# Enhanced response switching after negative feedback and novelty seeking in adolescence are associated with reduced representation of choice probability in medial frontal pole

Maria Waltmann[a,b]*, Nadine Herzog[b], Andrea M.F. Reiter [a,c], Arno Villringer[b,d], Annette Horstmann[b,e], Lorenz Deserno[a,b,f]

[a] Department of Child and Adolescent Psychiatry, Psychosomatics and Psychotherapy, Centre of Mental Health, University of Würzburg, Würzburg, Germany

[b] Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[c] CRC-940 Volition and Cognitive Control, Faculty of Psychology, Technical University of Dresden, Dresden, Germany

[d] MindBrainBody Institute, Berlin School of Mind and Brain, Charité-Universitätsmedizin Berlin and Humboldt-Universität zu Berlin, Berlin, Germany

[e] Department of Psychology and Logopedics, Faculty of Medicine, University of Helsinki, Helsinki, Finland

[f] Neuroimaging Center, Technical University of Dresden, Dresden, Germany

* Corresponding author

Address: Margarete-Höppel-Platz 1, 97080 Würzburg, Germany

E-Mail: waltmann_m@ukw.de

Phone: +49 (0931) 201 – 78700

**ABSTRACT**

Precisely charting the maturation of core neurocognitive functions such as reinforcement learning (RL) and flexible adaptation to changing action-outcome contingencies is key for developmental neuroscience. It can also help us understand how disruptions during development might contribute to the onset of psychopathology. However, research in this area is both sparse and conflicted, especially regarding potentially asymmetric development of learning for different motives (obtain wins vs avoid losses) and learning from valenced feedback (positive vs negative). In the current study, we investigated the development of RL from adolescence to adulthood, using a probabilistic reversal learning task modified to experimentally separate motivational context and feedback valence, in a sample of 95 healthy participants between 12 and 45. We show that adolescence is characterized by enhanced novelty seeking and response shifting after negative feedback, which leads to poorer returns when reward contingencies are stable. Computationally, this is accounted for by reduced impact of positive feedback on behavior. We also show, using fMRI, that activity of the medial frontopolar cortex reflecting choice probability is attenuated in adolescence. We argue that this can be interpreted as reflecting diminished confidence in upcoming choices. Interestingly, we find no age-related differences between learning in win and loss contexts.

**Keywords:** *Adolescence, Reversal Learning, Reinforcement Learning, Computational Modelling, fMRI*

# 1. INTRODUCTION

Adolescence is a pivotal period of neurocognitive development (Dahl et al., 2018). Precisely charting the maturation of critical neurocognitive functions such as reinforcement learning can help us, for example, tailor educational programs to different age groups, but also understand potentially consequential developmental difficulties. Thus, many psychiatric disorders with adolescent onset (Kessler et al., 2007; Paus et al., 2008) are associated with altered reinforcement learning (RL) and impaired adaptation to changing action-outcome contingencies (Crawley et al., 2020; Hauser et al., 2014; Mukherjee et al., 2020; Reddy et al., 2016; Reiter et al., 2016, 2017; Robinson et al., 2012; Schlagenhauf et al., 2014; Smith et al., 2015; Tezcan et al., 2017). Yet whether and how disruptions during development might contribute to their etiology is unclear, partly because our understanding of the normative development of reinforcement learning is still incomplete and conflicted.

Evidence from self-reports and functional neuroimaging (mainly during gambling/risk taking tasks) suggests that adolescence is characterized by heighted reward sensitivity and/or reduced punishment sensitivity (Barkley-Levenson & Galván, 2014; Davidow et al., 2016; Ernst et al., 2005; Galvan et al., 2006; Schreuders et al., 2018; although see e.g., Bjork et al., 2004). However, there is little indication that this straightforwardly translates to differences in reinforcement learning and instrumental behavior. Indeed, several studies indicate *diminished* impact of feedback (learnt values) on instrumental behavior in youths, evidenced by enhanced choice switching and stochasticity (e.g., Christakou et al., 2013; Crawley et al., 2020; Eckstein et al., 2021; Javadi et al., 2014; see Bolenz et al., 2017; Nussenbaum & Hartley, 2019 for reviews). Reports of age effects on the neural coding of reward prediction errors (RPEs) are largely incongruous (Christakou et al., 2013; Cohen et al., 2010; Hauser et al., 2015; Javadi et al., 2014; van den Bos et al., 2012), as are findings of age differences in learning rates for positive and negative feedback (Christakou et al., 2013; Jones et al., 2014; Rosenbaum et al., 2022; van den Bos et al., 2012). In a recent review, Nussenbaum & Hartley (2019) suggest that such inconsistencies might partially stem from different task demands. They argue that, rather than monotonically changing during development, learning rates and processes like RPE coding may become more task-optimal with age.

Though compelling, this does not explain inconsistencies between studies that employed tasks whose demands are largely similar. For example, Eckstein et al.

(2021) and Hauser et al. (2015) report opposite effects of age on learning rates for positive and negative feedback – yet both employed a probabilistic reversal learning task. Interestingly, in Eckstein et al.'s (2021) study, the outcomes available were wins and neutral events, while in Hauser et al.'s (2015) study, there were losses as well as wins. Given evidence that adolescents may learn better from wins than losses (Palminteri et al., 2016), this difference in motivational context may have been paramount. However, effects of motivational context (obtain wins vs avoid losses) and feedback valence (positive vs negative) were not differentiated in these studies. But learning in different motivational contexts and learning from unexpected positive and negative outcomes may have distinct developmental trajectories. Experimentally separating motivational context and feedback valence is therefore a necessary next step in charting the development of RL.

In the present study, we employed a mixed cross-sectional and longitudinal design to investigate the development of probabilistic reversal learning in a sample of adolescents and younger and older adults (12-45 years). We aimed, first, to replicate the relatively consistent previous finding of enhanced choice switching in adolescence compared to adulthood. Second, we aimed to disentangle learning in different motivational contexts (gain rewards vs avoid losses) from valenced feedback processing (positive vs. negative). To this end, we used a task modified to separate the influences of motivational context from feedback valence. We also introduced a post-task test designed to measure how well participants learned from wins compared to losses in the main task (Frank et al., 2004; Palminteri et al., 2016). In this task, familiar stimuli are pitted against one another and against novel stimuli. Based on the literature, we expected younger participants to perform worse when trying to avoid losses (Palminteri et al., 2016), and to process valenced feedback less optimally (i.e., less staying after positive, more switching after negative feedback) compared to older participants  (Crawley et al., 2020; Javadi et al., 2014). Third, we aimed to identify differences in computational processes that may underly age differences in behavior. Previous work indicated increased choice stochasticity/decreased feedback sensitivity (Eckstein et al., 2021; Javadi et al., 2014; Nussenbaum & Hartley, 2019) and decreased counterfactual inference in youths (Palminteri et al., 2016). Finally, we aimed to chart the development of the neural representations of RPEs and relative value (choice probability).

4

## 2. METHODS

### 2.1. Participants and procedure.

As part of a larger study, we recruited N = 95 right-handed healthy participants between the ages of 12 and 45 from the participant pool of the Max Planck Institute for Human Cognitive and Brain Sciences, as well as via advertisements in local schools, universities, GP practices, gyms, and shops. The study protocol consisted of a battery of interviews, questionnaires, physical examinations, neuropsychological assessments, and tasks (reported in full elsewhere). As part of this protocol, participants performed a probabilistic reversal learning task during functional magnetic resonance imaging (MRI) and completed a post-task probabilistic selection task (~30 min. after the end of the main task). Additionally, they completed the Trail-Making Test (Reitan, 1958), the digit-symbol-substitution task (Wechsler, 2008), a digitalized version of the digit span task (Wechsler, 2008) and a vocabulary test (Wortschatztest) (Schmidt & Metzler, 1992). A minimum of 6 months after their first visit (max 41 months, median = 8.71 months), participants were re-invited for a follow-up session in which they repeated the probabilistic reversal learning task and the post-task probabilistic selection task (without MRI measurement). Participants were compensated for their time (money or an Amazon voucher for minors) separately after the second and third study day. Information on demographics and neuropsychology is summarized in Table 1.

**Table 1. Demographics and Neuropsychological Assessment**

|                                  | Adolescents (Age ≤18) | Adults (Age >18) |
| -------------------------------- | --------------------- | ---------------- |
| N                                | 40                    | 55               |
| Age                              | 14.80 (±1.66)         | 28.68 (±5.58)    |
| Follow-up Interval (years)       | 1.08 (±0.75)          | 1.09 (±0.76)     |
| Drop-out                         | 7.5 %                 | 30.91 %          |
| Gender (% female)                | 50.00 %               | 60.00 %          |
| Years of education (full-time)   | 8.39 (±1.71)          | 17.29 (±3.89)    |

| | | |
|---|---|---|
| TMT-A (seconds) | 24.13 (±8.72) | 19.53 (±5.62) |
| TMT-B (seconds) | 52.94 (±25.42) | 39.47 (±11.63) |
| Digit Span Forward (levels achieved) | 5.85 (±1.14) | 6.69 (±1.53) |
| Digit Span Backwards (levels achieved) | 4.80 (±0.99) | 5.38 (±1.52) |
| Digit-Symbol-Substitution Task (symbols completed) | 67.60 (±14.90) | 82.04 (±15.31) |
| Wortschatztest (raw score) | 21.43 (±7.79) | 33.75 (±2.66) |

*N.B. Unless stated otherwise, values represent means, brackets contain standard deviations*

### 2.2. Task.

The probabilistic reversal learning task (PRLT) employed in this study (see also Boehme et al., 2015; Deserno et al., 2020; Reiter et al., 2016, 2017 for similar implementations), consists of two blocks of 140 trials in which participants make repeated binary choices between two cards. The cards are associated with different probabilities of winning (+10 cents) or not winning (±0 cents) (80%-20% and 20%-80%, respectively) in the win block, and of losing (-10 cents) or not losing (±0 cents) in the loss block (order counterbalanced). Neutral outcomes (±0 cents) signal negative feedback (no win) in the win condition, and positive feedback (no loss) in the loss condition. Independent from feedback valence, the motivational context in the two blocks is different: in the win condition, the goal is to collect as many rewards as possible; in the loss condition, the goal is to avoid losses.

In each trial, after making a choice by pressing a button (button box in the MRI, "n" and "m" keys on the PC for training), participants are shown a feedback screen (a picture of a 10-cents coin with a green plus sign for wins, a picture of a 0 cents coin for neutral outcomes, a picture of a 10-cents coin with a red minus sign for losses) for 0.5 seconds. Feedback (positive vs negative) is read out at each trial from a pre-defined schedule that was designed to match the reward/loss probabilities (i.e., for an 80%-win stimulus, 1 in every 5 choices was not rewarded). The feedback screen is followed by a variable inter-trial interval with a mean of 2.5 seconds, in which participants are shown a fixation cross (Fig. 1 – A, upper panel). After an initial acquisition phase (1st to 35th trial) the

cards' reward contingencies flip 5 times (after the 35th, 55th, 70th, 85th, and 105th trial), such that the previously more lucrative/less losing stimulus now becomes the more frequently neutral/losing one, and vice versa. For details, see Fig. 1 – A, lower panel.

In the bonus task, a probabilistic selection task, the stimuli from the two blocks of the PRLT and two novel stimuli (instructed as yielding a neutral outcome, i.e., neither win nor loss) are presented in all possible pairings, 3 times for each pair (totaling 45 trials). For each pair, participants have to pick the stimulus they thought most likely to produce a win / avoid a loss, without receiving feedback, equivalent to the "test phase" of the Frank probabilistic stimulus selection task (Frank et al., 2004) (Fig. 1 – B).
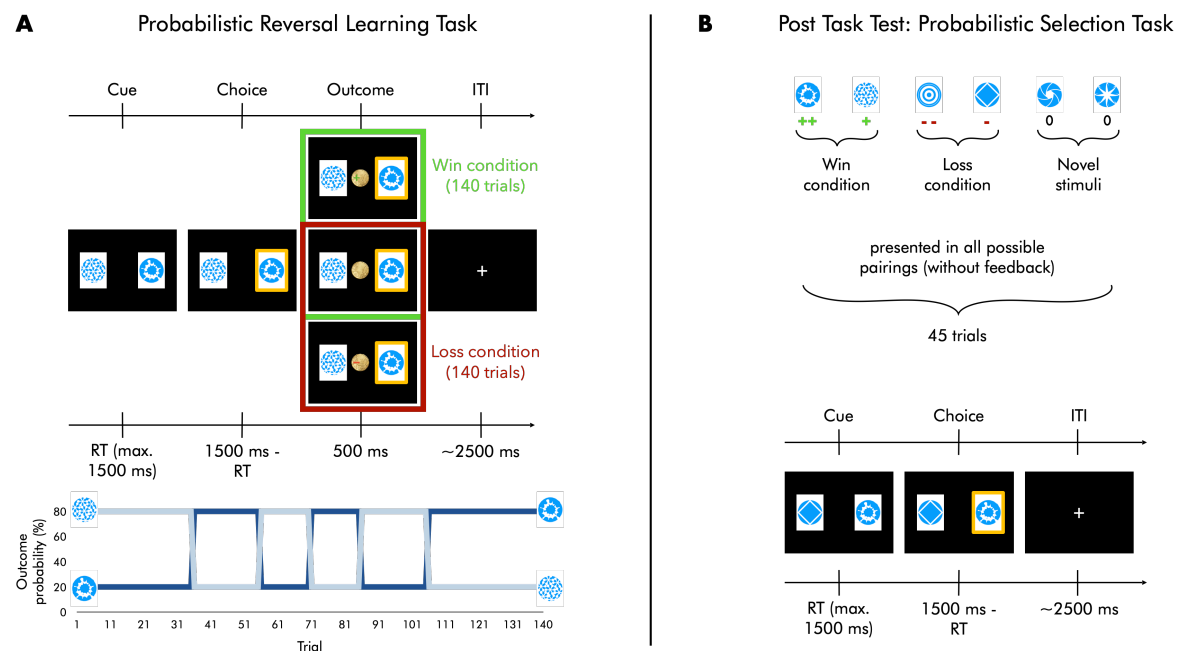


**Fig. 1**. **A**, upper panel – Design of the probabilistic reversal learning task (PRLT). In each condition (block), participants make 140 binary choices between two abstract stimuli (cards) with different probabilities of obtaining rewards, neutral outcomes, or losses (rewards and neutral outcomes in the win condition, neutral outcomes and losses in the loss condition). They are instructed to gain as much and lose as little money as possible, depending on condition. At each trial, the stimuli are shown for a maximum of 1500ms or until the participant responds. A frame then appears around the chosen card. This screen is shown for the remainder of 1500ms, i.e., for 1500ms minus the response time. Then, a feedback screen with either a picture of a 10-cents coin (wins), a picture of a 0-cents coin (neutral outcomes), or a picture of a minus 10-cents coin (losses) is shown for 500ms. Wins and neutral no-wins are available in the win condition and neutral no-loss and loss are available in the loss condition. Finally, participants see a fixation cross for a variable intertrial interval (mean 2500ms). **A**, lower panel – Reward contingencies. In the first 35 trials, the same stimuli each have a 20%- and 80%-win/loss probability, respectively. Their contingencies then reverse 5 times over the course of the task in a perfectly

anticorrelated manner, which requires participants to flexibly adapt their behavior in order to gain and avoid losing money. The task ends with another 35 trials, in which the reward contingencies no longer change. **B** – Post-task test: probabilistic selection task. Approximately 30 minutes after the PRLT, participants complete a short bonus task, in which the stimuli from both the win and the loss blocks and two novel stimuli (instructed as being neutral, i.e., yielding 0 cents outcomes) are presented in all 15 possible pairings (3 times for each pair). Participants are instructed to try to earn as much and lose as little money as possible as before. No feedback is provided.

The experiment was implemented in Psychtoolbox (3.0.13) using Octave (4.2.2). The PRLT was displayed on a white screen using a projector in the MRI, and on a monitor outside the MRI for training purposes. The same monitor was used to display the probabilistic selection task.

*2.3. Analysis of behavior – PRLT.*

We used trial-by-trial logistic mixed effects models to estimate accuracy (probability of choosing the currently more lucrative/less likely to lose card) and stay-switch behavior (probability of sticking with the same card as in the previous trial after positive and negative feedback), using the package *glmer* in R (version 4.1.0). As predictor variables, we included age (z-scored, per timepoint), condition (win vs. loss), and previous feedback (positive vs. negative) for stay-switch behavior. As an explorative analysis, we also looked at the effect of age on reaction times after positive and negative feedback, in the different conditions, using a linear mixed effects model. All our models employed a maximal random effects structure to the extent possible (Barr et al., 2013); we report the exact models in the supplement. Results were considered significant at $p < .05$, with p-values derived using Wald-Z tests in the case of GLMMs (as implemented in *glmer*) and Satterthwaite's method (as implemented in the *lmerTest* package) in the case of LMMs.

The task has different parts with stable or changing outcome probabilities, which can be captured in the regression models in different ways. Thus, models can differentiate between task phases (acquisition phase, i.e., the first 35 trials of each block, and reversal phase, i.e., the remaining trials; alternatively, stable phases, i.e., the first and last 35 trials, and reversal phase, i.e., the remaining trials) or between trial types (pre-reversal trials, i.e., the trials leading up to a reversal, and post reversal trials, i.e., the 5 trials directly following a reversal). We fit models including phase and trial type separately and compared them against each other as well as the simplest model

(without phase or trial-type) based on BICs. We report the results from the best fitting model (for model comparison, see Supplementary Table 1.).

*2.4. Analysis of behavior – Probabilistic selection task.*

Probabilistic stimulus selection tasks similar to the one we implemented have been analyzed in different ways. Thus, for example, Frank et al. (2004) calculated how often participants choose the best stimulus over the others and compared it to how often they avoid the worst stimulus, in novel pairs, to dissociate how well people learn from wins vs. losses. Palminteri et al., in a similar task, estimated the choice probability for each stimulus as a function of motivational context (obtain win vs. avoid loss) and whether or not the choice was "correct", i.e., the one which is more likely to lead to a win / avoid a loss (Palminteri et al., 2016). The Frank approach is "nested" in the Palminteri approach in so far as a difference between the rate at which individuals choose the best and avoid the worst stimulus would emerge as an interaction between motivational context and "correctness" or accuracy. We therefore implemented a strategy akin to Palminteri's, predicting choice rates of the known stimuli based on motivational context, accuracy, and age, adding another factor representing whether choices were between two known stimuli or between a known and a novel stimulus.

*2.5. Analysis of longitudinal development.*

Because it is conceivable that within-subject development effects differ depending on age (such that, for example, younger people change more from the first to the second assessment), we took models with significant age effects forward for further analysis in which we differentiated between cross-sectional (between-subject) age differences and longitudinal (within-subject) development. To do that, we included cross-sectional age variance (subjects' mean age across timepoints, z-scored) and longitudinal age variance (the difference between subjects' age at each time point and their individual mean age, z-scored) as separate variables in the model, where they were also allowed to interact (Neuhaus & Kalbfleisch, 1998; Vanes et al., 2020). Because these are post-hoc, confirmatory analyses, we only included predictors that significantly interacted with age in these models.

*2.6. Computational modelling of behavior.*

In order to identify individual differences in processes underlying behavior in this task, we fit 12 different reinforcement learning models based on Q-Learning (Watkins & Dayan, 1992) to the data. For detailed descriptions (including equations), please refer

to the supplement. Parameter estimation was performed using empirical Bayesian estimation in an expectation maximization procedure, implemented in MATLAB R2020b using the emfit toolbox (Huys et al., 2011, 2012; Huys & Schad, 2015) (details in the supplement). We performed model selection on the estimated models based on the integrated Bayesian Information Criterion (Huys et al., 2012) in the entire sample as well as separately for adolescents (participants ≤18) and adults (participants >18) to make sure both groups were best fit by the same model. The best model (overall and in both groups) proved to be a full double update model with separate reinforcement sensitivities ($\rho$) for positive and negative feedback, a single learning rate ($\alpha$) and a softmax decision policy (Eq. 1 through 3; p: probability, Q: expected value, a: action, t: trial).

$$p(a_i) = \frac{exp\left(Q_{a_i}\right)}{\sum_{j=1}^{K} exp\left(Q_{a_j}\right)} \qquad \text{Eq. 1}$$

$$Q_{a,t+1} = Q_{a,t} + \alpha\left(\rho_{+/-} * r - Q_{a,t}\right) \qquad \text{Eq. 2}$$

$$Q_{a_{unchosen},t+1} = Q_{a_{unchosen},t} + \alpha\left(\left(-\rho_{+/-} * r\right) - Q_{a_{unchosen},t}\right) \qquad \text{Eq. 3}$$

We took this model forward for further analysis, computing linear mixed effects models to gauge the effects of age, condition, and feedback on the fitted parameters. To ascertain that any age effects were not driven by age-related differences in model fit at chance level, we repeated all analyses excluding individuals with chance fit. As for the raw behavior, we also re-ran models differentiating between cross-sectional age-differences and longitudinal development where age effects came out significant.

*2.7. fMRI Preprocessing.*

For scanning sequences, please refer to the supplement. The fMRI data was preprocessed using SPM12 (http://www.fil.ion.ucl.ac.uk/spm/software/spm12) in MATLAB 2020b. First, the functional and structural images' origin was set to approximately the location of the anterior commissure in order to aid later co-registration and normalization. The functional images were then slice-time corrected and voxel-displacement maps were computed based on the field maps. Subsequently, they were realigned and unwarped, accounting for motion, distortion, and the interaction between motion and distortion, and spatially normalized to MNI (Montreal Neurological Institute) space based on the normalization parameters generated during

the segmentation of each participant's anatomical scan. Finally, they were smoothed using an isotropic Gaussian kernel of 8mm full width at half maximum.

*2.8. fMRI Analysis.*

Before 1st level statistical analysis, the data was high-pass filtered with a cut off at 128s. We then applied event-related analyses using the general linear model implemented in SPM12, modeling feedback onsets, cue onsets, missing trials, and the 6 movement parameters.

Parametric modulators were constructed and added to the model as follows. First, we derived, for each individual, trial-by-trial prediction errors (PEs) from the fitted computational models. To be able to differentiate the neural representation of actual and inferred (counterfactual) feedback, we computed both single and double update prediction errors. For the former, we used the single update (SU) model with separate reinforcement sensitivities for positive and negative feedback and a single learning rate (corresponding to Eq. 1 and 2, without Eq. 3 above; see supplement for details). Note that we fixed the positive reinforcement sensitivity to 1 and the negative reinforcement sensitivity to -1 to have the prediction errors on the same scale (bounded between +1 and -1), and to separate effects of the learning rate and reinforcement sensitivities. To capture the additional counterfactual information contained within prediction errors from the (winning) double update (DU) model, we generated trial-by-trial prediction errors from that model and subtracted the SU prediction errors (see Reiter et al., 2017 for a similar approach). The SU and DU prediction errors were included as orthogonalized parametric modulators on the feedback regressor. Second, we generated trial-by-trial choice probabilities for each individual based on the fitted parameters of the winning double update model. The inferred choice probability is a function of the relative expected values of the two options and can be interpreted as confidence in the upcoming choice. Third, from the choice probabilities, we constructed a control regressor reflecting trial-by-trial model-fit, where choices predicted with below-chance accuracy (<50%) were coded as 1 and 0 otherwise. The choice probabilities and model-fit regressors were included as orthogonalized parametric modulators on the cue regressor. This was done for both conditions (win and loss block) in a single model, where each block was modeled as a separate session. The regressors were convolved with the canonical hemodynamic response function in SPM12. For the second level analyses, we estimated random effects ANOVAs on the

contrast images of the parametric modulators with a condition factor (win/loss block) and a covariate reflecting age.

## 3. RESULTS

In this section, we only report significant age effects in detail. For full results tables from all models reported below, please refer to the supplement.

### 3.1. Behavior

#### 3.1.1. Accuracy.

The trial-type model differentiating between trials leading up to and following reversals best accounted for the data (Supplementary Table 1). It revealed a significant age x trial-type interaction effect (OR = 0.87, $z$ = -2.64, $p$ = .008), such that older participants tend to be more accurate in pre-reversal and less accurate in post-reversal trials than younger participants (Fig.2 – A). Simple effects analysis revealed that the interaction was driven primarily by the positive effect of age on accuracy in pre-reversal trials and less so by the negative effect of age on accuracy in post-reversal trials.
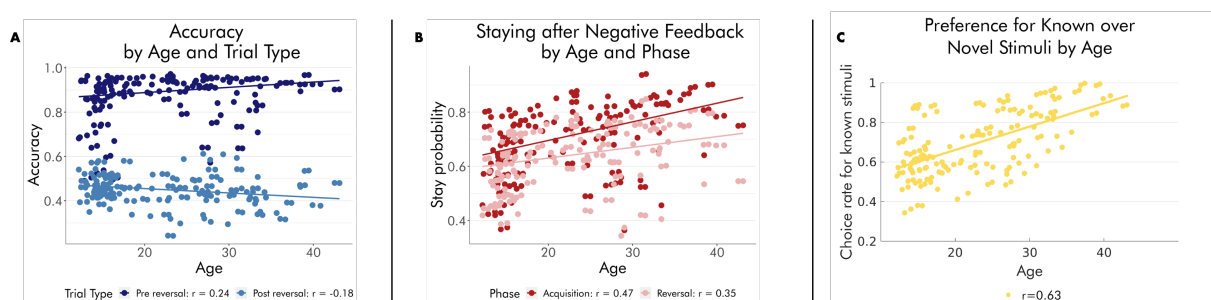


 **Fig. 2. A** – Predicted probability of choosing the more advantageous card, by age and trial-type, based on a generalized linear mixed effects model. Midnight-blue dots reflect accuracy in pre-reversal trials, steel-blue dots reflect accuracy in post-reversal trials. Correlation coefficients are between age and predicted values per trial-type. **B** – Predicted probability of staying with the same choice after negative feedback, by age and task phase, based on a generalized linear mixed effects model. Burgundy dots reflect switching in the acquisition phase of the task, rose dots reflect switching in the reversal phase. Correlation coefficients are between age and predicted values for each task phase. **C** – Predicted choice rates in the probabilistic selection task following the PRLT for known over novel stimuli, based on a linear mixed effects model. In all plots, there are up to two dots per person and color: one reflecting the initial session, one the follow up session (where data was available).

#### 3.1.2. Stay-Switch-Behavior.

The phase model differentiating between the acquisition phase and the remainder of trials best accounted for the data (Supplementary Table 1). It revealed a four-way

interaction between age, phase, condition, and previous feedback (OR = 0.95, $z$ = - 2.95, $p$ = .003) in addition to a three-way interaction between age, phase, and condition (OR = 0.96, $z$ = -2.11, $p$ = .04), a three-way interaction between age, previous feedback and phase (OR = 1.05, $z$ = 2.76, $p$ = .006), and a main effect of age (OR = 1.22, $z$ = 2.17, $p$ = .03). Unpacking this, simple effects analyses showed an age by phase interaction for staying after negative feedback, which was driven by a stronger positive effect of age in the acquisition than the reversal phase (Fig.2 – B). There were no condition-specific age effects on staying after negative feedback. Further simple effects analyses looking at staying after positive feedback showed that age had no effect during the acquisition phase of the loss condition and only marginal effects in the other conditions and phases (s. Supplementary Fig.1 and associated tables). Given that the effect of age on staying after positive feedback was not significant in any phase or condition, we refrain from interpreting it.

### 3.1.3. Explorative – reaction times.

The phase model differentiating between acquisition phase and the remainder of trials best accounted for the data (Supplementary Table 1). It revealed an interaction between age and previous feedback ($\beta$ = -0.03, $t$(45897) = -3.32, $p$ < .001), such that older participants responded faster than younger participants, in particular after positive feedback.

### 3.1.4. Probabilistic selection task.

Our model predicting the choice rate for each stimulus based on motivational context (i.e., win or loss stimulus in the PRLT), accuracy (i.e., better or worse stimulus in the PRLT), pairing (choice against a known or a novel stimulus) and age showed a main effect of motivational context ($\beta$ = 0.05, $t$(1288) = 4.203, $p$ < .001), such that participants more often chose stimuli from the win than from the loss block, as well as an interaction between age and pairing ($\beta$ = 0.0474, $t$(1288) = 5.0189, $p$ < .001), such that when faced with a known and an novel stimulus, younger participants more often chose the novel stimulus (regardless of motivational context) than older participants (s. Fig.2 – C).

### 3.2. Computational Modelling

### 3.2.1. RL model selection.

A full double update model with separate reinforcement sensitivities for positive and negative feedback and a single learning rate had the best evidence (lowest integrated BIC = 30,261, distance to next lowest ΔBIC = 209) across the whole sample, as well as in adolescents (participants ≤18) and adults (participants >18) considered separately (Supplementary Figures 3 through 5). This model updates the values for the chosen and unchosen options to the same extent (double update) and equally fast after positive and negative feedback (single learning rate), but allows for differential impact of positive and negative feedback on expected values and choices (separate reinforcement sensitivities for positive and negative feedback).

*3.2.2. RL parameters – reinforcement sensitivity.*

A linear mixed effects model predicting reinforcement sensitivity parameter values from age, feedback and condition revealed an interaction between age and previous feedback ($\beta$ = 0.11, $t$(672) = 2.5862, $p$ = .01), such that older participants were more sensitive to positive but not negative feedback. This did not change when we excluded individuals fit at or below chance level. In an explorative analysis, we compared parameters from the single and double update models, showing that under the single update model, there was an age effect on both positive and negative feedback sensitivity. This suggests that the specificity to positive feedback is a feature of the double update model (refer to the supplement for details).
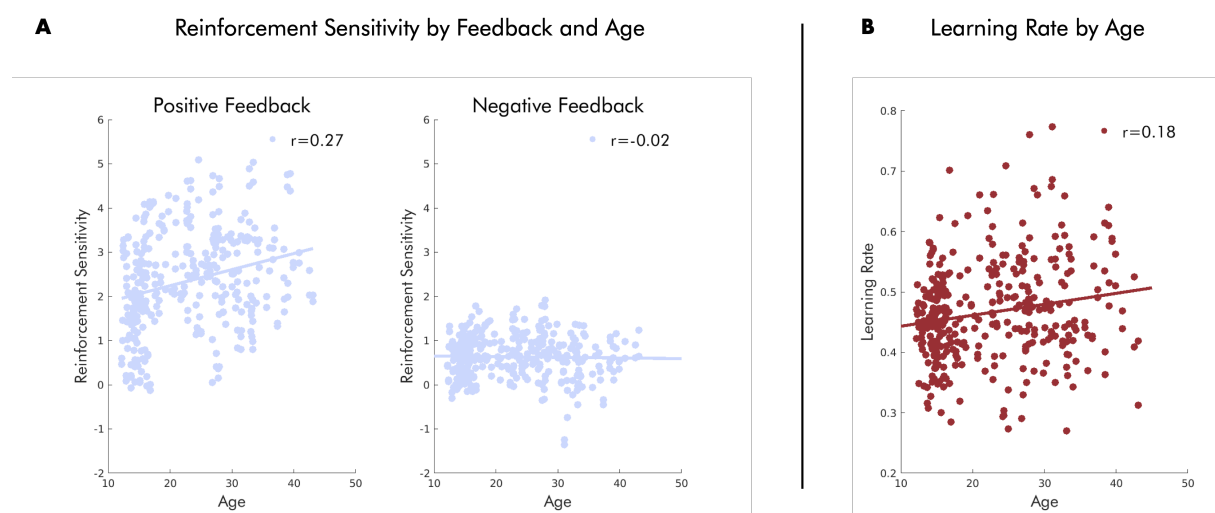


**Fig. 3. A** – Reinforcement sensitivity (averaged across conditions), by age and feedback. There are up to two dots per plot and person: one reflecting the learning rate in the initial session, one in the follow up session (where data was available). Left panel: reinforcement sensitivity for positive feedback; right panel: reinforcement sensitivity for negative feedback. The correlation coefficients reflect the

relationship between sensitivity values and age. **B** – Learning rate (averaged across conditions), by age. The correlation coefficient reflects the relationship between learning rates and age

### 3.2.3. RL parameters – learning rate.

A linear mixed effects model predicting the learning rate from age and condition revealed no effects of age (all $p > .39$). This did not change when we excluded individuals fit at or below chance level.

### 3.2.4. RL recovery and posterior predictive checks.

In order to ensure that the model fit our subjects' behavior well on a qualitative level, we simulated 100 datasets based on the fitted model parameters of each subject. The recovered data generally captured the participants' parameters well and reproduced the observed effects of age (Supplementary Figures 6 and 7).

### 3.3. Differential contributions of within- and between-subject development.

Given that within-subject development effects may differ depending on age, we repeated all our analyses differentiating between (cross-sectional) age-differences and (longitudinal) development. The results suggest that the age effects reported above were driven primarily by cross-sectional variance (for detailed results, please refer to the supplement).

### 3.4. fMRI

### 3.4.1. Prediction error coding.

As expected based on previous studies (e.g., Abler et al., 2006; McClure et al., 2004; O'Doherty et al., 2007), participants showed robust correlations between prediction errors derived from the single update model and BOLD signals in the striatum at the group level (Fig. 4, top row, full results tables in supplement). Prediction errors incorporating counterfactual information (derived from the double-update model) were coded mostly in the vmPFC, hippocampus and PCC (Fig. 4, middle row, full results tables in supplement). There was no evidence of age differences in single or double update prediction error coding, and neither changed depending on condition (motivational context).

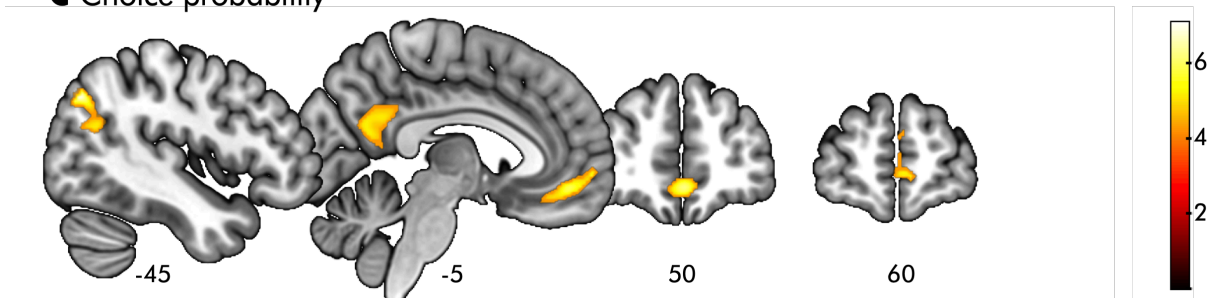# Group-level effects of trial-by-trial regressors derived from computational modelling

**A** Single update prediction error
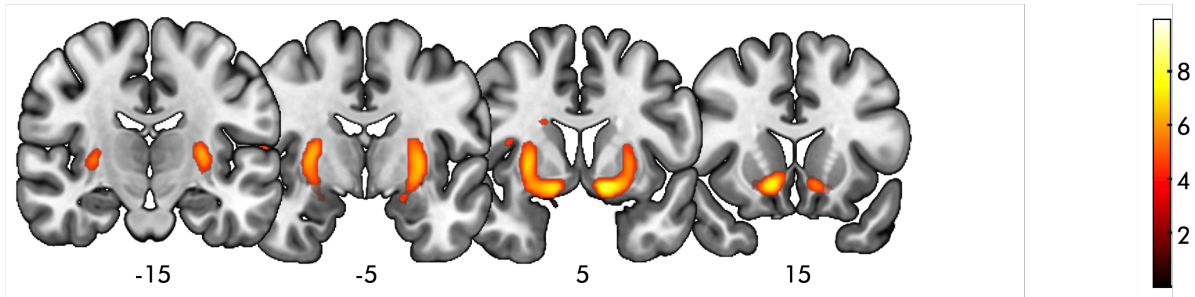


**B** Double update prediction error - single update prediction error
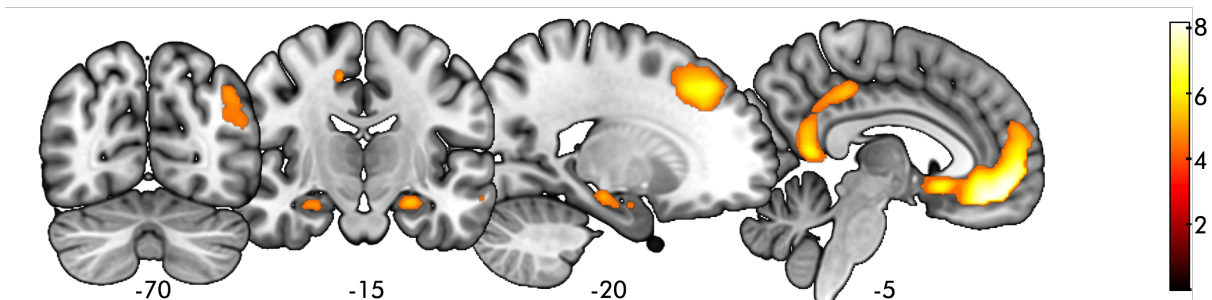


**C** Choice probability



**Fig. 4.** Group-level (positive) effects of regressors derived from computational modelling. **A** – prediction errors derived from a single update RL-model. **B** – additional (counterfactual) information incorporated in double update prediction errors (calculated as the difference between PEs derived from the double and single update models). **C** – choice probability as derived from the double update model. All maps are thresholded at $p_{FWE} < .05$ (no minimum cluster size). Blob colors represent t-values.

### 3.4.2. fMRI – Choice probability coding.

At the group level, trial-by-trial choice probability was correlated positively with BOLD signal in the (v)mPFC and PCC (Fig. 4, bottom row, full results tables in supplement). Choosing an option not predicted by our reinforcement learning model (predicted choice probability of <=.5) was correlated with activation in the bilateral insula, dorsal

ACC, and premotor areas (full results tables in supplement). There were no effects of condition on the neural coding of either choice probability or trial-by-trial model fit.

The positive neural correlate of choice probability (i.e., chosing the option that was suggested to be the better one by the DU model) was modulated by age in the medial prefrontal cortex/frontal pole, such that older participants showed stronger neural representation of choice probability in this area (Fig. 5 – A, [-4,64,12], k=17, $t$ = 3.51, $p_{FWE}$ = .03, small-volume corrected using the group-level activation map). Next, we looked at brain-behaviour relationships related to this age effect of choice probability coding in the medial frontal pole. Indeed, parameter values extracted at the peak coordinate correlated significantly with staying after negative feedback and preference for known over novel stimuli in the post-task test (Fig. 5 – B).

**A**  Positive Effect of Age on Neural Correlate of Choice Probability in the Medial Frontopolar Cortex
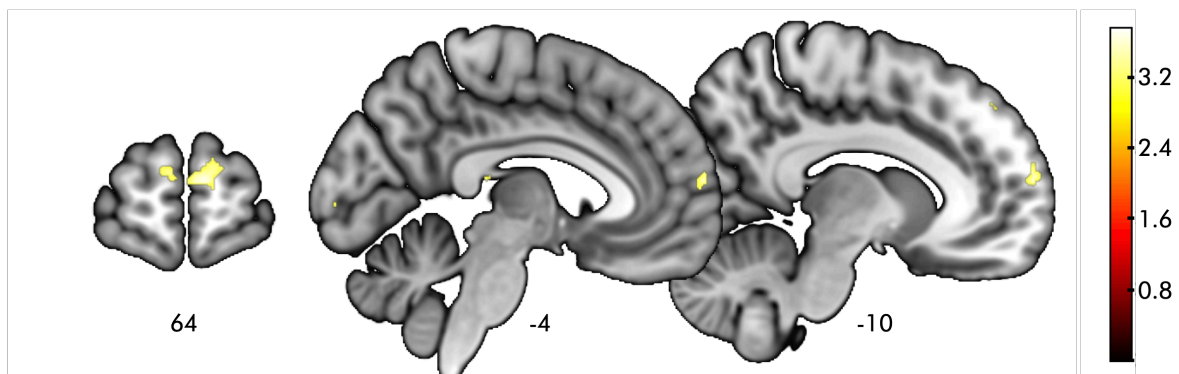


**B**  Correlations between Neural Correlate of Choice Probability in the Medial Frontopolar Cortex and Task Behavior
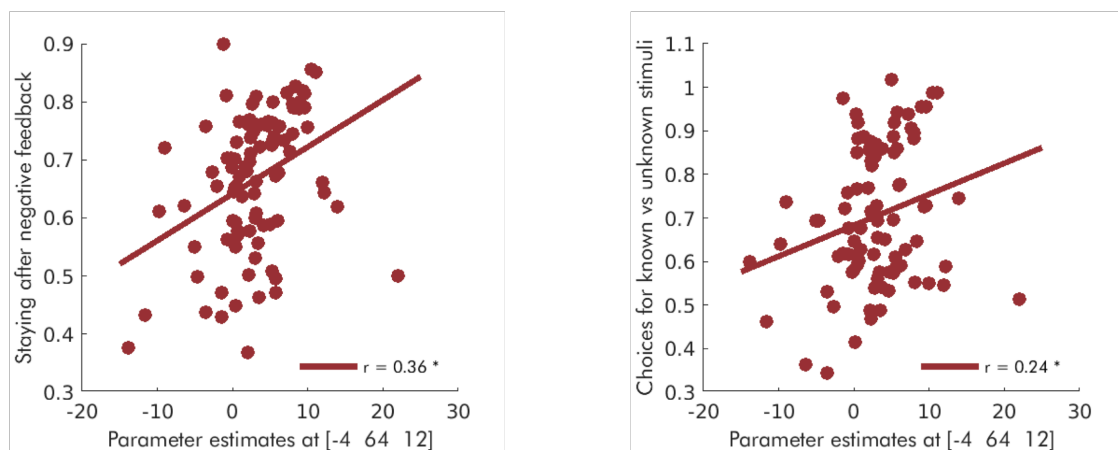


**Fig. 5. A** – Association between the positive neural correlates of model-predicted choice probability and age. Blob colors represent t-contrast values, thresholded at p<.001 uncorrected for visualization. **B** –

left panel. Correlation between parameters at the cluster peak ([-4,64,12]) and staying after negative feedback. **B** – right panel. Correlation between parameters at the cluster peak ([-4,64,12]) and choices for known over unknown stimuli in the post-task test. Outliers > 3 standard deviations from the mean were removed before plotting.


## 4. DISCUSSION

In this study, we show that performance during stable phases of the probabilistic reversal learning task, i.e., prior to reversals, improves linearly with age. Our results indicate that this is driven by excessive response switching following negative feedback. Computationally, this could be accounted for by lower sensitivity to positive feedback in younger participants: thus, in younger participants, positive feedback had less of an impact on the expected values of the two choice options (and the difference between them), such that negative feedback in subsequent trials induced switching more readily. In the brain, there was no evidence of differences in reward prediction error coding between adolescents and adults. However, reduced sensitivity to positive feedback was reflected in diminished activation of the medial frontopolar cortex as a function of choice probability in youths. Interestingly, we found no age-related differences between learning in win and loss contexts, nor differences in the extent to which adolescents and adults used inferred counterfactual feedback, in either behavior or fMRI.

Our behavioral results are in line with evidence showing similarly enhanced switching (less win/stay and/or more lose/shift behavior) (Crawley et al., 2020; Javadi et al., 2014; Van Den Bos et al., 2009) and greater choice stochasticity/reduced feedback sensitivity (Christakou et al., 2013; Crawley et al., 2020; Decker et al., 2015; Javadi et al., 2014; Moutoussis et al., 2021; Rodriguez Buritica et al., 2019; although see Davidow et al., 2016) in younger (adolescent) individuals. We extend this literature by differentiating between sensitivity to positive and negative feedback. Thus, we provide evidence that enhanced switching behavior might be computationally accounted for by insufficient sensitivity to positive feedback rather than enhanced sensitivity to negative feedback or overall lower feedback sensitivity. This interpretation is supported by our explorative analysis of reaction times: congruent with previous research (Decker et al., 2016b; Eckstein et al., 2021), it shows that younger participants respond more slowly than older participants, especially after positive feedback. According to drift-diffusion accounts (McDougle & Collins, 2021; Mormann et al., 2010; Pedersen et al., 2017), it

takes longer to sample noisy information. Hence, this may be indicative of relatively elevated uncertainty as to the value of choice options our younger participants, which has previously been shown to decrease across adolescence (Reiter et al., 2021).

Moreover, we point to a neural correlate of these behavioral effects, showing reduced coding of trial-by-trial choice probability in the medial frontopolar cortex in youth. This signal can be read as confidence in an upcoming choice and was associated with key behavioral readouts, i.e., switching after negative feedback and the relative preference for known over novel stimuli in a post-task test. The medial frontopolar cortex has previously been implicated in tracking choice probabilities (Daw et al., 2006). It has been proposed to be involved in arbitrating between exploration and exploitation, specifically by monitoring the relative value of current behavior and triggering exploration (Mansouri et al., 2017). In line with this role, the medial PFC's connectivity has been shown to be associated with choice stochasticity (Moutoussis et al., 2021). In this sense, the involvement of this region supports our interpretation that reduced sensitivity to feedback, and consequently relative value, might drive adolescent over-switching in the PRLT. Importantly, this region and its connectivity, along with other regions of the PFC, are known to mature substantially and asymmetrically relative to subcortical structures in adolescence (Casey et al., 2008; Dahl et al., 2018; Dumontheil et al., 2008).

Our behavioral findings may thus be associated with the stage of development of the adolescent brain (although we do not explicitly test this). Alternatively or epiphenomenally, it might reflect an adaptive response to adolescents' specific (social) environment. Thus, adolescents' choice behavior may be uniquely adapted towards navigating environments full of novel stimuli and volatile affordances (Eckstein et al., 2021; Hartley & Somerville, 2015). Reduced sensitivity to positive feedback allows for rapid and flexible responses in case reward contingencies change or new opportunities arise. In our task, this is not always helpful as most trials occur in relatively stable phases, where exploration comes at a steep performance cost. But the (social) environment of youths might be (perceived as) one in which reward contingencies arise and change rapidly and unpredictably. In such environments, exploration and continuous readiness to modify behavior is the most optimal course of action. Consistent with this, younger participants more frequently chose new over known stimuli (regardless of whether the known stimuli were win or loss stimuli) in a post-task test, indicating increased novelty-seeking.

Interestingly, our analysis suggested that both adolescents' and adults' task behavior was best fit by a model incorporating full counterfactual inference. This is somewhat surprising, since counterfactual learning relies on the utilization of inferred knowledge about the environment, which has been found to increase from adolescence to adulthood (Decker et al., 2016a; Palminteri et al., 2016). In addition, this process is thought to primarily recruit prefrontal brain structures, which are known to exhibit protracted development well into adulthood (Casey et al., 2008). At the same time, two previous studies on probabilistic reversal learning in youths similarly reported model selection favoring double update models (Eckstein et al., 2021; Hauser et al., 2015; but see Boehme et al., 2017 for evidence of effects of pubertal status). This suggests that comparatively simple counterfactual inference might already be nearly fully functional in adolescents, even though they might not always be able to optimally use it. In the future, more sophisticated methods to investigate counterfactual learning (e.g., Boorman et al., 2011; Li & Daw, 2011) may be helpful to precisely characterize its development.

Contrary to our hypotheses, we found no differential effects of motivational context across the age range. Instead, our data suggests that participants of all ages found the win condition "easier". Thus, participants switched less and responded more quickly in the win condition than the loss condition. In line with this, the computational modelling showed clear condition effects on both the reinforcement sensitivities and the learning rate, such that parameters were more optimal (more extreme sensitivities and learning rates) in the win condition. The observed absence of interactions between age and motivational context is somewhat at odds with evidence of enhanced reward sensitivity in adolescents (Somerville et al., 2010; Somerville & Casey, 2010) as well as previous evidence of altered performance in loss contexts (Palminteri et al., 2016; although see Bolenz & Eppinger, 2022). It is possible that such effects are subtle and our study was insufficiently powered to detect them; alternatively, heighted reward sensitivity in adolescence might not straightforwardly translate to differential learning from wins and losses. Further studies disentangling feedback valence and motivational context will be needed to clarify this point.

Differential analyses looking into interactions between age and longitudinal change showed little within-subject development and suggested that the age effects we did observe primarily stemmed from between-subject differences. Future studies should attempt to distill within-subjects development and its interaction with age by sampling

from a narrower age range (e.g., Ziegler et al., 2019) and/or extending the follow up interval.

In conclusion, the current study adds to a growing body of evidence showing that the development of reinforcement learning from adolescence to adulthood is characterized by decreasing novelty seeking and response shifting, especially after negative feedback, leading to poorer returns in environments with stable reward contingencies in youths. We show that this can be computationally accounted for by increasing sensitivity to positive feedback. The behavioral effects were linked to diminished activity of the medial frontopolar cortex reflecting trial-by-trial choice probability in adolescents, putatively reflecting confidence in the upcoming choice. Future studies should further elucidate the exact time course and the drivers of normative RL development, both proximal (what are the underlying cognitive processes?) and ontogenic (what are the underlying psychobiological maturation processes?), to identify vulnerable periods in which disruption could cause future mental health problems.

## 5. Data and Code availability

The raw behavioral data and analysis scripts underlying the analyses in this article are available on the Open Science Framework (https://osf.io/ptxs6).

## 6. Acknowledgements and Declarations

### 6.2. Author contributions

LD and AH designed the study and acquired funding; NH and MW acquired the data; MW and LD analyzed the data and wrote the original draft; MW, NH, AH, AV, AR, LD reviewed and edited the manuscript.

### 6.4. Declaration of interests

MW, NH, AH, AV, AR and LD declare no biomedical financial interests or potential conflicts of interest.

## 7. REFERENCES

Abler, B., Walter, H., Erk, S., Kammerer, H., & Spitzer, M. (2006). Prediction error as a linear function of reward probability is coded in human nucleus accumbens. *NeuroImage*, *31*(2), 790–795. https://doi.org/https://doi.org/10.1016/j.neuroimage.2006.01.001

Barkley-Levenson, E., & Galván, A. (2014). Neural representation of expected value in the adolescent brain. *Proceedings of the National Academy of Sciences*, *111*(4), 1646–1651. https://doi.org/10.1073/pnas.1319762111

Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 10.1016/j.jml.2012.11.001. https://doi.org/10.1016/j.jml.2012.11.001

Bjork, J. M., Knutson, B., Fong, G. W., Caggiano, D. M., Bennett, S. M., & Hommer, D. W. (2004). Incentive-elicited brain activation in adolescents: similarities and differences from young adults. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *24*(8), 1793–1802. https://doi.org/10.1523/JNEUROSCI.4862-03.2004

Boehme, R., Deserno, L., Gleich, T., Katthagen, T., Pankow, A., Behr, J., Buchert, R., Roiser, J. P., Heinz, A., & Schlagenhauf, F. (2015). Aberrant Salience Is Related to Reduced Reinforcement Learning Signals and Elevated Dopamine Synthesis Capacity in Healthy Adults. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience*, *35*(28), 10103–10111. https://doi.org/10.1523/JNEUROSCI.0805-15.2015

Boehme, R., Lorenz, R. C., Gleich, T., Romund, L., Pelz, P., Golde, S., Flemming, E., Wold, A., Deserno, L., Behr, J., Raufelder, D., Heinz, A., & Beck, A. (2017). Reversal learning strategy in adolescence is associated with prefrontal cortex activation. *European Journal of Neuroscience*, *45*(1), 129–137. https://doi.org/10.1111/ejn.13401

Bolenz, F., & Eppinger, B. (2022). Valence bias in metacontrol of decision making in adolescents and young adults. *Child Development*, *93*(2), e103–e116. https://doi.org/https://doi.org/10.1111/cdev.13693

Bolenz, F., Reiter, A. M. F., & Eppinger, B. (2017). Developmental Changes in Learning: Computational Mechanisms and Social Influences . In *Frontiers in Psychology* (Vol. 8). https://www.frontiersin.org/articles/10.3389/fpsyg.2017.02048

Boorman, E. D., Behrens, T. E., & Rushworth, M. F. (2011). Counterfactual Choice and Learning in a Neural Network Centered on Human Lateral Frontopolar Cortex. *PLOS Biology*, *9*(6), e1001093. https://doi.org/10.1371/journal.pbio.1001093

Casey, B. J., Jones, R. M., & Hare, T. A. (2008). The adolescent brain. *Annals of the New York Academy of Sciences*, *1124*, 111–126. https://doi.org/10.1196/annals.1440.010

Christakou, A., Gershman, S. J., Niv, Y., Simmons, A., Brammer, M., & Rubia, K. (2013). Neural and Psychological Maturation of Decision-making in Adolescence and Young Adulthood. *Journal of Cognitive Neuroscience*, *25*(11), 1807–1823. https://doi.org/10.1162/jocn_a_00447

Cohen, J. R., Asarnow, R. F., Sabb, F. W., Bilder, R. M., Bookheimer, S. Y., Knowlton, B. J., & Poldrack, R. A. (2010). A unique adolescent response to reward prediction errors. *Nature Neuroscience*, *13*(6), 669–671. https://doi.org/10.1038/nn.2558

Crawley, D., Zhang, L., Jones, E. J. H., Ahmad, J., Oakley, B., San José Cáceres, A., Charman, T., Buitelaar, J. K., Murphy, D. G. M., Chatham, C., den Ouden, H., Loth, E., & group,  the E.-A. L. (2020). Modeling flexible behavior in childhood to adulthood shows age-dependent learning mechanisms and less optimal learning in autism in each age group. *PLOS Biology*, *18*(10), e3000908. https://doi.org/10.1371/journal.pbio.3000908

Dahl, R. E., Allen, N. B., Wilbrecht, L., & Suleiman, A. B. (2018). Importance of investing in adolescence from a developmental science perspective. *Nature*, *554*(7693), 441–450. https://doi.org/10.1038/nature25770

Davidow, J. Y., Foerde, K., Galván, A., & Shohamy, D. (2016). An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. *Neuron*, *92*(1), 93–99. https://doi.org/10.1016/j.neuron.2016.08.031

Daw, N. D., O'Doherty, J. P., Dayan, P., Seymour, B., & Dolan, R. J. (2006). Cortical substrates for exploratory decisions in humans. *Nature*, *441*(7095), 876–879. https://doi.org/10.1038/nature04766

Decker, J. H., Lourenco, F. S., Doll, B. B., & Hartley, C. A. (2015). Experiential reward learning outweighs instruction prior to adulthood. *Cognitive, Affective, & Behavioral Neuroscience*, *15*(2), 310–320. https://doi.org/10.3758/s13415-014-0332-5

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016a). From Creatures of Habit to Goal-Directed Learners: Tracking the Developmental Emergence of Model-Based Reinforcement Learning. *Psychological Science*, *27*(6), 848–858. https://doi.org/10.1177/0956797616639301

Decker, J. H., Otto, A. R., Daw, N. D., & Hartley, C. A. (2016b). From Creatures of Habit to Goal-Directed Learners. *Psychological Science*, *27*(6), 848–858. https://doi.org/10.1177/0956797616639301

Deserno, L., Boehme, R., Mathys, C., Katthagen, T., Kaminski, J., Stephan, K. E., Heinz, A., & Schlagenhauf, F. (2020). Volatility Estimates Increase Choice Switching and Relate to Prefrontal Activity in Schizophrenia. *Biological Psychiatry: Cognitive Neuroscience and Neuroimaging*, *5*(2), 173–183. https://doi.org/https://doi.org/10.1016/j.bpsc.2019.10.007

Dumontheil, I., Burgess, P. W., & Blakemore, S.-J. (2008). Development of rostral prefrontal cortex and cognitive and behavioural disorders. *Developmental Medicine and Child Neurology*, *50*(3), 168–181. https://doi.org/10.1111/j.1469-8749.2008.02026.x

Eckstein, M. K., Master, S. L., Dahl, R. E., Wilbrecht, L., & Collins, A. G. E. (2021). Reinforcement Learning and Bayesian Inference Provide Complementary Models for the Unique Advantage of Adolescents in Stochastic Reversal. *BioRxiv*, 2020.07.04.187971. https://doi.org/10.1101/2020.07.04.187971

Ernst, M., Nelson, E. E., Jazbec, S., McClure, E. B., Monk, C. S., Leibenluft, E., Blair, J., & Pine, D. S. (2005). Amygdala and nucleus accumbens in responses to

receipt and omission of gains in adults and adolescents. *NeuroImage*, *25*(4), 1279–1291. https://doi.org/10.1016/j.neuroimage.2004.12.038

Frank, M. J., Seeberger, L. C., & Reilly, R. C. O. (2004). *By Carrot or by Stick : Cognitive Reinforcement Learning in Parkinsonism. December*, 1940–1943.

Galvan, A., Hare, T. A., Parra, C. E., Penn, J., Voss, H., Glover, G., & Casey, B. J. (2006). Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *The Journal of Neuroscience : The Official Journal of the Society for Neuroscience, 26*(25), 6885–6892. https://doi.org/10.1523/JNEUROSCI.1062-06.2006

Hartley, C. A., & Somerville, L. H. (2015). The neuroscience of adolescent decision-making. *Current Opinion in Behavioral Sciences, 5*, 108–115. https://doi.org/10.1016/j.cobeha.2015.09.004

Hauser, T. U., Iannaccone, R., Ball, J., Mathys, C., Brandeis, D., Walitza, S., & Brem, S. (2014). Role of the Medial Prefrontal Cortex in Impaired Decision Making in Juvenile Attention-Deficit/Hyperactivity Disorder. *JAMA Psychiatry, 71*(10), 1165–1173. https://doi.org/10.1001/jamapsychiatry.2014.1093

Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D., & Brem, S. (2015). Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *NeuroImage, 104*, 347–354. https://doi.org/https://doi.org/10.1016/j.neuroimage.2014.09.018

Huys, Q. J. M., Cools, R., Gölzer, M., Friedel, E., Heinz, A., Dolan, R. J., & Dayan, P. (2011). Disentangling the Roles of Approach, Activation and Valence in Instrumental and Pavlovian Responding. *PLOS Computational Biology, 7*(4), e1002028. https://doi.org/10.1371/journal.pcbi.1002028

Huys, Q. J. M., Eshel, N., O'Nions, E., Sheridan, L., Dayan, P., & Roiser, J. P. (2012). Bonsai Trees in Your Head: How the Pavlovian System Sculpts Goal-Directed Choices by Pruning Decision Trees. *PLOS Computational Biology, 8*(3), e1002410. https://doi.org/10.1371/journal.pcbi.1002410

Huys, Q. J. M., & Schad, D. (2015). *No Title*. Emfit Matlab Script. https://github.com/mpc-ucl/emfit

Javadi, A. H., Schmidt, D. H. K., & Smolka, M. N. (2014). Adolescents Adapt More Slowly than Adults to Varying Reward Contingencies. *Journal of Cognitive Neuroscience, 26*(12), 2670–2681. https://doi.org/10.1162/jocn_a_00677

Jones, R. M., Somerville, L. H., Li, J., Ruberry, E. J., Powers, A., Mehta, N., Dyke, J., & Casey, B. J. (2014). Adolescent-specific patterns of behavior and neural activity during social reinforcement learning. *Cognitive, Affective, & Behavioral Neuroscience, 14*(2), 683–697. https://doi.org/10.3758/s13415-014-0257-z

Kessler, R. C., Amminger, G. P., Aguilar-Gaxiola, S., Alonso, J., Lee, S., & Ustün, T. B. (2007). Age of onset of mental disorders: a review of recent literature. *Current Opinion in Psychiatry, 20*(4), 359–364. https://doi.org/10.1097/YCO.0b013e32816ebc8c

Li, J., & Daw, N. D. (2011). Signals in Human Striatum Are Appropriate for Policy Update Rather than Value Prediction. *The Journal of Neuroscience, 31*(14), 5504 LP – 5511. https://doi.org/10.1523/JNEUROSCI.6316-10.2011

Mansouri, F. A., Koechlin, E., Rosa, M. G. P., & Buckley, M. J. (2017). Managing

competing goals — a key role for the frontopolar cortex. *Nature Reviews Neuroscience*, *18*(11), 645–657. https://doi.org/10.1038/nrn.2017.111

McClure, S. M., York, M. K., & Montague, P. R. (2004). The Neural Substrates of Reward Processing in Humans: The Modern Role of fMRI. *The Neuroscientist*, *10*(3), 260–268. https://doi.org/10.1177/1073858404263526

McDougle, S. D., & Collins, A. G. E. (2021). Modeling the influence of working memory, reinforcement, and action uncertainty on reaction time and choice during instrumental learning. *Psychonomic Bulletin & Review*, *28*(1), 20–39. https://doi.org/10.3758/s13423-020-01774-z

Mormann, M. M., Malmaud, J., Huth, A., Koch, C., & Rangel, A. (2010). The drift diffusion model can account for the accuracy and reaction time of value-based choices under high and low time pressure. *Judgment and Decision Making*, *5*(6), 437–449.

Moutoussis, M., Garzón, B., Neufeld, S., Bach, D. R., Rigoli, F., Goodyer, I., Bullmore, E., Guitart-Masip, M., & Dolan, R. J. (2021). Decision-making ability, psychopathology, and brain connectivity. *Neuron*, *109*(12), 2025-2040.e7. https://doi.org/10.1016/j.neuron.2021.04.019

Mukherjee, D., Filipowicz, A. L. S., Vo, K., Satterthwaite, T. D., & Kable, J. W. (2020). Reward and punishment reversal-learning in major depressive disorder. *Journal of Abnormal Psychology*, *129*(8), 810–823. https://doi.org/10.1037/abn0000641

Neuhaus, J. M., & Kalbfleisch, J. D. (1998). Between- and within-cluster covariate effects in the analysis of clustered data. *Biometrics*, *54*(2), 638–645.

Nussenbaum, K., & Hartley, C. A. (2019). Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience*, *40*, 100733. https://doi.org/10.1016/j.dcn.2019.100733

O'Doherty, J. P., Hampton, A., & Kim, H. (2007). Model-Based fMRI and Its Application to Reward Learning and Decision Making. *Annals of the New York Academy of Sciences*, *1104*(1), 35–53. https://doi.org/https://doi.org/10.1196/annals.1390.022

Palminteri, S., Kilford, E. J., Coricelli, G., & Blakemore, S.-J. (2016). The Computational Development of Reinforcement Learning during Adolescence. *PLOS Computational Biology*, *12*(6), e1004953. https://doi.org/10.1371/journal.pcbi.1004953

Paus, T., Keshavan, M., & Giedd, J. N. (2008). Why do many psychiatric disorders emerge during adolescence? *Nature Reviews Neuroscience*, *9*(12), 947–957. https://doi.org/10.1038/nrn2513

Pedersen, M. L., Frank, M. J., & Biele, G. (2017). The drift diffusion model as the choice rule in reinforcement learning. *Psychonomic Bulletin & Review*, *24*(4), 1234–1251. https://doi.org/10.3758/s13423-016-1199-y

Reddy, L. F., Waltz, J. A., Green, M. F., Wynn, J. K., & Horan, W. P. (2016). Probabilistic Reversal Learning in Schizophrenia: Stability of Deficits and Potential Causal Mechanisms. *Schizophrenia Bulletin*, *42*(4), 942–951. https://doi.org/10.1093/schbul/sbv226

Reitan, R. M. (1958). Validity of the Trail Making Test as an Indicator of Organic Brain Damage. *Perceptual and Motor Skills*, *8*(3), 271–276.

https://doi.org/10.2466/pms.1958.8.3.271

Reiter, A. M. F., Deserno, L., Kallert, T., Heinze, H.-J., Heinz, A., & Schlagenhauf, F. (2016). Behavioral and Neural Signatures of Reduced Updating of Alternative Options in Alcohol-Dependent Patients during Flexible Decision-Making. *Journal of Neuroscience*, *36*(43), 10935–10948. https://doi.org/10.1523/JNEUROSCI.4322-15.2016

Reiter, A. M. F., Heinze, H. J., Schlagenhauf, F., & Deserno, L. (2017). Impaired Flexible Reward-Based Decision-Making in Binge Eating Disorder: Evidence from Computational Modeling and Functional Neuroimaging. *Neuropsychopharmacology*, *42*(3), 628–637. https://doi.org/10.1038/npp.2016.95

Reiter, A. M. F., Moutoussis, M., Vanes, L., Kievit, R., Bullmore, E. T., Goodyer, I. M., Fonagy, P., Jones, P. B., & Dolan, R. J. (2021). Preference uncertainty accounts for developmental effects on susceptibility to peer influence in adolescence. *Nature Communications*, *12*(1), 3823. https://doi.org/10.1038/s41467-021-23671-2

Robinson, O. J., Cools, R., Carlisi, C. O., Sahakian, B. J., & Drevets, W. C. (2012). Ventral striatum response during reward and punishment reversal learning in unmedicated major depressive disorder. *The American Journal of Psychiatry*, *169*(2), 152–159. https://doi.org/10.1176/appi.ajp.2011.11010137

Rodriguez Buritica, J. M., Heekeren, H. R., & van den Bos, W. (2019). The computational basis of following advice in adolescents. *Journal of Experimental Child Psychology*, *180*, 39–54. https://doi.org/https://doi.org/10.1016/j.jecp.2018.11.019

Rosenbaum, G. M., Grassie, H. L., & Hartley, C. A. (2022). Valence biases in reinforcement learning shift across adolescence and modulate subsequent memory. *ELife*, *11*, e64620. https://doi.org/10.7554/eLife.64620

Schlagenhauf, F., Huys, Q. J. M., Deserno, L., Rapp, M. A., Beck, A., Heinze, H.-J., Dolan, R., & Heinz, A. (2014). Striatal dysfunction during reversal learning in unmedicated schizophrenia patients. *NeuroImage*, *89*, 171–180. https://doi.org/https://doi.org/10.1016/j.neuroimage.2013.11.034

Schmidt, K.-H., & Metzler, P. (1992). *Wortschatztest: WST*. Beltz.

Schreuders, E., Braams, B. R., Blankenstein, N. E., Peper, J. S., Güroğlu, B., & Crone, E. A. (2018). Contributions of Reward Sensitivity to Ventral Striatum Activity Across Adolescence and Early Adulthood. *Child Development*, *89*(3), 797–810. https://doi.org/10.1111/cdev.13056

Smith, P., Benzina, N., Vorspan, F., Mallet, L., & N'Diaye, K. (2015). Compulsivity and probabilistic reversal learning in OCD and cocaine addiction. *European Psychiatry*, *30*(S2), S110–S111. https://doi.org/DOI: 10.1016/j.eurpsy.2015.09.210

Somerville, L. H., & Casey, B. J. (2010). Developmental neurobiology of cognitive control and motivational systems. *Current Opinion in Neurobiology*, *20*(2), 236–241. https://doi.org/https://doi.org/10.1016/j.conb.2010.01.006

Somerville, L. H., Jones, R. M., & Casey, B. J. (2010). A time of change: behavioral and neural correlates of adolescent sensitivity to appetitive and aversive environmental cues. *Brain and Cognition*, *72*(1), 124–133.

https://doi.org/10.1016/j.bandc.2009.07.003

Tezcan, D., Tumkaya, S., & Bora, E. (2017). Reversal learning in patients with obsessive-compulsive disorder (OCD) and their unaffected relatives: Is orbitofrontal dysfunction an endophenotype of OCD? *Psychiatry Research*, *252*, 231–233. https://doi.org/https://doi.org/10.1016/j.psychres.2017.03.001

van den Bos, W., Cohen, M. X., Kahnt, T., & Crone, E. A. (2012). Striatum–Medial Prefrontal Cortex Connectivity Predicts Developmental Changes in Reinforcement Learning. *Cerebral Cortex*, *22*(6), 1247–1255. https://doi.org/10.1093/cercor/bhr198

Van Den Bos, W., Güroğlu, B., Van Den Bulk, B., Rombouts, S., & Crone, E. (2009). Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing . In *Frontiers in Human Neuroscience* (Vol. 3). https://www.frontiersin.org/article/10.3389/neuro.09.052.2009

Vanes, L. D., Moutoussis, M., Ziegler, G., Goodyer, I. M., Fonagy, P., Jones, P. B., Bullmore, E. T., Consortium, N., & Dolan, R. J. (2020). White matter tract myelin maturation and its association with general psychopathology in adolescence and early adulthood. *Human Brain Mapping*, *41*(3), 827–839. https://doi.org/https://doi.org/10.1002/hbm.24842

Watkins, C. J. C. H., & Dayan, P. (1992). Q-learning. *Machine Learning*, *8*(3), 279–292. https://doi.org/10.1007/BF00992698

Wechsler, D. (2008). Wechsler adult intelligence scale–Fourth Edition (WAIS–IV). *San Antonio, TX: NCS Pearson*, *22*, 498.

Ziegler, G., Hauser, T. U., Moutoussis, M., Bullmore, E. T., Goodyer, I. M., Fonagy, P., Jones, P. B., Lindenberger, U., & Dolan, R. J. (2019). Compulsivity and impulsivity traits linked to attenuated developmental frontostriatal myelination trajectories. *Nature Neuroscience*, *22*(6), 992–999. https://doi.org/10.1038/s41593-019-0394-3