta_{j,1} = \beta_j$ and $\theta_{j,2} = \text{logit}(\alpha_j)$ (the transformed learning rate), respectively. The conditional distributions of the behavioral parameters are, hence,

$$\theta_{j,1} \sim N(\kappa_1 + B_1 x_j, \Psi_1^{-1}),$$

$$\theta_{j,2} \sim N(\kappa_2 + B_2 x_j, \Psi_2^{-1}),$$

where $x_j = 0$ and $x_j = 1$ represent adolescents and adults, respectively. We observed an age-related shift in the mean posterior densities of the behavioral parameters, specifically learning rate and inverse temperature (Fig. 5, Table 1). Adolescents exhibited lower values for both parameters compared to adults, supporting our hypotheses that adolescents rely less on feedback when making decisions, as reflected in their lower learning rates, and are more variable choice behavior, indicated by lower inverse temperature.

We subsequently examined whether the associations between reward-learning parameters and rsFC differed by age group. Individual MAP estimates of the reward-learning parameters (learning rate and inverse temperature) were extracted and entered into multivariate linear regression models, with reward-learning parameter, age group,

and their interaction as independent variables, and functional connectivity as the dependent variable. Separate models were run for learning rate and inverse temperature. Across the 42 interaction tests, two effects reached nominal significance at the uncorrected threshold of $p < .05$: ACC-Put ($\beta = -1.73$, $SE = .68$, $p = .013$) and OFC-Put ($\beta = -1.90$, $SE = .68$, $p = .007$), but they did not survive correction for multiple comparisons using the false discovery rate (FDR) procedure.

3.4. Validation of the hierarchical Bayesian model

To validate the hierarchical Bayesian model, we conducted simulation studies using parameters similar to those estimated from empirical data. The hierarchical approach showed improved bias, RMSE, and out-of-sample predictive accuracy over conventional maximum likelihood estimation (full results in Supplementary Table S1 and Section S4).

3.5. Developmental differences in the win-stay lose-switch strategy

We also compared the probability of staying with an option after winning and switching after losing. Compared to adolescents ($Md = 0.206$, $IQR = 0.2384$), adults ($Md = 0.086$, $IQR = 0.147$) were more likely to stay with the winning option ($W = 643.5$, $p < .001$) (Fig. 6), but we detected no group differences in switching after losing (adults: $Md = 0.698$, $IQR = 0.139$; adolescents: $Md = 0.706$, $IQR = 0.114$) ($W = 417.5$, $p = 0.9875$).

4. Discussion

This study compared adults and adolescents on reward learning parameters and their associations with resting-state functional connectivity. In line with previous research, we found that adolescents had lower learning rate and inverse temperature than adults, indicating slower value updating [17,18] and more variable choice pattern [9,10]. In the whole sample, learning rate, but not inverse temperature, was associated with resting-state functional connectivity primarily involving the ACC, and adolescents had stronger associations between learning rate and fronto-striatal connectivity. These findings suggest distinct reward learning mechanisms in adolescence, and link reward learning to resting-state functional connectivity.

4.1. Functional connectivity and reward learning parameters

The present study demonstrates a link between functional connectivity strength and reward learning. In line with our hypothesis, learning rate was associated with functional connectivity strength within the reward circuit, providing insights into the neural mechanisms underlying feedback processing. The ACC emerged as a critical node in the connectivity patterns associated with learning rate. Prior work suggests that prefrontal regions can signal the relative salience of outcomes to modulate learning rate [65], supporting the notion that learning rate is neurally instantiated rather than purely abstract. In our case, the ACC may serve an analogous role as a hub linking salience and reward networks. The ACC is also a key hub of the salience network, which coordinates attention toward motivationally relevant stimuli and facilitates interactions between reward-processing and cognitive control networks [66,67]. This brain region plays a crucial role in monitoring outcomes, detecting conflicts, and signaling the need for behavioral adjustments [68]. Its prominent involvement in the learning rate-connectivity associations suggests that this region may be particularly important for integrating feedback information and updating value representations.

In contrast to our hypothesis, we detected no associations between inverse temperature (indicative of more variable choice pattern) parameters and functional connectivity, which may indicate that behavioral variability or exploratory behavior is supported by different neural mechanisms than those captured by our functional connectivity

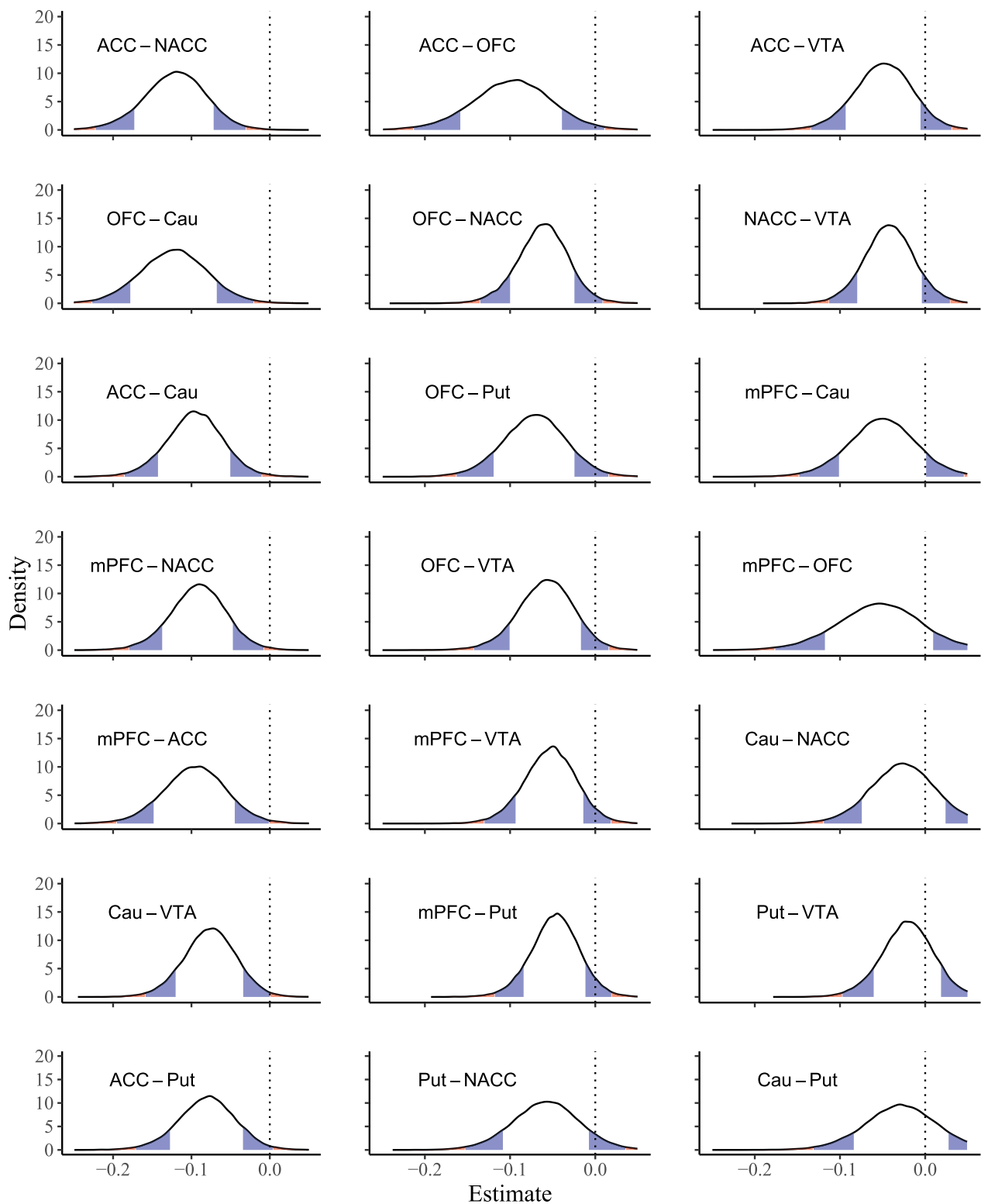


Fig. 3. Posterior densities of associations ($A_{1,k}$) between learning rate and resting-state functional connectivity (rsFC) pairs $n_{j,k}$. There were negative associations between learning rate and ACC-NACC, OFC-Cau, ACC-Cau, mPFC-NACC and mPFC-ACC functional connectivity pairs, as indicated by more than 99% of the density of the probability distribution to the left of zero i.e., $P(A_{1,k} < 0) > .99$. Each plot shows the estimated probability distribution ("posterior density") for the association between learning rate and a specific rsFC pair from the Bayesian model. The x-axis reflects possible values of the association (regression coefficient), and the y-axis reflects the relative likelihood of each value given the data. Densities primarily to the left of zero indicate a negative association: higher functional connectivity is linked to a lower learning rate. Color shading marks the 'tails' of the distribution: red regions represent the extreme percentiles (0–1% and 99–100%), which are the least probable; blue regions 1–10% and 90–99%. The dashed vertical lines correspond to the parameter equal to zero. The densities are sorted from up to down and then from left to right in increasing probability larger than zero. Included brain regions: anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), medial prefrontal cortex (mPFC), Putamen (Put), Caudate (Cau), ventral tegmental area (VTA), nucleus accumbens (NACC).

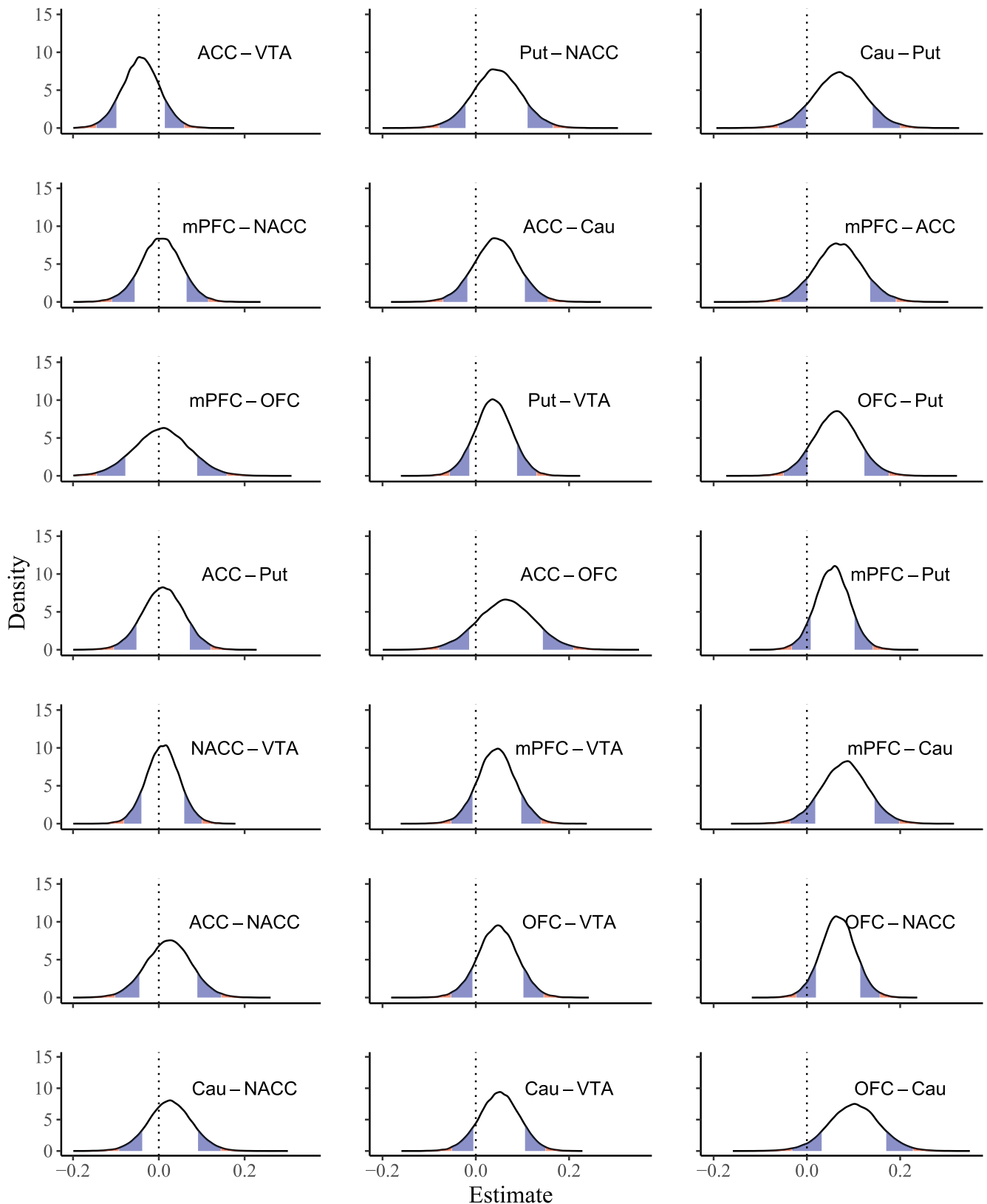


Fig. 4. Posterior densities of associations ($\lambda_{2,k}$) between learning rate and resting-state functional connectivity (rsFC) pairs $n_{j,k}$. None of the effects had a posterior probability exceeding 99 % of being strictly positive or strictly negative. Each plot shows the estimated probability distribution ("posterior density") for the association between inverse temperature and a specific rsFC pair from the Bayesian model. The x-axis reflects possible values of the association (regression coefficient), and the y-axis reflects the relative likelihood of each value given the data. Color shading marks the 'tails' of the distribution: red regions represent the extreme percentiles (0–1 % and 99–100 %, which are the least probable; blue regions 1–10 % and 90–99 %). The dashed vertical lines correspond to the parameter equal to zero. The densities are sorted from up to down and then from left to right in increasing probability larger than zero. Included brain regions: anterior cingulate cortex (ACC), orbitofrontal cortex (OFC), medial prefrontal cortex (mPFC), Putamen (Put), Caudate (Cau), ventral tegmental area (VTA), nucleus accumbens (NACC).

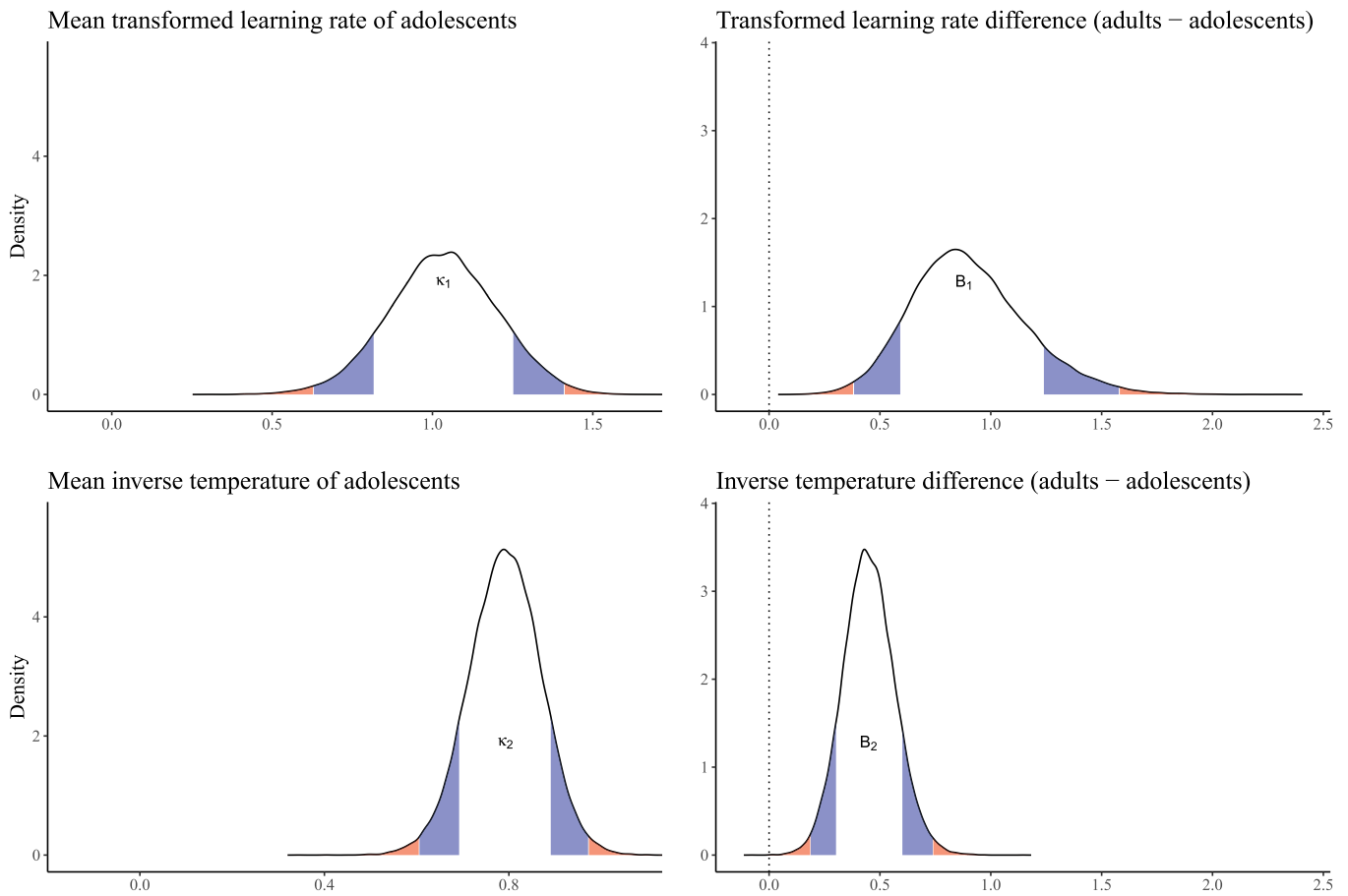


Fig. 5. Posterior densities of learning rate and inverse temperature parameters. Posterior densities of κ_1 and κ_2 corresponding to group-means of transformed learning rate and inverse temperature for adolescents, and of B_1 and B_2 corresponding to the difference between age groups (adults-adolescents) on transformed learning rate and inverse temperature, respectively. The transformed learning rate refers to $\text{logit}(\alpha_j) \in (-\infty, +\infty)$. Red regions correspond to percentiles 0–1 and 99–100, and blue regions correspond to percentiles 1–10 and 90–99. The dashed vertical lines correspond to a parameter value equal to zero. The posterior densities exhibit evidence for positive parameter values of all four parameters, i.e., $P(\kappa_1 > 0) > .99$ and $P(\kappa_2 > 0) > .99$ suggest that the learning rate and inverse temperature of adolescents are both larger than zero on average, and $P(B_1 > 0) > .99$ and $P(B_2 > 0) > .99$ suggest that the learning rate and inverse temperature both are larger in adults compared to adolescents on average. The MAP (maximum a posterior) estimates of learning rates of adolescents and adults are 0.73 and 0.87, respectively.

Table 1

Posterior estimates of learning rate and inverse temperature parameters. Parameter, posterior mean, HDI (Highest Density Interval; 90 % low and high limits) and posterior standard deviations of elements in κ_1 and κ_2 (mean transformed learning rate and inverse temperature of adolescents, respectively) and B_1 and B_2 (the difference in transformed learning rate and inverse temperature between adults and adolescents, respectively). The transformed learning rate refers to $\text{logit}(\alpha_j) \in (-\infty, +\infty)$. This is equivalent of $\alpha_j \in (0, 1)$. The posterior distribution of both B_1 and B_2 have posterior means 0.45 and 0.92, respectively, with HDI not including zero. This provides evidence that adults have higher learning rate and inverse temperature than adolescents. The MAP (maximum a posterior) estimates of learning rates (untransformed) of adolescents and adults are 0.73 and 0.87, respectively.

Parameter	Mean	HDI (low)	HDI (high)	SD
Mean transformed adolescent learning rate (κ_1)	1.0189	0.7502	1.2913	0.1644
Mean adolescent inverse temperature (κ_2)	0.7929	0.6620	0.9150	0.0775
Difference in transformed learning rate (B_1) adults - adolescents	0.9210	0.5067	1.3484	0.2605
Difference in inverse temperature (B_2) adults - adolescents	0.4500	0.2634	0.6406	0.1154

measures. Previous research suggests that uncertainty-directed exploration may be more supported by the rostralateral prefrontal cortex

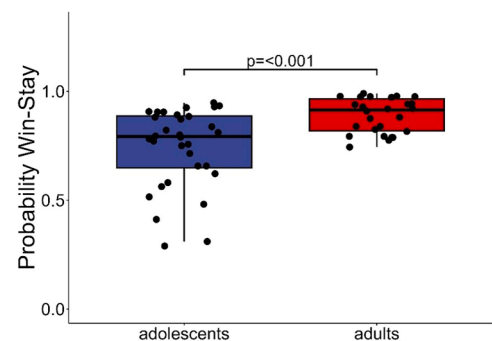


Fig. 6. Developmental differences in win-stay strategy. Box and whisker plot showing the probability of staying with a winning strategy, when participants repeated their previous choice following a reward in a two-arm bandit task. Each dot represents an individual participant. Adults (red) had a significantly higher win-stay probability than adolescents (blue) ($p < 0.001$, two-tailed t -test). Box plots display the median, interquartile range (IQR), and range.

(RLPFC) and random exploration more by dorsolateral prefrontal cortex [69], regions not included in our analyses but implicated in relational reasoning that tracks relative uncertainty around choice options and promotes exploration [6,70]. Here, we focused on the reward circuitry to keep the number of association tests down in the relatively small

sample, but future studies of the neural underpinnings of exploration-exploitation trade-offs may benefit from adding additional brain regions in larger sample sizes.

4.2. Developmental differences in functional connectivity and reward learning parameters

Supporting our hypothesis that adolescents differ from adults in how resting-state functional connectivity is linked to reward learning parameters, learning rate was more strongly associated with ACC-Putamen and OFC-Putamen functional connectivity in adolescents. Functionally, these circuits support different but complementary aspects of reward learning. The orbitofrontal cortex (OFC) encodes expected reward value and rapidly updates predictions based on new information, transmitting value signals to striatal targets such as the putamen [71,72]. The anterior cingulate cortex (ACC) monitors outcomes and guides behavioral adjustment [68]. Its connectivity with the putamen facilitates the integration of reward feedback into action selection and habit formation. The putamen tracks the likelihood that conditioning stimuli lead to correct motor responses, acting as a sensorimotor hub that integrates stimulus-reward associations with motor and cognitive processes [73]. These associations likely reflect the complex way that brain networks develop and function during adolescence. Different brain regions mature at different rates, with subcortical reward-processing areas developing earlier than prefrontal control regions, creating a functional imbalance. During normal adolescent brain development, neural networks undergo important refinement processes, including synaptic pruning and myelination, which lead to more efficient brain function [24]. These developmental changes result in brain activity patterns that become more focal and fine-tuned with age, while regions not essential for task performance show diminished activity [25]. This refinement process suggests that optimal brain function emerges through the strengthening of relevant connections and the elimination of unnecessary ones, rather than simply through increased overall connectivity [24].

4.3. Developmental differences in reward learning behavior

Furthermore, adolescents exhibited significantly lower learning rates and inverse temperature values compared to adults, supporting our hypothesis. These computational parameters reflect fundamental differences in how feedback is processed and utilized for future decision-making across developmental stages. The lower learning rates observed in adolescents suggest a reduced ability to integrate feedback effectively when updating beliefs about reward contingencies. This finding aligns with previous developmental studies showing that adolescents are slower to update expected values compared to adults [17, 18]. The reduced learning rate may reflect ongoing maturation of neural circuits involved in value updating and feedback processing, particularly the fronto-striatal networks that undergo substantial development during adolescence, as also noted in our results. Similarly, the lower inverse temperature values in adolescents indicate a more variable choice behavior, characterized by more frequent suboptimal choices and less reliance on learned values to guide behavior. This bias was further substantiated by the win-stay lose-switch analysis, which revealed that adolescents stayed less with winning options than adults. Hence, using two independent analytical methods, we found distinct developmental differences that highlight increased behavioral variability, often associated with exploratory tendencies, across adolescent development [9, 10]. The increased variability and reduced feedback integration observed in adolescents may contribute to both adaptive and maladaptive outcomes during this developmental period. Increased choice variability may reflect exploratory tendencies, that facilitates learning and promotes autonomy, it may also increase vulnerability to risky behaviors when combined with heightened reward sensitivity [5]. However, it remains unclear whether the lower learning rates in adolescents truly reflect a diminished capacity to integrate feedback, or whether

adolescents are able to integrate feedback but instead prioritize exploration over exploitation due to heightened novelty seeking and reduced sensitivity to potential risks. From a developmental standpoint, this distinction is important, as it suggests that adolescence may be characterized by a shifting balance between increased behavioral variability (often linked to exploration) and learning that together may support adaptive flexibility in changing environments [32].

4.4. Methodological considerations and model validation

We developed and tested a novel hierarchical Bayesian framework. This approach offers several advantages over traditional maximum likelihood methods for analyzing sparse behavioral data. Our validation analyses demonstrated superior performance of the hierarchical Bayesian method in terms of parameter estimation accuracy and out-of-sample prediction. The convergence diagnostics and cross-validation results support the reliability of our modeling approach. The integration of behavioral and neuroimaging data within a unified computational framework represents a significant methodological advance. This joint modeling approach allows for direct examination of brain-behavior relationships while accounting for the hierarchical structure of the data and individual differences in both domains.

4.5. Limitations and future directions

Several limitations should be acknowledged. First, the cross-sectional design limits our ability to make causal inferences about developmental trajectories. Longitudinal studies would provide stronger evidence for developmental changes in reward learning and neural connectivity. Second, the rsFC measures may not fully capture the dynamic neural processes involved in reward learning. Future studies could benefit from task-based connectivity analyses or measures of neural flexibility. The hierarchical Bayesian model provided accurate estimates of the parameters we investigated here, but the relatively small sample size limited us in exploring additional associations given the large number of free parameters that would be needed to for example uncover age-group differences in the relationships between reward learning parameters and functional connectivity within the Bayesian model. Hence, it does seem like a promising analysis framework but would benefit from larger sample sizes. Relatedly, although asymmetries in learning from rewarding versus punishing feedback are an important aspect of developmental reinforcement learning, adding separate parameters for reward- and loss-based updating resulted in unstable model convergence in exploratory analyses. Future studies with larger samples and tasks optimized for differentiating valence-specific learning should aim to investigate these mechanisms directly. When estimating learning rate and inverse temperature on an individual level, independently, there is a risk to run into identifiability issues. Different sets of learning parameters can give rise to similar behaviors. The proposed Bayesian model provides priors on the individual level, hence, respecting the uncertainty in learning parameters in an optimal way. However, future studies involve thorough investigations of when the learning parameters are separable and when they are part of a common construct. Additionally, the use of contingency reversals in a relatively short task may have limited the robustness of the computational modeling. Although reversal-induced changes in behavior can be informative, including pupil dilation effects, a simpler two-armed bandit design without reversals may have provided more stable estimates of learning parameters in this sample size. Future studies could increase trial numbers to improve model precision.

The adolescent sample was predominantly composed of 13-year-olds, which may not capture the full spectrum of adolescent brain development. The cortical-subcortical developmental changes that characterize different phases of adolescence may not be fully represented in this sample, potentially limiting the generalizability of our findings across the broader adolescent period.

In the analysis regarding developmental differences in associations between learning parameters and functional connectivity, we entered MAP estimates in a multivariate analysis instead of expanding the Bayesian model to allow for differences in associations between age groups. This was due to the small sample size and large number of parameters. A larger sample size would provide the possibility to employ a fully Bayesian model in this exploratory analysis.

5. Conclusion

This study provides novel insights into the developmental neuroscience of reward learning by integrating computational modeling with neuroimaging data. The findings demonstrate clear behavioral differences between adolescents and adults in reward-guided decision-making, characterized by increased behavioral variability and reduced feedback integration in adolescents. At the neural level, learning rate, but not inverse temperature, was associated with functional connectivity strength within reward circuits, with the ACC playing a prominent role and developmental differences revealing greater dependence on fronto-striatal connectivity in adolescents, consistent with the ongoing maturation of reward-related neural networks.

These results contribute to our understanding of typical adolescent development and have implications for identifying factors that promote adaptive decision-making during this critical period. The methodological approach developed here provides a framework for future studies examining brain-behavior relationships in developmental populations.

CRedit authorship contribution statement

Nils Kroemer: Writing – review & editing, Methodology. **Malin Gingnell:** Writing – review & editing, Resources, Project administration, Methodology, Investigation, Funding acquisition, Conceptualization. **Andreas Frick:** Writing – review & editing, Visualization, Supervision, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Johan Vegelius:** Writing – original draft, Visualization, Methodology, Formal analysis, Conceptualization. **Ebba Widegren:** Writing – review & editing, Project administration, Methodology, Investigation, Conceptualization. **Zsófia Karlocai:** Writing – original draft, Investigation, Formal analysis, Conceptualization. **Johan Lundin Kleberg:** Writing – review & editing, Software, Methodology. **David Fällmar:** Writing – review & editing, Investigation.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the authors used ChatGPT (OpenAI, GPT-5) and Perplexity AI to assist with improving the clarity and readability of the manuscript. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the content of the published article.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

The authors are grateful to the participants, whose involvement made this research possible. This work was supported by a grant from Riksbankens Jubileumsfond awarded to Dr. Frick and Dr. Gingnell (grant number P17-0256:1). Additional support for Dr. Frick came from Kjell och Märta Beijers stiftelse, the Swedish Research Council (grant number 2021-03106), and the AI4Research initiative at Uppsala

University. Dr. Fällmar received funding from the Swedish Society for Medical Research (SSMF, grant number PD21-0136) and Hjärnfonden (grant number PS2021-0026). Computational and data management resources were provided by the National Academic Infrastructure for Supercomputing in Sweden (NAISS), partly funded by the Swedish Research Council under grant no. 2022-06725. Furthermore, the authors gratefully acknowledge the support of the German Research Foundation (IRTG2804; Women's Mental Health Across the Reproductive Years). The funding organizations had no influence on the study design, data collection, analysis, interpretation, manuscript preparation, or the decision to submit this article for publication.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at [doi:10.1016/j.bbr.2025.116008](https://doi.org/10.1016/j.bbr.2025.116008).

Data Availability

Due to the sensitive nature of the data collected in this study, raw data will not be shared. Processed data will be shared upon reasonable request.

References

- [1] D.S. Fareri, L.N. Martin, M.R. Delgado, Reward-related processing in the human brain: developmental considerations, *Dev. Psychopathol.* 20 (4) (2008) 1191–1211.
- [2] C. Laube, W. Van Den Bos, Y. Fandakova, The relationship between pubertal hormones and brain plasticity: implications for cognitive training in adolescence, *Dev. Cogn. Neurosci.* 42 (2020) 100753.
- [3] J.H. Decker, F.S. Lourenco, B.B. Doll, C.A. Hartley, Experiential reward learning outweighs instruction prior to adulthood, *Cogn. Affect Behav. Neurosci.* 15 (2) (2015) 310–320.
- [4] J.J. Arnett, Adolescent storm and stress, reconsidered, *Am. Psychol.* 54 (5) (1999) 317–326.
- [5] G. Icenogle, E. Cauffman, Adolescent decision making: a decade in review, *J. Res. Adolesc.* 31 (4) (2021) 1006–1022.
- [6] K. Nussenbaum, C.A. Hartley, Reinforcement learning across development: what insights can we draw from a decade of research? *Dev. Cogn. Neurosci.* 40 (2019) 100733.
- [7] B.J. Casey, A.O. Cohen, A. Galvan, The beautiful adolescent brain: an evolutionary developmental perspective, *Ann. New Y. Acad. Sci.* 1546 (1) (2025) 58–74.
- [8] R.C. Wilson, A. Geana, J.M. White, E.A. Ludvig, J.D. Cohen, Humans use directed and random exploration to solve the explore–exploit dilemma, *J. Exp. Psychol. Gen.* 143 (6) (2014) 2074–2081.
- [9] M. Jepma, J.V. Schaaf, I. Visser, H.M. Huizenga, Effects of advice on experienced-based learning in adolescents and adults, *J. Exp. Child Psychol.* 211 (2021) 105230.
- [10] L. Xia, S.L. Master, M.K. Eckstein, B. Baribault, R.E. Dahl, L. Wilbrecht, et al., Modeling changes in probabilistic reinforcement learning during adolescence, in: S. Palminteri (Ed.), *PLoS Comput Biol.* 17, 2021 e1008524.
- [11] L. Steinberg, A social neuroscience perspective on adolescent risk-taking, *Dev. Rev.* 28 (1) (2008) 78–106.
- [12] M.K. Eckstein, S.L. Master, R.E. Dahl, L. Wilbrecht, A.G.E. Collins, Reinforcement learning and Bayesian inference provide complementary models for the unique advantage of adolescents in stochastic reversal, *Dev. Cogn. Neurosci.* 55 (2022) 101106.
- [13] A. Lloyd, R. McKay, C.L. Sebastian, J.H. Balsters, Are adolescents more optimal decision-makers in novel environments? Examining the benefits of heightened exploration in a patch foraging paradigm, *Dev. Sci.* 24 (4) (2021) e13075.
- [14] D. Wang, W. Wei, L. Li, X. Wang, J. Liang, Rethinking exploration–exploitation trade-off in reinforcement learning via cognitive consistency, *Neural Netw.* 187 (2025) 107342.
- [15] W. Van Den Bos, R. Bruckner, M.R. Nassar, R. Mata, B. Eppinger, Computational neuroscience across the lifespan: promises and pitfalls, *Dev. Cogn. Neurosci.* 33 (2018) 42–53.
- [16] K. Shimomura, K. Morita, The effect of reward magnitude on different types of exploration in human reinforcement learning, *Comput. Brain Behav.* 8 (1) (2025) 147–161.
- [17] J.Y. Davidow, C. Insel, L.H. Somerville, Adolescent development of value-guided goal pursuit, *Trends Cogn. Sci.* 22 (8) (2018) 725–736.
- [18] A.H. Javadi, D.H.K. Schmidt, M.N. Smolka, Adolescents adapt more slowly than adults to varying reward contingencies, *J. Cogn. Neurosci.* 26 (12) (2014) 2670–2681.
- [19] A. Christakou, S.J. Gershman, Y. Niv, A. Simmons, M. Brammer, K. Rubia, Neural and psychological maturation of decision-making in adolescence and young adulthood, *J. Cogn. Neurosci.* 25 (11) (2013) 1807–1823.

- [20] A. Gopnik, Childhood as a solution to explore–exploit tensions, *Philos. Trans. R. Soc. B* 375 (1803) (2020) 20190502.
- [21] R.S. Sutton, A.G. Barto, *Reinforcement learning: An introduction*, 1st ed., 1, MIT Press, 1998.
- [22] M. Cassotti, O. Houdé, S. Moutier, Developmental changes of win-stay and loss-shift strategies in decision making, *Child Neuropsychol.* 17 (4) (2011) 400–411.
- [23] G. Chierchia, M. Soukupová, E.J. Kilford, C. Griffin, J. Leung, S. Palminteri, et al., Confirmatory reinforcement learning changes with age during adolescence, *Dev. Sci.* 26 (3) (2023) e13330.
- [24] B.J. Casey, R.M. Jones, L.H. Somerville, Braking and accelerating of the adolescent brain, *J. Res. Adolesc.* 21 (1) (2011) 21–33.
- [25] Galvan, Adolescent development of the reward system. *Front Hum Neurosci* [Internet]. 2010 [cited 2025 Aug 14]; Available from: (<http://journal.frontiersin.org/article/10.3389/fnhum.2010.006.2010/abstract>).
- [26] A.F.P. Sanders, M.P. Harms, S. Kandala, S. Marek, L.H. Somerville, S. Y. Bookheimer, et al., Age-related differences in resting-state functional connectivity from childhood to adolescence, *Cereb. Cortex.* 33 (11) (2023) 6928–6942.
- [27] S.D. Klein, P.F. Collins, V. Lozano Wun, P. Grund, M. Luciana, Fronto-striatal networks undergo functional specialization during adolescence that follows a ventral–dorsal gradient: developmental trajectories and longitudinal associations, *J. Neurosci.* 45 (15) (2025) e1233232025.
- [28] Y. Xiao, D. Alkire, D. Moraczewski, E. Redcay, Developmental differences in brain functional connectivity during social interaction in middle childhood, *Dev. Cogn. Neurosci.* 54 (2022) 101079.
- [29] A. Ojha, A.C. Parr, W. Foran, F.J. Calabro, B. Luna, Puberty contributes to adolescent development of fronto-striatal functional connectivity supporting inhibitory control, *Dev. Cogn. Neurosci.* 58 (2022) 101183.
- [30] M.L. Rosen, M.A. Sheridan, K.A. Sambrook, M.J. Dennison, J.L. Jenness, M. K. Askren, et al., Salience network response to changes in emotional expressions of others is heightened during early adolescence: relevance for social functioning, *Dev. Sci.* 21 (3) (2018) e12571.
- [31] R.G. Lewis, E. Florio, D. Punzo, E. Borrelli, The Brain's Reward System in Health and Disease, *Adv. Exp. Med Biol.* 1344 (2021) 57–69.
- [32] C.A. Hartley, L.H. Somerville, The neuroscience of adolescent decision-making, *Curr. Opin. Behav. Sci.* 5 (2015) 108–115.
- [33] E. Gozdas, S.K. Holland, M. Altaye, CMIND Authorship Consortium, Developmental changes in functional brain networks from birth through adolescence, *Hum. Brain Mapp.* 40 (5) (2019) 1434–1444.
- [34] M.G. Tapia Medina, R. Cosío-Guirado, M. Perú-Cebollero, C. Cañete-Massé, E. R. Villuendas-González, J. Guardia-Olmos, The clinical relevance of healthy neurodevelopmental connectivity in childhood and adolescence: a meta-analysis of resting-state fMRI, *Front Neurosci.* 19 (2025) 1576932.
- [35] C. Gentili, E. Di Rosa, I. Podina, R. Popita, B. Voinescu, D. David, Resting state predicts neural activity during reward-guided decision making: An fMRI study on Balloon Analogue Risk Task, *Behav. Brain Res.* 417 (2022) 113616.
- [36] K. Choi, E. Piasini, E. Díaz-Hernández, L.V. Cifuentes, N.T. Henderson, E.N. Holly, et al., Distributed processing for value-based choice by prelimbic circuits targeting anterior-posterior dorsal striatal subregions in male mice, *Nat. Commun.* 14 (1) (2023) 1920.
- [37] A.C.K. Van Duijvenvoorde, B. Westhoff, F. De Vos, L.M. Wierenga, E.A. Crone, A three-wave longitudinal study of subcortical–cortical resting-state connectivity in adolescence: Testing age- and puberty-related changes, *Hum. Brain Mapp.* 40 (13) (2019) 3769–3783.
- [38] E. Widegren, J. Vegelius, M.A. Frick, A.A. Roy, S. Möller, J.L. Kleberg, et al., Fear extinction retention in children, adolescents, and adults, *Dev. Cogn. Neurosci.* 71 (2025) 101509.
- [39] M. Guath, J.L. Kleberg, J. Weis, E. Widegren, M. Frick, S. Möller, et al., Pupil dilation during negative prediction errors is related to brain choline concentration and depressive symptoms in adolescents, *Behav. Brain Res.* 436 (2023) 114060.
- [40] D.V. Sheehan, K.H. Sheehan, R.D. Shytle, J. Janavs, Y. Bannon, J.E. Rogers, et al., Reliability and validity of the mini international neuropsychiatric interview for children and adolescents (MINI-KID), *J. Clin. Psychiatry* 71 (03) (2010) 313–326.
- [41] Nieto-Castanon A., Whitfield-Gabrieli S. CONN functional connectivity toolbox: RRID SCR_009550, release 20 [Internet]. 2020. Available from: 10.56441/hilbertpress.2048.3738.
- [42] A. Nieto-Castanon, FMRI minimal preprocessing pipeline. In: *Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN*, Hilbert Press, 2020, pp. 3–16.
- [43] J.L.R. Andersson, C. Hutton, J. Ashburner, R. Turner, K. Friston, Modeling Geometric Deformations in EPI Time Series, *NeuroImage* 13 (5) (2001) 903–919.
- [44] Karl J. Friston, J. Ashburner, C.D. Frith, J. -B. Poline, J.D. Heather, R.S. J. Frackowiak, Spatial registration and normalization of images, *Hum. Brain Mapp.* 3 (3) (1995) 165–189.
- [45] R. Henson, C. Buechel, O. Josephs, K.J. Friston, The slice-timing problem in event-related fMRI, *NeuroImage* (1999).
- [46] R. Sladky, K.J. Friston, J. Tröstl, R. Cunnington, E. Moser, C. Windischberger, Slice-timing effects and their correction in functional MRI, *NeuroImage* 58 (2) (2011) 588–594.
- [47] Whitfield-Gabrieli, Nieto-Castanon, Ghosh. *Artifact detection tools (ART)*. Cambridge, MA; 2011.
- [48] Nieto-Castanon A. Preparing fMRI Data for Statistical Analysis [Internet]. arXiv: 2022 [cited 2025 Aug 15]. Available from: (<https://arxiv.org/abs/2210.13564>).
- [49] J.D. Power, A. Mitra, T.O. Laumann, A.Z. Snyder, B.L. Schlaggar, S.E. Petersen, Methods to detect, characterize, and remove motion artifact in resting state fMRI, *NeuroImage* 84 (2014) 320–341.
- [50] V.D. Calhoun, T.D. Wager, A. Krishnan, K.S. Rosch, K.E. Seymour, M.B. Nebel, et al., The impact of T1 versus EPI spatial normalization templates for fMRI data analyses, *Hum. Brain Mapp.* 38 (11) (2017) 5331–5342.
- [51] J. Ashburner, A fast diffeomorphic image registration algorithm, *NeuroImage* 38 (1) (2007) 95–113.
- [52] J. Ashburner, K.J. Friston, Unified segmentation, *NeuroImage* 26 (3) (2005) 839–851.
- [53] A. Nieto-Castanon, FMRI denoising pipeline. *Proceeding of Handbook of Functional Connectivity Magnetic Resonance Imaging Methods in CONN*, Hilbert Press, 2020, pp. 17–25.
- [54] K.J. Friston, S. Williams, R. Howard, R.S.J. Frackowiak, R. Turner, Movement-Related effects in fMRI time-series, *Magn. Reson. Med.* 35 (3) (1996) 346–355.
- [55] M.N. Hallquist, K. Hwang, B. Luna, The nuisance of nuisance regression: spectral misspecification in a common approach to resting-state fMRI preprocessing reintroduces noise and obscures functional connectivity, *NeuroImage* 82 (2013) 208–225.
- [56] Y. Behzadi, K. Restom, J. Liu, T.T. Liu, A component based noise correction method (CompCor) for BOLD and perfusion based fMRI, *NeuroImage* 37 (1) (2007) 90–101.
- [57] X.J. Chai, A.N. Castañón, D. Öngür, S. Whitfield-Gabrieli, Anticorrelations in resting state networks without global signal regression, *NeuroImage* 59 (2) (2012) 1420–1428.
- [58] A.C. Trutti, L. Fontanesi, M.J. Mulder, P.L. Bazin, B. Hommel, B.U. Forstmann, A probabilistic atlas of the human ventral tegmental area (VTA) based on 7 Tesla MRI data, *Brain Struct. Funct.* 226 (4) (2021) 1155–1167.
- [59] A. Nieto-Castanon, Functional Connectivity measures. *Proceeding of Handbook of functional connectivity Magnetic Resonance Imaging methods in CONN*, Hilbert Press, 2020, pp. 26–62.
- [60] R Core Team, R: A Language and Environment for Statistical Computing [Internet], R Foundation for Statistical Computing, Vienna, Austria, 2024. (<https://www.R-project.org/>) (Available from).
- [61] C.J.C.H. Watkins, P. Dayan, Q-learning, *Mach. Learn* 8 (3–4) (1992) 279–292.
- [62] S.J. Gershman, Do learning rates adapt to the distribution of rewards? *Psychon. Bull. Rev.* 22 (5) (2015) 1320–1327.
- [63] K. Katahira, The statistical structures of reinforcement learning with asymmetric value updates, *J. Math. Psychol.* 87 (2018) 31–45.
- [64] H. Zhu, S. Lee, Statistical analysis of nonlinear factor analysis models, *Brit J. Math. Stat. Sci.* 52 (2) (1999) 225–242.
- [65] V.M. K Nambodiri, T. Hobbs, I. Trujillo-Pisanty, R.C. Simon, M.M. Gray, G. D. Stuber, Relative salience signaling within a thalamo-orbitofrontal circuit governs learning rate, *Curr. Biol.* 31 (23) (2021) 5176–5191.e5.
- [66] W. Snyder, L.Q. Uddin, J.S. Nomi, Dynamic functional connectivity profile of the salience network across the life span, *Hum. Brain Mapp.* 42 (14) (2021) 4740–4749.
- [67] W.W. Seeley, V. Menon, A.F. Schatzberg, J. Keller, G.H. Glover, H. Kenna, et al., Dissociable Intrinsic Connectivity Networks for Salience Processing and Executive Control, *J. Neurosci.* 27 (9) (2007) 2349–2356.
- [68] M.F.S. Rushworth, M.P. Noonan, E.D. Boorman, M.E. Walton, T.E. Behrens, Frontal cortex and reward-guided learning and decision-making, *Neuron* 70 (6) (2011) 1054–1069.
- [69] M.S. Tomov, V.Q. Truong, R.A. Hundia, S.J. Gershman, Dissociable neural correlates of uncertainty underlie different exploration strategies, *Nat. Commun.* 11 (1) (2020) 2371.
- [70] D. Badre, B.B. Doll, N.M. Long, M.J. Frank, Rostrolateral prefrontal cortex and individual differences in uncertainty-driven exploration, *Neuron* 73 (3) (2012) 595–607.
- [71] J.S. Riceberg, M.L. Shapiro, Orbitofrontal cortex signals expected outcomes with predictive codes when stable contingencies promote the integration of reward history, *J. Neurosci.* 37 (8) (2017) 2010–2021.
- [72] E.T. Rolls, W. Cheng, J. Feng, The orbitofrontal cortex: reward, emotion and depression, *Brain Commun.* 2 (2) (2020) fcaa196.
- [73] A. Brovelli, B. Nazarian, M. Meunier, D. Boussaoud, Differential roles of caudate nucleus and putamen during instrumental learning, *NeuroImage* 57 (4) (2011) 1580–1590.