# Electrical brain activations in young children during a probabilistic reward-learning task are associated with behavioral strategy.

**Yu Sun Chung[1*], Berry van den Berg[2], Kenneth C. Roberts[3], Marty G. Woldorff [1,3], Michael S. Gaffrey[1,4,5]**

[1] Department of Psychology and Neuroscience, Duke University, Reuben-Cooke Building, 417 Chapel Drive, Durham, NC 27708, USA

[2] Experimental Psychology, University of Groningen, The Netherlands

[3] Center for Cognitive Neuroscience, Department of Psychiatry, Psychology & Neuroscience and Neurobiology, Duke University, Durham, NC, 27708
USA.

[4] Children's Wisconsin, 9000 W. Wisconsin Avenue, Milwaukee, WI, 53226

[5] Medical College of Wisconsin, Division of Pediatric Psychology and Developmental Medicine, Department of Pediatrics, 8701 Watertown Plank Road, Milwaukee, WI, 53226

*** Correspondence:**
Corresponding Author
Yu Sun Chung, Ph.D. [1]
Reuben-Cooke Building
417 Chapel Drive
Campus Box 90086
Duke University
Department of Psychology and Neuroscience
Durham, NC 27708
Email: yusun08@gmail.com

**Keywords**: reinforcement learning, reward-related positivity, attention, N2pc, P300, win-stay-lose-switch strategy

## Abstract

Both adults and children learn through feedback which events and choices in the environment are associated with higher probability of reward. This probability reward-learning ability is thought to be supported by the development of fronto-striatal reward circuits. Recent developmental studies have applied computational models of reward learning to investigate such learning in children. However, there has been limited development of task tools capable of measuring the cascade of neural reward-learning processes in children. Using a child-version of a probabilistic reward-learning task while recording event-related-potential (ERP) measures of electrical brain activity, this study examined key processes of reward learning in preadolescents (n=30), namely: (1) reward-feedback sensitivity, as measured by the early reward-related frontal ERP positivity, (2) rapid attentional shifting of processing toward favored visual stimuli, as measured by the N2pc component, and (3) longer-latency attention-related responses to reward feedback as a function of behavior strategies (i.e., Win-Stay-Lose-Shift), as measured by the central-parietal P300. Consistent with our prior work in adults, the behavioral findings indicate that preadolescents could learn stimulus-reward outcome associations, but at varying levels of performance. Neurally, poor preadolescent learners (those with slower learning rates) showed greater reward-related positivity amplitudes relative to good learners, suggesting greater reward sensitivity. We also found attention shifting towards to-be-chosen stimuli, as evidenced by the N2pc, but not to more highly rewarded stimuli. Lastly, we found an effect of behavioral learning strategies (i.e., Win-Stay-Lose-Shift) on the feedback-locked P300 over the parietal cortex. These findings provide novel insights into the key neural processes underlying reinforcement learning in preadolescents.

## 1. Introduction

From an early age, children learn about the world around them through trial and error. That is, through repeated experience, children progressively learn what actions are likely to result in relatively positive or negative outcomes given a specific environment and/or set of choices. The ability to learn associations between choice behavior and outcomes and change patterns of subsequent choice behavior in light of this information, has been called probabilistic reinforcement learning (PRL). Computational models of PRL separate such learning into computationally-driven components associated with both behaviorally and neurobiologically distinguishable processes. Specifically, PRL model involves a sequence of neurocognitive processes, including evaluation of the relative value associated with available choices, updating of stimulus-reward outcome associations based on feedback, and shaping of selective attention toward a more-likely-winning stimulus to guide future choice (Anderson, 2017). However, while the neural correlates of PRL have been well studied in adults, relatively few studies have investigated this question in adolescents or children (reviewed by (DePasque & Galván, 2017). This is a critical knowledge of gap because the study of PRL in children is highly promising for identifying specific neurocognitive processes that may account for differences in these important cognitive functions across development (Nussenbaum & Hartley, 2019b) and/or may predict risk for depression, which is heightened during adolescence (Chen et al., 2015; Keren et al., 2018a).

At the behavioral level, PRL is evident during childhood, adolescence, and adulthood (Raab & Hartley, 2018). However, findings from studies using computational models of PRL to examine how value-based learning ability changes with age have been mixed. As reviewed by (Nussenbaum & Hartley, 2019a), some developmental studies report that learning rates increase from childhood to adulthood (Davidow et al., 2016; Master et al., 2020) while others report that they do not change with age (Palminteri et al., 2016). Such mixed behavior findings may reflect individual differences in learning rates that may come from neurobiological alterations in one or all of the processes underlying PRL. For example, learning rate captures individual differences in the value-updating process, reflecting the degree to which reward feedback and reward prediction errors are incorporated into the updated value estimates during learning. A number of empirical studies using functional Magnetic Resonance Imaging (fMRI) have suggested that reward prediction errors are signaled by mid-brain dopaminergic, fronto-striatal reward brain circuits (Averbeck & O'Doherty, 2022; Nasser et al., 2017; Schultz, 2016), including in adolescents (Christakou et al., 2013; Cohen et al., 2010; Hauser et al., 2015). However, the sluggishness of fMRI signals limits the ability of this method to the different subprocesses underlying learning on a trial-by-trial basis, which requires more precise temporal resolution.

The high temporal resolution of electroencephalogram (EEG) data makes it particularly well-suited for studying the neural subprocesses associated with PRL. More specifically, the ability to measure event-related potentials (ERPs) time-locked to stimulus and reward-feedback events provides an exceptional opportunity to delineate the temporal cascade of cognitive processes related to reward processing within a trial (Amodio et al., 2014). One ERP component extensively examined in reward studies is the reward positivity (RewP), a frontal, positive-polarity, brain wave measured in response to gain minus loss feedback, occurring approximately 250-350ms after that feedback. With respect to RL theory (Holroyd & Coles, 2002; Sutton & Barto, 2018), the RewP is thought to reflect a reward prediction error signal that is sensitive to outcome valence and is larger for unexpected positive events relative to unexpected negative events (Holroyd & Coles, 2002). Although the RewP (or its inverse, the Feedback-related Negativity, FN**,** calculated as loss-minus gain in some earlier research (Krigolson, 2018)) as a measure of response to reward has been investigated in children, its role in PRL during this age has not.

As one learns which stimuli are more likely to generate reward outcomes, one tends to pay more attention to the stimuli associated with higher probability of leading to reward. Such attention shifting toward reward-associated stimuli has been indexed by the lateralized attention-shifting-sensitive N2pc component in prior work with adults (San Martin et al., 2016; Hickey et al., 2010). In contrast, there have been only a handful of ERP studies that have examined the N2pc in children, either in the context of a reward task or more generally during visual search (Couperus & Quirk, 2015; Li et al., 2022; Shimi et al., 2014; Sun et al., 2018; Turoman et al., 2021) and/or working memory tasks (Rodríguez-Martínez et al., 2021; Shimi et al., 2015). Consistent with prior work with adults, existing studies with children indicate the elicitation of the N2pc component during non-PRL, attentional tasks, indicating that this component can be used as a marker of attentional section in children like in adults (Couperus & Quirk, 2015; Li et al., 2022; Shimi et al., 2014; Sun et al., 2018; Turoman et al., 2021). However, the presence of this attentional neural process has not been directly tested within the framework of computational PRL model. Accordingly, we know little about the functionality of the N2pc ERP component in children within a PRL framework – that is, whether it may reflect an index of early attentional shifting as a function of reward learning in children, as shown previously in adults.

Lastly, the P300 has been thought more generally to reflect longer-latency attentional processes involved in information updating of salient outcomes during evaluative processing (for reviews, see Glazer et al., 2018, San Martín, 2012). The P300 is a centro-parietally distributed, positive-going wave that peaks between 400 and 600 ms following a stimulus (San Martín et al., 2013; Wu & Zhou, 2009). It has been successfully used to investigate PRL with adults (e.g., (Donaldson et al., 2016). More specifically, prior work has shown that P300 is sensitive to a later, top-down-control process of outcome evaluation, suggesting that outcome-related feedback information in a probability choice task (e.g., probability and magnitude) is used to guide future decisions for the goal of reward maximization (San Martín et al., 2013, Wu and Zhou, 2009, Zhang et al., 2013). However, like the N2pc, the functionality of the feedback locked P300 (reward gain vs. loss feedback) during PRL in children has been rarely examined.

The purpose of the current study was to establish the feasibility and initial validity of a developmentally appropriate probabilistic reward learning task, which we have termed the C-PLearn task (for Child Probability Learning task). Based on our results, we believe that this task would be a promising tool for studying both typical and atypical development of key neural processes underlying PRL. For example, alterations in PRL abilities have been known to be critical components of hedonic functioning that predicts depression risk for children and adolescents (Keren et al., 2018b; Morris et al., 2015; Saulnier et al., 2023). The design of the C-PLearn task was based on our prior work investigating neural correlates of PRL processes in adults (van den Berg et al., 2019b). In the present study, we examined key processes of PRL in a group of preadolescent children, namely (1) reward-feedback sensitivity, as measured by the RewP, (2) early attentional processing toward stimuli with higher reward value or likelihood as a shaped by reward learning, as measured by the N2pc, (3) longer-latency attentional processing in response to the feedback as a function of behavioral learning strategies (i.e., Win-Stay-Lose-Shift, WSLS (Hernstein et al., 2000)), as measured by the central-parietal P300. More specifically, first we wanted to replicate the presence of the RewP (reflecting reward-feedback sensitivity and reward prediction errors), the N2pc (reflecting the shifting of attention to certain stimuli versus others), and the P300 (reflecting longer-latency attention-related cognitive processes) as a function of reward learning for preadolescents within the PRL framework. Second, we predicted that preadolescents who didn't learn the stimulus-outcome associations very well (i.e., "bad" learners) would show smaller RewP amplitudes, consistent with the RL framework, compared to "good" learners, and/or may adopt less effective behavior strategy. Finally, we

hypothesized that the central-parietal, feedback-locked P300 may reflect a modulation of later cognitive processing to update reward value as a function of behavior strategy for PRL learning (e.g., WSLS) in preadolescents, an idea that has been supported by the literature in adults (Donaldson et al., 2016; von Borries et al., 2013).

## 2. Methods

### 2.1. Participants

Thirty children between 8 to 12 years old (mean age = 9.94 years ±1.26 (SD); 15 female; 30 right-handed; 20% Hispanic or Latino ethnicity, 80% non-Hispanic white) were recruited from a Research Participation Database at Duke University maintained by the Human Subjects Coordinator for the Department of Psychology and Neuroscience. In addition, methods included posting flyers on social media sites and distributing flyers to local schools were used to recruit participants.

Children who were color blind, reported a current and/or historical diagnosis of any psychiatric disorder, had a history of neurological disorder or insult (e.g., seizures, hydrocephalus, cerebral palsy, brain tumor, extended loss of consciousness, head trauma), took psychotropic mediation for a mood or behavioral difficulty, or left-handed were excluded from study participation. At the end of EEG sessions, a $30 Amazon e-gift card was sent to parents via email for their participation in the study. Children also received candy and/or small toys after each completed session. Additional compensation was provided ($10 Amazon e-gift card per hour) to parents if testing sessions exceeded 2.5 hours. This study was conducted in accordance with protocols approved by the Duke Institutional Review Board.

### 2.2. The Child Probabilistic Reward Learning Task (C-PLearn)

The C-PLearn task used in the current study for children was based on our adult PLearn task previously used successfully in college students (van den Berg et al., 2019b). Participants completed the C-PLearn, followed by a face-vs-house neural localizer task (for localizing the face-selective neural activity on the scalp), while EEG was recorded. The task was programmed and run on the OpenSesame platform for behavioral research (Mathôt et al., 2012), and event codes were sent using the Python ‘egi’ package.

The C-PLearn consisted of 20 blocks of 18 trials each, slightly different than the 20 trials per block in the original adult version to accommodate reduced attention spans and lower tolerance for prolonged experimental tasks in children. Also, the probability of reward was set to 0.77 (instead of ranging randomly from 0.5 to 0.75 as in the original adult version) to mitigate child frustration and potential confusion noted during early piloting sessions using the original adult reward probability levels. At the beginning of each trial, participants were presented with two images, one of a face and one of a house, and were instructed to choose between the two. They were also instructed to try to learn during each block whether faces or houses were more likely to result in winning points, and that winning more points would allow them to choose small toys with differing levels of cost. Each block was followed by feedback as to how many points they had accrued up until that point. After a practice period in which the subject needed to demonstrate understanding of how to do the task, the test session started with the C-PLearn task runs.

### 2.3. Computational modeling of reward learning behavior

To calculate each subject’s reward learning behavior within the PRL framework, we defined a win-stay (WS) trial as those in which the previous trial ($n - 1$) was rewarded (won) and the choice on the current trial $n$ is the same as the choice on trial $n - 1$ (i.e., “stayed” with the same choice of face

or house made on the previous trial). We then calculated WS probabilities as the proportion of trials with WS behavior given a previously rewarded trial (i.e., after receiving gain feedback). Also, we defined a lose shift (LS) trial as those in which the previous trial ($n - 1$) was not rewarded (i.e., receiving loss feedback) and the choice on the current trial $n$ differs from the choice on trial $n - 1$. We then calculated LS probabilities as a proportion of trials with LS behavior given that the previous trial was unrewarded (i.e., after receiving loss feedback). Similarly, WS probabilities were calculated as a proportion of trials with WS behavior given that the previous trial was rewarded (i.e., after receiving gain feedback).

2.4. EEG Data Collection and Processing

Continuous EEG was recorded using EGI's Net Station acquisition software version 5.4.2 and Net Amps 400 series amplifiers with the associated 128-channel Geodesic Sensor Net system, at a sampling rate of a 1000 Hertz (Hz per channel). Data was referenced online to the vertex (Cz) during acquisition, and impedances were maintained below 50 kilohms throughout the paradigm.

Offline EEG data preprocessing was performed using EEGLAB version 2021 (Delorme & Makeig, 2004) in MATLAB R2019b (The MathWorks Inc, Natick, MA) using custom scripts (codes are available at:https://github.com/DEEDLabEEG). According to the Maryland analysis of developmental EEG pipeline recommended for pediatric populations (Debnath et al., 2020), we followed several preprocessing steps as follows. First, twenty-four channels located on the outer ring of the sensor net were removed from the analyses. As participants were required to wear masks during the EEG sessions according to Duke COVID-related policy, these channels contained many artifacts during the recordings in these children due to poor connections. The data were then downsampled to 250Hz and segments of data without relevant events (i.e., breaks) were removed. Using the ERPLAB v8.2 plugin (Lopez-Calderon & Luck, 2014), a 0.1 to 30 Hz, 4$^{th}$ order Butterworth, bandpass filter was applied. CleanLine (Kappenman & Luck, 2012) was used to remove any remaining 60 Hz line noise (Mitra & Bokil, 2007). All data were re-referenced to the average of the mastoids. Bad channels were removed by running the Clean Rawdata plugin; a channel was considered bad if 1) it was flat for more than five seconds, 2) contained more than four standard deviations of line noise relative to its signal, or 3) correlated at less than .8 to nearby channels.

According to several recommendations (Debener et al., 2010; Debnath et al., 2020), a "copy" of the data was made and then, the following steps were applied (Delorme & Makeig, 2004): 1) the high-pass filter at a 1 Hz (4$^{th}$ order Butterworth) , 2) Artifact Subspace Reconstruction (ASR;(Mullen et al., 2015) with the burst criterion set to 20 (Chang et al., 2018) to remove large artifacts with Clean Rawdata, and 3) extended infomax Independent Component Analysis (ICA; (Lee et al., 1999) with PCA dimension reduction (50 components). The resulting ICA matrix was then copied over to the original full-length data (i.e., the data just before the copy was made and the 1 Hz high-pass filter was applied). The ICLabel plugin (Pion-Tonachini et al., 2019) automatically removed independent components with a probability greater than .7 of being associated with eye movements and blinks. Application of the ICA matrix to the full-length data and subsequent removal of eye-related components allowed for the preservation of data that would have otherwise been removed by ASR or other artifact removal methods. Bad channels that were previously removed from the analyses were interpolated back into the data set using spherical splines (Perrin, 1989).

For both the C-PLearn task and the localizer task, epochs were extracted from 400 ms before until 800 ms after the onset of the stimulus (either the cue-pairs or feedback screens in the main task or face/house images in the localizer task). All epochs were baseline corrected from -200 ms to the onset of the stimuli. Artifact rejection using the TBT plugin (Trial-By-Trial basis; (Ben-Shachar, 2018)) removed epochs with at least 10 channels meeting the following criteria: 1) peak-to-peak amplitudes exceeding 100 μV within 200 ms windows sliding across the epoch by 20 ms increments, 2) voltages below -150 or greater than +150 μV, or 3) joint probabilities above three standard

deviations for local and global thresholds. If less than 10 channels in an epoch met criteria for rejection, the epoch was not removed, but the identified channels were interpolated for that epoch only. But if there were more than 10 channels in the epoch that met criteria for rejection, the epoch was not included. Lastly, the averaged ERPs for each bin (i.e., Gain, Loss) were computed. Mean of the number of accepted Gain and Loss trials per subject were 184.3 and 128.2, respectively.

2.5. Data binning and averaging

The epoched EEG data were binned according to feedback (gains and losses) and choice (face and house), which resulted in an average number of trials for each subject (Brown et al.) for face gain [90.53 (± 24.97)], face loss [62.13(±13.73)], house gain [93.76(±19.60)], and house loss [66.06 (±14.76)], after rejection of noisy epochs. As expected, due to learning, there were significantly more gain trials than loss trials [$F(1,29) = 43.52$, $p < .001$]. On the other hand, there were no significant differences in the number of gain or loss trials for choosing a face vs a house ([$F(1,29) = 0.46$, $p = .50$].

2.6. Analysis of cue-evoked ERP data

We calculated the cue-evoked N2pc response to the cue-pair at the beginning of each trial, to assess attentional orientation to the two stimulus types as a function of what they will later choose that trial, following a similar procedure used in our prior work in adults (van den Berg et al., 2019b). More specifically, the N2pc neural responses were assayed by the standard N2pc contralateral-vs-ipsilateral analysis this component (Luck & Kappenman, 2012), that is by subtracting the activity in the contralateral channels (relative to the chosen side) minus the ipsilateral channels and then collapsing over the left and right sides.

2.7. Statistical Analysis

For statistical analysis, we first calculated the time-locked ERPs as a function of the various event types and conditions. Based on previous literature, the reward-related Positivity (RewP) was measured from 275 to 375 ms in a fronto-central ROI (E10, E11, E16). The P300 was measured from 400 to 600 ms in a central-parietal ROI (Left/Right: E37/E87, E31/E80, E53/E86, E60/E85, E67/E77, E61/E78, E54/E79; middle line: E129, E55, E62, E71; (Fischer & Ullsperger, 2013; San Martín, 2012; San Martín et al., 2013). Subsequently, the extracted values were analyzed with a repeated-measures analyses of variance (ANOVAs) to test for statistical significance ($p < .05$) using the SPSS 2021 version.

The EEG data collected during the independent localizer task were used to confirm that the stimuli that we used in this study elicited face-vs-house activity differences as has been observed in the existing literature and to delineate potential influences of reward learning on stimulus-specific regions. To do so, we first visually inspected all posterior electrodes (i.e., over visual cortex) for the ERP differences between the neural responses to face and house images to make sure these effects were representative of previously reported findings with adults, including the hallmark, face-selective, lateral inferior occipital N170 responses. The N170 was measured from 170 to 220 ms after onset of the cue image, from corresponding left and right occipital ROIs (E66/E84, E67/E91, E61/E92).

The cue-related attentional bias (N2pc) was measured from 175 to 225 ms after onset of the cue image pair, from corresponding left and right occipital ROIs (E50/E101, E58/96; (Hickey et al., 2010; Kappenman & Luck, 2012; San Martín et al., 2016)). Subsequently, we also calculated the difference in voltage contralateral vs ipsilateral relative to the side on which the set-winner was presented.

To calculate each subject's individual behavioral learning rate, we used the same formula (see (van den Berg et al., 2019b) for detailed formula) and calculated on a trial-by-trial number basis

using a mixed-modeling approach using the *lme4* statistical package (Bates et al., 2015). A varying slope of condition per subject (the random effect) was included in the model if the Akaike Information Criterion (a measure of the quality of the model) improved. Statistical significance was set at $p < 0.05$, with the Satterthwaite's degrees of freedom method as given by the R package *l merTest* (Kuznetsova et al., 2017).

## 3. Results

### 3.1. Behavioral Results

Consistent with our prior work in adults (van den Berg et al., 2019b), and as predicted, children demonstrated behaviorally reinforcement learning ability based on trial-by-trial reward-related feedback. The probability learning behavior across blocks in children for choosing the probabilistic set-winner ranged from 0.40 to 0.73 successful, indicating that, as expected, the children were able to integrate reward information throughout the block through reinforcement learning. As presented in Figure 2.a, at the beginning of each 18-trial set, the participants chose the most likely winner for that set at chance level (0.45-0.50), as would be expected. But by the end of the set, this proportion of choosing set-winner increased up to ~0.75 on average, indicating that the children had used the feedback across the trials of that set to learn to choose the more likely probability-based winner (Fig2a). That said, the individual learning curves revealed very large individual difference in reward learning, as the final proportion of choosing set winners in a particular block ranged from 0.26 to 1.0 across the subjects and blocks. Due to this large range, we wanted to explore whether 'good' vs. 'bad' learners might show different patterns of reward-related ERP components, in particular by dividing the subjects using a median split based on their learning performance (Fig1).

As presented in Figure 2, good vs. bad learners adopted different behavior strategies during the C-PLearn task. Good learners were more likely to choose the same stimulus type (i.e., Win-Stay, WS behavior strategy) on the trial after one with a gain feedback ($t(28) = 3.92$, $p < .001$, Cohen's $d = 1.43$). In sharp contrast, bad learners were more likely to change their response choice (i.e., Lose-Shift, LS behavior strategy) on the next trial after loss feedbacks ($t(28) = -2.89$, $p = 0.007$, Cohen's $d = 1.05$).

##### figure 1 here #####.

##### figure 2 here #####.

### 3.2. Electrophysiological Measures

#### 3.2.1. The Reward-related Positivity (RewP)

Consistent with prior work, we confirmed the presence of the RewP around 275-375 ms after gain vs. loss feedback in our frontal ROI (E10, E11, E16) for young children (see Figure 3.1). According to the independent *t*-tests, bad learners had greater RewP amplitudes in the frontal ROIs compared to good learners [$t(27) = -2.14$, $p = .04$, Cohen's $d = 0.79$] (see Figure 3.2 )

##### figures 3.1 and 3.2 here #####.

#### 3.2.2. P300 during reward gain vs. loss feedback as a function of behavior strategy

We examined updating reward values as a function of reward learning, as reflected by behaviorally dependent modulation of the P300. Figure 5.1 shows the P300 during gain vs. loss feedback as a function of behavior strategy (i.e., Switch vs. Stay) between good and bad learners. Using P300 amplitudes in the central ROIs between 400 and 600 ms for good and bad learners, we conducted repeated measures ANOVAs with two within-subject factors (i.e., behavior strategy: switch vs stay, feedback: gain vs. loss) and one between-subject factor (i.e., good vs. bad learners). The ANOVAs were conducted to examine whether bad learners showed greater P300 amplitudes

than did good learners during gain vs. loss feedback as a function of switch behavior strategy compared to good learners. As shown in Figure 4.2(a), bad learners (i.e., those with low learning rates), relative to good learners, showed higher P300 amplitude differences in central, parietal ROIs when they switched their responses on the next trial after gain feedback vs. after loss feedback. The repeated ANOVA analysis revealed no main effect on the P300 of learner type [$F(1, 27) = 1.45$, $p = .23$, $\eta_p^2$: 0.05]. We did observe, however, main effects of behavior strategy [$F(1, 27) = 9.11$, $p = .005$, $\eta_p^2$: 0.25] and feedback [$F(1, 27) = 7.00$, $p = .01$, $\eta_p^2$: 0.21], as presented in Figure 4.2.(b) and (c), respectively, but no interaction of feedback by behavior strategy [$F(1, 27) = 0.34$, $p = .56$, $\eta_p^2$: 0.01] or of behavior strategy by learner type [$F(1, 27) = 0.38$, $p = .54$, $\eta_p^2$: 0.01] or feedback by behavior strategy by learner type [$F(1, 27) = 0.09$, $p = .76$, $\eta_p^2$: 0.003] (see Figure 4.2). According to the *post-hoc* paired *t*-tests, after receiving gain vs. loss feedback, children showed greater P300 amplitude differences in the central, parietal ROIs when they switched their choice on the subsequent trial compared to when they chose the same responses [$t(28)= 2.68$, $p = .012$, Cohen's $d = 0.32$ ] (see Figure 4.2.b). Also, P300 amplitudes in the same central and parietal ROIs were higher when children received loss feedback compared to gain feedback, collapsed across both good and bad learners [$t(28)= 3.03$, $p = .005$, Cohen's $d = 0.39$] (see Figure 4.2.c)

##### figure 4.1  and 4.2 here #####.

3.2.4. N2pc reflects attentional bias towards choice behavior

In addition, in order to examine attention shift towards behavior choice regardless of whether or not reward learning had been consolidated, we analyzed the effect of choice on the N2pc over the course of the 18-trial set. We collapsed over whether the chosen image was that of a face or a house to increase signal-to-noise ratio. We analyzed the relationship between the N2pc elicited by the cue-pair stimulus relative to the choice made later in the trial, regardless of what the set-winner was. As expected, we found the presence of N2pc around 175-225 ms after the onset of each cue-pair stimulus in the ROIs (E50/E58) [$t(29)= , p <.008$]. These results indicate that children's reward learning behavior was reflected by the modulation of the N2pc, indexing early attentional bias towards their choice behavior.  Like our prior work with college students (van den Berg et al., 2019), we did not analyze choice behavior in the analysis of the neural N2pc. Because if one calculates the N2pc amplitudes and its changes with learning across a block separately for when faces vs. houses were the set-winner, the lateralized attention-related N2pc activity could be potentially conflated with lateralized face-selective N170 activity. Independent *t*-test on the N2pc amplitudes measured in the ROIs from 175 to 275 ms between good and bad learners revealed no statistical differences ($t(28) = -1.76$, $p=.09$), even though the behavior indicated that they had learned the reward association.

##### figure 5 here #####.

## 4. Discussion

Consistent with our prior work in adults (van den Berg et al., 2019b), our behavioral C-PL earn findings indicate that at a group level, preadolescents are able to learn which stimulus type in a pair (a face vs. a house) is more likely to result in a higher probability of reward over a series of trials. As we observed large individual differences in individual learning rates, however, we divided all participants into good (i.e., with high learning rates) vs. bad learners (i.e., with low learning rates). Further, our behavior strategy analysis revealed that bad learners used a different strategy than did good learners, namely a "Lose-Shift" behavior strategy rather than a "Gain-Stay" strategy. In line with prior work with adults (Worthy & Maddox, 2014), choosing the same choice on the next trial after receiving reward feedback, namely, "Gain-Stay" strategy appears to be the better heuristic learning strategy for preadolescent children in binary decision-making situations.

At the neural level, the current study's findings support C-PLearn as a developmentally appropriate task for identifying and measuring key ERPs associated with PRL processes and behavior in preadolescents. In contrast to our predictions, we found that poor preadolescent learners with low learning rates showed greater RewP ERP amplitudes than good learners, indicating greater reward-feedback sensitivity and/or greater prediction errors. However, our behavior results revealed that poor learners used a different strategy than did good learners, namely a "Lose-Shift" behavior strategy, in line with prior work with adults (Worthy & Maddox, 2014), rather than a "Gain-Stay" strategy. Furthermore, we found early attention modulation toward the stimulus to be chosen, as evidenced by the N2pc, although not toward the stimulus of higher likelihood of reward, which has been shown in adult subjects (van den Berg et al. 2019a). Lastly, we demonstrated the contextual modulation of feedback-locked P300 over the parietal cortex as a function of behavior strategy in preadolescents. These results provide novel insights into the associations between RL-related ERP components and computationally-driven reward learning behavior (i.e., Win-Stay-Lose-Shift, WSLS: (Hernstein et al., 2000). As a result, the current study supports the use of C-PLearn to measure key behavioral and neural correlates of PRL processing in preadolescent children.

To our knowledge, this is the first study, not just to examine the key neural processes underlying PRL learning (i.e., reward-related feedback processing and attention shaping as a function of reward learning) in preadolescents, but also with a combination of PRL and WSLS models. Most existing ERP studies with children have focused on the RewP effect on general reward feedback processing (e.g., (Burani et al., 2019; Ethridge et al., 2017; Ferdinand et al., 2016; Kessel et al., 2016; Kessel et al., 2019; Lukie et al., 2014). However, little is known about the RewP in the context of reinforcement learning in preadolescents. This is critical gap of knowledge in the field because poor PRL learning may be a key process for underlying adolescent-onset of depression (Chen et al., 2015; O'Callaghan & Stringaris, 2019). Although the exact underlying mechanisms are not yet clear, a meta-analytic review provides strong evidence that depression is associated with impaired brain signals of reward prediction error and with poor reinforcement learning behavior (Halahakoon et al., 2020). We extended the current literature by examining the RewP effect in the domain of RL processing. Different from our predictions, we found the presence of a small frontal RewP ERP following reward feedback for preadolescents, which was mostly driven by *poor* learners, as described more below. Given that we observed large individual differences in computationally-driven behavior learning rates in the present study, we compared the RewP effect between "good" and "bad" learners (i.e., those with lower learning rates), as defined by their computationally-derived learning rates. This adaptation of WSLS model in current study revealed that poor preadolescent learners tended to "switch" their choice following losing feedback. These findings are consistent with prior work with college students(Worthy & Maddox, 2014). WSLS is a rule-based strategy that has been

shown to be commonly used in binary outcome choice tasks (e.g., Otto et al., 2011). According to (Worthy & Maddox, 2014), college students with higher probabilities of shifting following a loss trial (i.e., "Lose-Shift" behavior strategy) showed poorer learning performance. Along with prior work with adults, current findings add converging evidence that combining the WSLS and PRL models may provide a better account of binary decision-making behavior even in preadolescents compared to a single PRL model.

One key component of PRL processing is attentional shifts towards stimuli associated with higher probability of getting rewards. Previously learned reward values can have a pronounced effect on the allocation of selective attention, which has been indexed by an increasing amplitude of the attention-sensitive, lateralized negative deflection (the N2pc) contralateral to the set-winner in the cue-pair presentation. The prior ERP studies in adults showed that once a stimulus-reward association has been learned, it tends to induce a larger attentional shift when presented and also predicts the likelihood of subsequent choice behavior ((Hickey et al., 2010; Hickey & van Zoest, 2012; San Martín et al., 2016). Such pronounced attentional shift in the context of reward learning, as indexed by the N2pc has been little investigated in children. Only a handful of ERP studies with children reported the role of N2pc during visual search (Couperus & Quirk, 2015; Li et al., 2022; Shimi et al., 2014; Sun et al., 2018; Turoman et al., 2021) and/or in working memory tasks(Rodríguez-Martínez et al., 2021; Shimi et al., 2015). In addition to the RewP effect, the current study revealed the modulation of N2pc ERP amplitude indexing early attentional processing towards the choice behavior (which stimulus would be chosen on that trial). Our prior work in adults (van den Berg et al., 2019a) revealed the presence of significant N2pc to the to-be-chosen stimulus *and* to the more rewarded stimulus in a set, particularly on the later trials of the set, indicating early attention modulation as a function of reward learning. However, we were not able to replicate such reward-learning findings in the attention-shifting in the present study with preadolescents. Rather, we found a modulation of attention bias only towards the chosen side, which was elicited over the whole course of learning. We speculate that bad learners' limited attention towards rewarded stimuli might have diluted the presence of small N2pc amplitudes during later vs. early trials of learning in the good learners. This speculation should be tested in future studies with large samples of both good and bad child learners.

Although we found that, like college students, preadolescents learned the stimulus-reward outcome association at a group level, our data indicated large individual differences in behavioral learning rates. With the computationally-driven PRL framework, our behavior data indicated major differences in strategy between good and bad learners, with good learners using win-stay strategy and bad learners using more of a lose-switch strategy. PRL theory generally assumes that the most recent outcomes exert the most influence on the current choice, such that behavior on a given trial, $n$, depends particularly upon the choice and outcome on the preceding trial, $n - 1$. This relationship has been formalized as a simple, heuristic learning strategy called Win-Stay-Lose-Shift (WSLS: (Hernstein et al., 2000). In our prior work in college students (San Martín et al., 2013), the amplitude of the feedback-locked P300 over central and parietal cortex, but not the FN (RewP), predicted behavioral adjustments (i.e., lose-switching strategy) on the next trial. These results are in line with the context-updating P300 hypothesis (Nieuwenhuis et al., 2005): decisions on each trial are informed by an internal model of the symbol/probability(win) contingencies, and the P300 amplitude reflects the extent of the feedback-triggered revision of such a model. Thus, the feedback-locked P300 distributed over central and parietal cortex may reflect adjustment processes to maximize gain and/or minimize loss. In line with this, the present study also found a modulation of P300 over central and parietal cortex as a function of behavioral strategy and feedback type. The feedback-

locked P300 amplitudes were greater on lose-switching trials compared to win-staying trials, and during loss vs. gain reward feedback.

In contrast to our expectation, we did not find differences in the feedback-locked P300 amplitudes located over central and parietal cortex between good and bad learners even though our data indicated a modulation of the P300 amplitudes as a function of behavioral strategy and feedback types more generally. We cannot rule out a possibility that there was not enough power to detect the interaction effect of P300 amplitudes and subgroup types due to limited number of participants. But we speculate that better reward learning in children may be more driven by reward sensitivity, rather than later attention processing related to the updating of salient reward outcomes, as evidenced by significant differences in the RewP amplitude patterns between the two subtypes of learners.

The main limitation of the current study is a relatively small sample size for examining individual level differences, which may have resulted in our missing more nuanced distinctions in our behavioral measures of RL behavior and brain function (e.g., no significant N2pc to the stimulus with a higher likelihood of reward). In addition, our sample was limited to preadolescent children. As a result, future work will be needed to replicate the current findings and test their generalizability in other age ranges and in larger groups reflective of the general US population in terms of race and ethnicity.

In summary, results of the current study provide novel insights into how a cascade of core neurocognitive processes underlying RL are associated with preadolescents' behavior learning strategy: 1) poor preadolescent learners (i.e., with low learning rates) showed greater reward-feedback sensitivity, which might reflect greater reward-prediction error compared to good learners, as evidenced by the RewP, 2) there was an early attention modulation towards the stimulus that would be chosen on a trial, as evidenced by the N2pc, and 3) updating reward values as a function of behavior strategy, as evidenced by the modulation of P300 amplitudes depending on lose-switch vs. win-stay strategy in the different learning groups. The alterations in these cognitive processes have been known to be a key mechanism underlying adolescent-onset depression (Chen et al., 2015). Adolescents have showed heightened vulnerability for depression with the onset of puberty when there is a significant reorganization of frontal-striatal reward brain circuitry, as implicated by differences in RL behavior (Kessler et al., 2001; Walker et al., 2017). The current study shows that the CPLearn task with the use of EEG measures of brain activity may be a promising tool to further investigate the relationship between attention and reward sensitivity in preadolescents, which is turn may advance our understanding of the factors underpinning the risk for depression and other cognition-development issues in this key demographic group.

## Conflict of Interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest

## Author Contributions

Chung Y.S. led all aspects of the study, including the development of the task paradigm, the collection of the data, the analyses of the data, the interpretation of the results, and the writing of the manuscript. Van den Berg B. and Roberts K.C. contributed to the development of the C-PLearn task, the preprocessing pipelines, and the data analyses. Woldorff M.G. contributed to the development of the task paradigm, provided guidance on the EEG data analyses, and contributed to the interpretation of the results and feedback on the manuscript. Gaffrey MS. contributed to the development of the task paradigm, guidance on the analyses, interpretation of the results, and feedback on the manuscript, and provided the research funding to support the study.

**Funding**

This study was supported by the National Institute of Mental Health, USA (R01MH110488 to MSG).

**Acknowledgments**

We would like to thank to children and their parents who participated and completed this study despite challenges due to the COVID pandemic (i.e., wearing face masks throughout all sessions). Also, we would like to specially thank to research staff, Amanda Neal and Summer Lawrence for their effort and support of data collection. Without their help, we could not have successfully completed this project.

## References


Amodio, D. M., Bartholow, B. D., & Ito, T. A. (2014). Tracking the dynamics of the social brain: ERP approaches for social cognitive and affective neuroscience. *Soc Cogn Affect Neurosci*, *9*(3), 385-393. https://doi.org/10.1093/scan/nst177

Anderson, B. A. (2017). Going for It: The Economics of Automaticity in Perception and Action. *Curr Dir Psychol Sci*, *26*, 140-145.

Averbeck, B. A.-O., & O'Doherty, J. P. (2022). Reinforcement-learning in fronto-striatal circuits. (1740-634X (Electronic)).

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). Package lme4. *Journal of Statistical Software*, *67*, 1-91.

Ben-Shachar, M. (2018). TBT: Reject and Interpolate channels on a trial-by-trial basis. Zenodo. *https://doi.org/10.5281/zenodo.1241518*.

Brown, V. M., Zhu, L., Solway, A., Wang, J. M., McCurry, K. L., King-Casas, B., & Chiu, P. H. (2021). Reinforcement Learning Disruptions in Individuals With Depression and Sensitivity to Symptom Change Following Cognitive Behavioral Therapy. *JAMA Psychiatry*, *78*(10), 1113-1122. https://doi.org/10.1001/jamapsychiatry.2021.1844

Burani, K., Mulligan, E. M., Klawohn, J., Luking, K. R., Nelson, B. D., & Hajcak, G. (2019). Longitudinal increases in reward-related neural activity in early adolescence: Evidence from event-related potentials (ERPs). *Dev Cogn Neurosci*, *36*, 100620. https://doi.org/10.1016/j.dcn.2019.100620

Chang, C. Y., Hsu, S. H., Pion-Tonachini, L., & Jung, T. P. (2018). Evaluation of Artifact Subspace Reconstruction for Automatic EEG Artifact Removal. *Annu Int Conf IEEE Eng Med Biol Soc*, *2018*, 1242-1245. https://doi.org/10.1109/embc.2018.8512547

Chen, C., Takahashi, T., Nakagawa, S., Inoue, T., & Kusumi, I. (2015). Reinforcement learning in depression: A review of computational research. *Neurosci Biobehav Rev*, *55*, 247-267. https://doi.org/10.1016/j.neubiorev.2015.05.005

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. *J Cogn Neurosci*, *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. *Nat Neurosci*, *13*(6), 669-671. https://doi.org/10.1038/nn.2558

Couperus, J. W., & Quirk, C. (2015). Visual search and the N2pc in children. *Atten Percept Psychophys*, *77*(3), 768-776. https://doi.org/10.3758/s13414-015-0833-5

Davidow, J. Y., Foerde, K., Galván, A., & Shohamy, D. (2016). An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. (1097-4199 (Electronic)).

Debener, S., Thorne, J. D., Schneider, T. R., & Viola, F. C. (2010). Using ICA for the Analysis of Multi-Channel EEG Data.

Debnath, R. A.-O., Buzzell, G. A., Morales, S. A.-O. X., Bowers, M. E., Leach, S. C., & Fox, N. A. (2020). The Maryland analysis of developmental EEG (MADE) pipeline. (1540-5958 (Electronic)).

Delorme, A., & Makeig, S. (2004). EEGLAB: an open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J Neurosci Methods*, *134*(1), 9-21. https://doi.org/10.1016/j.jneumeth.2003.10.009

DePasque, S., & Galván, A. (2017). Frontostriatal development and probabilistic reinforcement learning during adolescence. (1095-9564 (Electronic)).

Donaldson, K. R., Ait Oumeziane, B., Hélie, S., & Foti, D. (2016). The temporal dynamics of reversal learning: P3 amplitude predicts valence-specific behavioral adjustment. *Physiol Behav*, *161*, 24-32. https://doi.org/10.1016/j.physbeh.2016.03.034

Ethridge, P., Kujawa, A., Dirks, M. A., Arfer, K. B., Kessel, E. M., Klein, D. N., & Weinberg, A. (2017). Neural responses to social and monetary reward in early adolescence and emerging adulthood. *Psychophysiology*, *54*(12), 1786-1799. https://doi.org/10.1111/psyp.12957

Ferdinand, N. K., Becker, A. M. W., Kray, J., & Gehring, W. J. (2016). Feedback processing in children and adolescents: Is there a sensitivity for processing rewarding feedback? *Neuropsychologia*, *82*, 31-38. https://doi.org/10.1016/j.neuropsychologia.2016.01.007

Fischer, A. G., & Ullsperger, M. (2013). Real and fictive outcomes are processed differently but converge on a common adaptive mechanism. *Neuron*, *79*(6), 1243-1255. https://doi.org/10.1016/j.neuron.2013.07.006

Halahakoon, D. C., Kieslich, K., O'Driscoll, C., Nair, A., Lewis, G., & Roiser, J. P. (2020). Reward-Processing Behavior in Depressed Participants Relative to Healthy Volunteers: A Systematic Review and Meta-analysis. *JAMA Psychiatry*, *77*(12), 1286-1295. https://doi.org/10.1001/jamapsychiatry.2020.2139

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/10.1016/j.neuroimage.2014.09.018

Hernstein, R., Rachlin, H., & DI, L. (2000). The matching law. *Cambridge, MA: Harvard University Press*.

Hickey, C., Chelazzi, L., & Theeuwes, J. (2010). Reward changes salience in human vision via the anterior cingulate. *J Neurosci*, *30*(33), 11096-11103. https://doi.org/10.1523/jneurosci.1026-10.2010

Hickey, C., & van Zoest, W. (2012). Reward creates oculomotor salience. *Curr Biol*, *22*(7), R219-220. https://doi.org/10.1016/j.cub.2012.02.007

Holroyd, C. B., & Coles, M. G. H. (2002). The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev*, *109*(4), 679-709. https://doi.org/10.1037/0033-295x.109.4.679

Kappenman, E. S., & Luck, S. J. (2012). The Oxford Handbook of Event-Related Potential Components. *New York, NY: Oxford University Press*.

Keren, H., O'Callaghan, G., Vidal-Ribas, P., Buzzell, G. A., Brotman, M. A., Leibenluft, E., Pan, P. M., Meffert, L., Kaiser, A., Wolke, S., Pine, D. S., & Stringaris, A. (2018a). Reward Processing in Depression: A Conceptual and Meta-Analytic Review Across fMRI and EEG Studies. *Am J Psychiatry*, *175*(11), 1111-1120. https://doi.org/10.1176/appi.ajp.2018.17101124

Keren, H., O'Callaghan, G., Vidal-Ribas, P., Buzzell, G. A., Brotman, M. A., Leibenluft, E., Pan, P. M., Meffert, L., Kaiser, A., Wolke, S., Pine, D. S., & Stringaris, A. (2018b). Reward Processing in Depression: A Conceptual and Meta-Analytic Review Across fMRI and EEG Studies. (1535-7228 (Electronic)).

Kessel, E. M., Dougherty, L. R., Kujawa, A., Hajcak, G., Carlson, G. A., & Klein, D. N. (2016). Longitudinal Associations Between Preschool Disruptive Mood Dysregulation Disorder Symptoms and Neural Reactivity to Monetary Reward During Preadolescence. *J Child Adolesc Psychopharmacol*, *26*(2), 131-137. https://doi.org/10.1089/cap.2015.0071

Kessel, E. M., Nelson, B. D., Finsaas, M., Kujawa, A., Meyer, A., Bromet, E., Carlson, G. A., Hajcak, G., Kotov, R., & Klein, D. N. (2019). Parenting style moderates the effects of exposure to natural disaster-related stress on the neural development of reactivity to threat and reward in children. *Dev Psychopathol*, *31*(4), 1589-1598. https://doi.org/10.1017/s0954579418001347

Kessler, R. C., Avenevoli, S., & Ries Merikangas, K. (2001). Mood disorders in children and adolescents: an epidemiologic perspective. *Biol Psychiatry*, *49*(12), 1002-1014. https://doi.org/10.1016/s0006-3223(01)01129-5

Krigolson, O. E. (2018). Event-related brain potentials and the study of reward processing: Methodological considerations. (1872-7697 (Electronic)).

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). ImerTest Package: tests in linear mixed effects models. *Journal of Statistical Software*, *82*.

Lee, T. W., Girolami, M., & Sejnowski, T. J. (1999). Independent component analysis using an extended infomax algorithm for mixed subgaussian and supergaussian sources. *Neural Comput*, *11*(2), 417-441. https://doi.org/10.1162/089976699300016719

Li, B., Guo, J., Zhao, C., Luo, X., Kong, Y., Chen, Y., Liu, H., Sun, L., & Song, Y. (2022). Lack of an association between anticipatory alpha oscillations and attentional selection in children with attention-deficit/hyperactivity disorder. *Clin Neurophysiol*, *138*, 25-37. https://doi.org/10.1016/j.clinph.2022.02.026

Lopez-Calderon, J., & Luck, S. J. (2014). ERPLAB: an open-source toolbox for the analysis of event-related potentials. *Front Hum Neurosci*, *8*, 213. https://doi.org/10.3389/fnhum.2014.00213

Luck, S., & Kappenman, E. (2012). ERP components and selective attention In:

The Oxford Handbook of Event-Related Potential Components. *New York: Oxford University Press.*

Lukie, C. N., Montazer-Hojat, S., & Holroyd, C. B. (2014). Developmental changes in the reward positivity: an electrophysiological trajectory of reward processing. *Dev Cogn Neurosci*, *9*, 191-199. https://doi.org/10.1016/j.dcn.2014.04.003

Master, S. L., Eckstein, M. K., Gotlieb, N., Dahl, R., Wilbrecht, L., & Collins, A. G. E. (2020). Distentangling the systems contributing to changes in learning during adolescence. (1878-9307 (Electronic)).

Mathôt, S., Schreij, D., & Theeuwes, J. (2012). OpenSesame: An open-source, graphical experiment builder for the social sciences. *Behav Res*, *44*(314-324). https://doi.org/https://doi.org/10.3758/s13428-011-0168-7

Mitra, P., & Bokil, H. (2007). Observed Brain Dynamics. *Oxford University Press*, Chapter 7.3.4.

Morris, B. H., Bylsma Lm Fau - Yaroslavsky, I., Yaroslavsky I Fau - Kovacs, M., Kovacs M Fau - Rottenberg, J., & Rottenberg, J. (2015). Reward learning in pediatric depression and anxiety: preliminary findings in a high-risk sample. (1520-6394 (Electronic)).

Mullen, T. R., Kothe, C. A., Chi, Y. M., Ojeda, A., Kerth, T., Makeig, S., Jung, T. P., & Cauwenberghs, G. (2015). Real-Time Neuroimaging and Cognitive Monitoring Using Wearable Dry EEG. *IEEE Trans Biomed Eng*, *62*(11), 2553-2567. https://doi.org/10.1109/tbme.2015.2481482

Nasser, H. M., Calu, D. J., Schoenbaum, G., & Sharpe, M. J. (2017). The Dopamine Prediction Error: Contributions to Associative Models of Reward Learning. (1664-1078 (Print)).

Nieuwenhuis, S., Aston-Jones, G., & Cohen, J. D. (2005). Decision making, the P3, and the locus coeruleus-norepinephrine system. *Psychol Bull*, *131*(4), 510-532. https://doi.org/10.1037/0033-2909.131.4.510

Nussenbaum, K., & Hartley, C. A. (2019a). Reinforcement learning across development: What insights can we draw from a decade of research? (1878-9307 (Electronic)).

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

O'Callaghan, G., & Stringaris, A. (2019). Reward Processing in Adolescent Depression Across Neuroimaging Modalities. *Z Kinder Jugendpsychiatr Psychother*, *47*(6), 535-541. https://doi.org/10.1024/1422-4917/a000663

Palminteri, S., Kilford, E. A.-O., Coricelli, G., & Blakemore, S. J. (2016). The Computational Development of Reinforcement Learning during Adolescence. (1553-7358 (Electronic)).

Perrin, F., Pernier, J., Bertrand, O., & Echallier, J. F. . (1989). Spherical splines for scalp potential and current density mapping. *Electroencephalography and clinical neurophysiology*, *72(2)*, 184-187.

Pion-Tonachini, L., Kreutz-Delgado, K., & Makeig, S. (2019). ICLabel: An automated electroencephalographic independent component classifier, dataset, and website. *Neuroimage*, *198*, 181-197. https://doi.org/10.1016/j.neuroimage.2019.05.026

Raab, H. A., & Hartley, C. A. (2018). Chapter 13 - The Development of Goal-Directed Decision-Making. In R. Morris, A. Bornstein, & A. Shenhav (Eds.), *Goal-Directed Decision Making* (pp. 279-308). Academic Press. https://doi.org/https://doi.org/10.1016/B978-0-12-812098-9.00013-9

Rodríguez-Martínez, E. I., Arjona Valladares, A., Gómez-González, J., Diaz-Sanchez, J. A., & Gómez, C. M. (2021). Neurophysiological differences between ADHD and control children and adolescents during the recognition phase of a working memory task. *Neurosci Res*, *164*, 46-54. https://doi.org/10.1016/j.neures.2020.03.011

San Martín, R. (2012). Event-related potential studies of outcome processing and feedback-guided learning. *Front Hum Neurosci*, *6*, 304. https://doi.org/10.3389/fnhum.2012.00304

San Martín, R., Appelbaum, L. G., Huettel, S. A., & Woldorff, M. G. (2016). Cortical Brain Activity Reflecting Attentional Biasing Toward Reward-Predicting Cues Covaries with Economic Decision-Making Performance. *Cereb Cortex*, *26*(1), 1-11. https://doi.org/10.1093/cercor/bhu160

San Martín, R., Appelbaum, L. G., Pearson, J. M., Huettel, S. A., & Woldorff, M. G. (2013). Rapid brain responses independently predict gain maximization and loss minimization during economic decision making. *J Neurosci*, *33*(16), 7011-7019. https://doi.org/10.1523/jneurosci.4242-12.2013

Saulnier, K. G., Marr, N. S., van Geen, C., Babinski, D. E., & Mukherjee, D. (2023). Reinforcement-based responsiveness, depression, and anhedonia: A multi-method investigation of intergenerational risk. (1879-1379 (Electronic)).

Schultz, W. (2016). Dopamine reward prediction error coding. (1958-5969 (Electronic)).

Shimi, A., Kuo, B. C., Astle, D. E., Nobre, A. C., & Scerif, G. (2014). Age group and individual differences in attentional orienting dissociate neural mechanisms of encoding and maintenance in visual STM. *J Cogn Neurosci*, *26*(4), 864-877. https://doi.org/10.1162/jocn_a_00526

Shimi, A., Nobre, A. C., & Scerif, G. (2015). ERP markers of target selection discriminate children with high vs. low working memory capacity. *Front Syst Neurosci*, *9*, 153. https://doi.org/10.3389/fnsys.2015.00153

Sun, M., Wang, E., Huang, J., Zhao, C., Guo, J., Li, D., Sun, L., Du, B., Ding, Y., & Song, Y. (2018). Attentional selection and suppression in children and adults. *Dev Sci*, *21*(6), e12684. https://doi.org/10.1111/desc.12684

Sutton, R. S., & Barto, A. G. (2018). Reinforcement Learning: An Introduction. *MIT Press, Cambridge, MA, Second Edition*.

Turoman, N., Tivadar, R. I., Retsa, C., Maillard, A. M., Scerif, G., & Matusz, P. J. (2021). The development of attentional control mechanisms in multisensory environments. *Dev Cogn Neurosci*, *48*, 100930. https://doi.org/10.1016/j.dcn.2021.100930

van den Berg, B., Geib, B. R., San Martín, R., & Woldorff, M. G. (2019a). A key role for stimulus-specific updating of the sensory cortices in the learning of stimulus-reward associations. (1749-5024 (Electronic)).

van den Berg, B., Geib, B. R., San Martín, R., & Woldorff, M. G. (2019b). A key role for stimulus-specific updating of the sensory cortices in the learning of stimulus-reward associations. *Soc Cogn Affect Neurosci*, *14*(2), 173-187. https://doi.org/10.1093/scan/nsy116

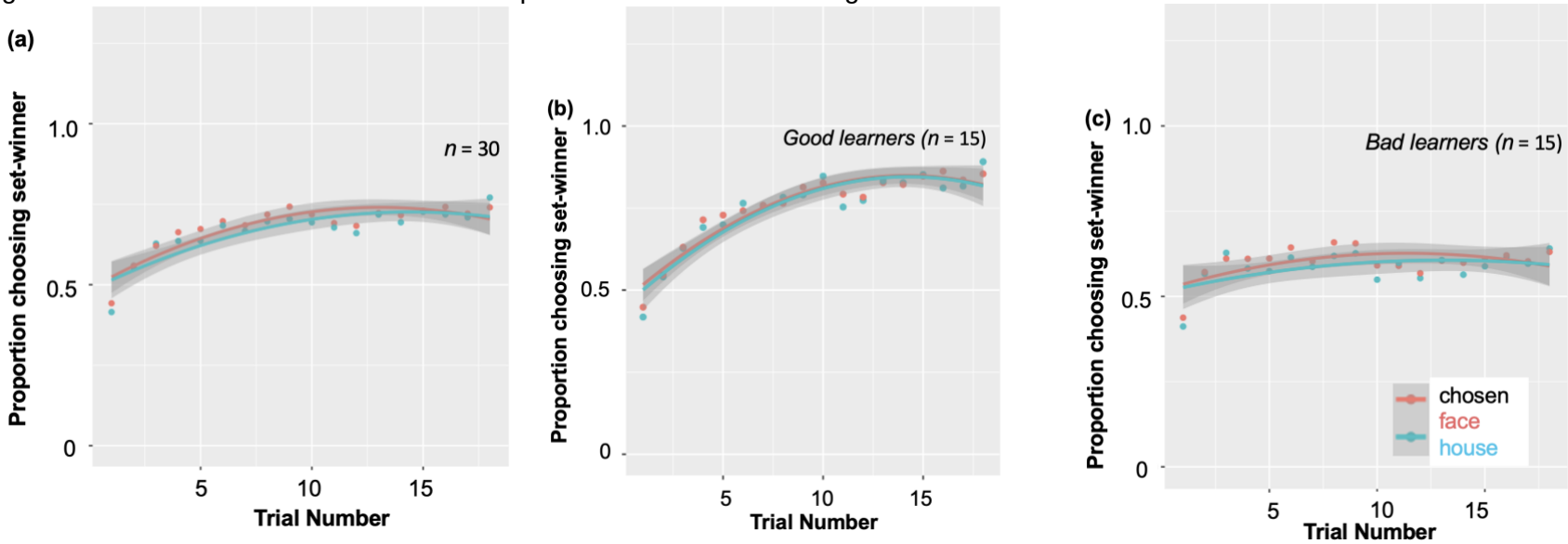
von Borries, A. K., Verkes, R. J., Bulten, B. H., Cools, R., & de Bruijn, E. R. (2013). Feedback-related negativity codes outcome valence, but not outcome expectancy, during reversal learning. *Cogn Affect Behav Neurosci*, *13*(4), 737-746. https://doi.org/10.3758/s13415-013-0150-1

Walker, D. M., Bell, M. R., Flores, C., Gulley, J. M., Willing, J., & Paul, M. J. (2017). Adolescence and Reward: Making Sense of Neural and Behavioral Changes Amid the Chaos. *J Neurosci*, *37*(45), 10855-10866. https://doi.org/10.1523/jneurosci.1834-17.2017

Worthy, D. A., & Maddox, W. T. (2014). A Comparison Model of Reinforcement-Learning and Win-Stay-Lose-Shift Decision-Making Processes: A Tribute to W.K. Estes. *J Math Psychol*, *59*, 41-49. https://doi.org/10.1016/j.jmp.2013.10.001

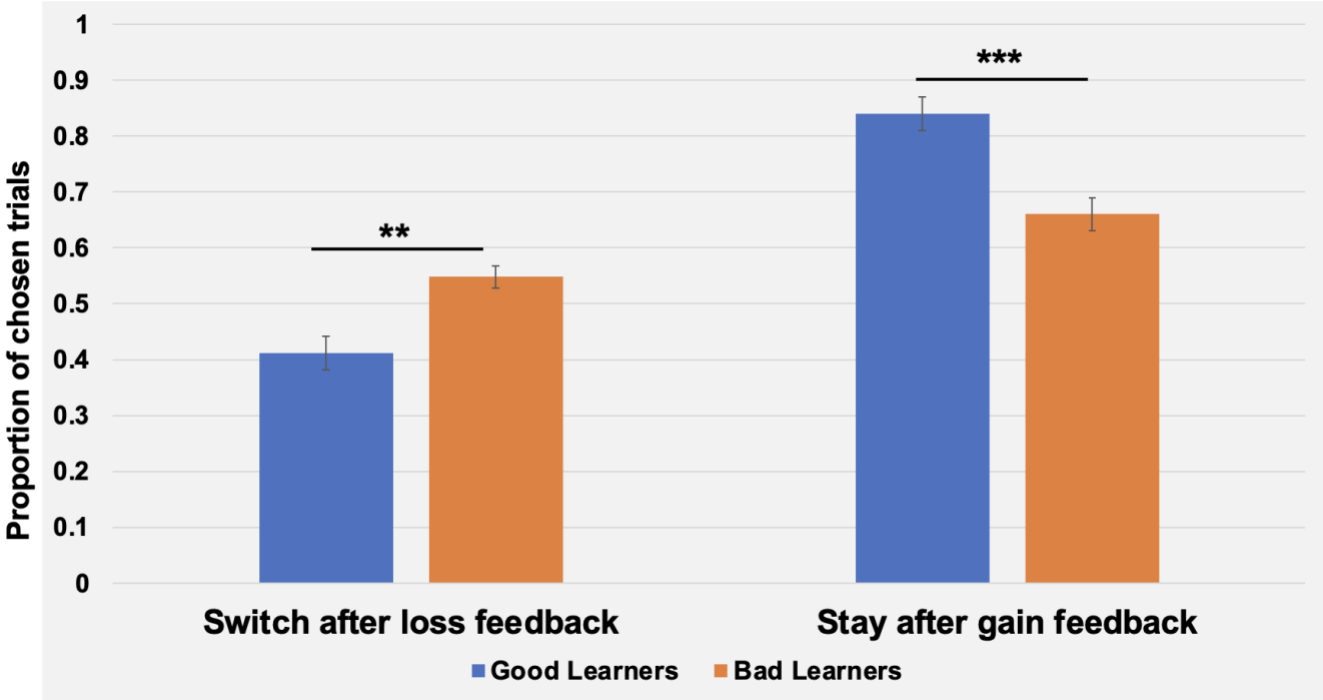
Wu, Y., & Zhou, X. (2009). The P300 and reward valence, magnitude, and expectancy in outcome evaluation. *Brain Res*, *1286*, 114-122. https://doi.org/10.1016/j.brainres.2009.06.032

Figure 1. Individual differences in behavioral probabilistic reward learning



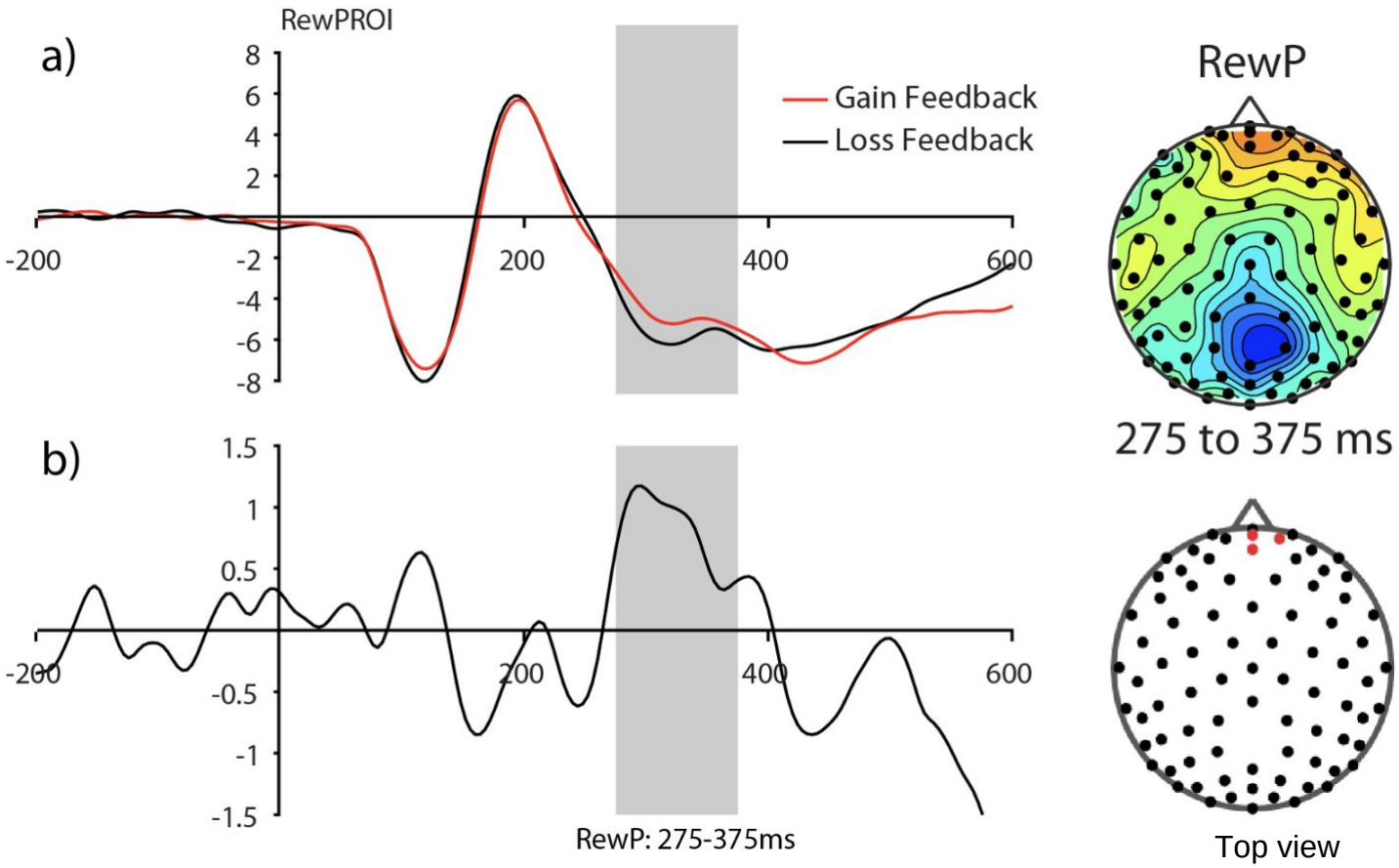

*Note.* Behavioral reward learning: During 18 blocks, the proportion of choosing set-winner for all participants ranged from 0.40 to 0.73. The individual learning curves revealed very large individual difference in reward learning, as the final proportion of choosing set winners in a particular block ranged from 0.26 to 1.0. Due to this large individual difference observed, we divided participants into good vs. bad learners with a median split method.

Figure 2. Behavioral strategies during the Child-version of Probabilistic Reward Learning (C-PLearn)



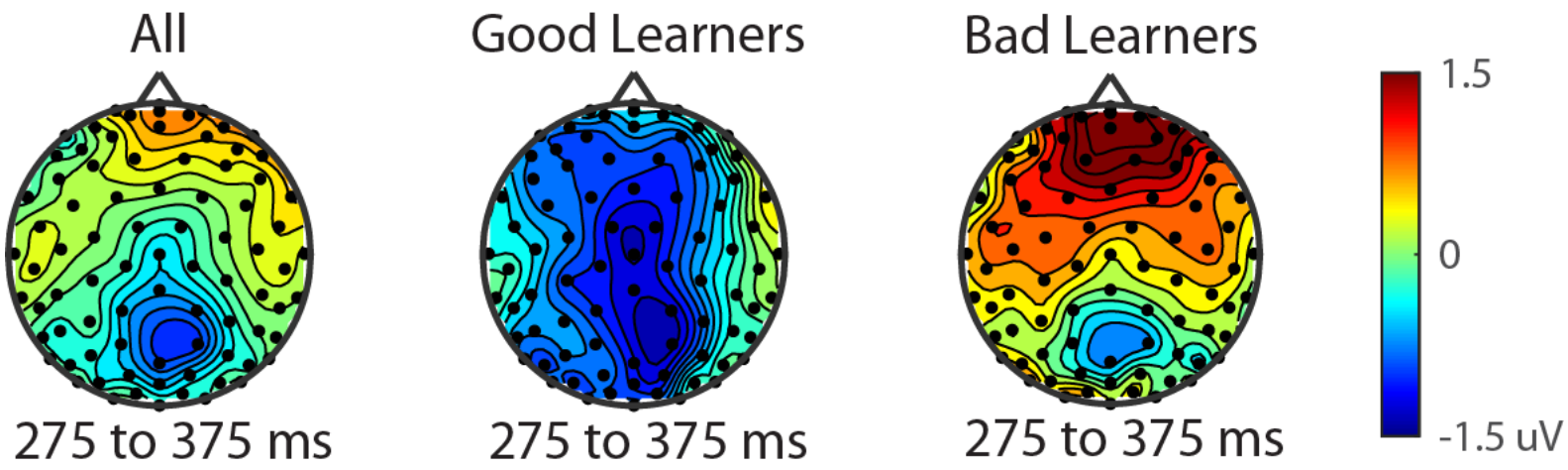

*Note.* Bad learners with low learning rates are more likely to change their choice on the next trial following loss feedback (i.e., Lose-Shift strategy). On the other hand, good learners with high learning rates are more like to make the same choice on the next trial following gain feedback (i.e., Win-Stay strategy). The error bars indicate standard error. ***p* = .007, ****p* <.001.

Figure 3.1. The Frontal reward-related potential (gain vs. loss feedback) in preadolescents (n=30):



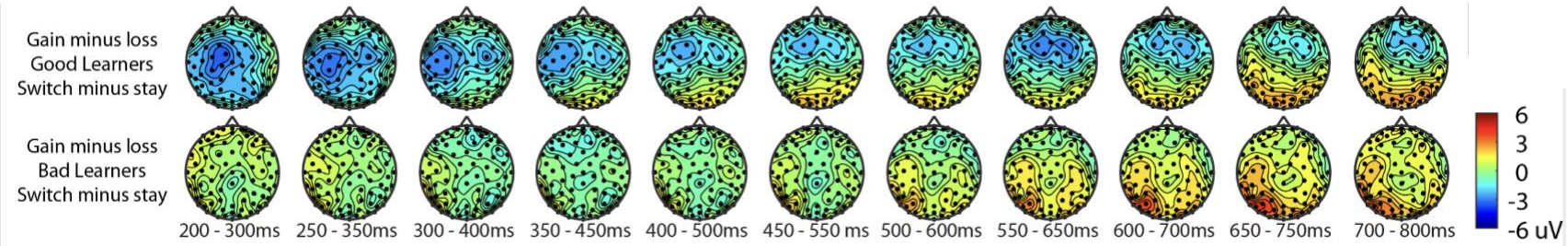

*Note.* (a) Grand average feedback-cue evoked responses for Gain and Loss reward feedback from the fronto-central ROI (E10, E11, E16) (see red dots in the topography for the location of the ROI).
(b) The reward-related potential difference waveform during Gain minus Loss reward feedback. The reward-related potential was measured from 275 to 375 ms in a fronto-central ROI.

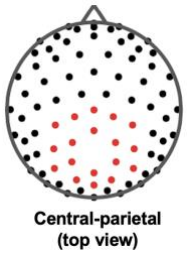
Figure 3.2. The reward-related positivity between good and bad learners




*Note.* The comparison of the reward-related potential ERP marker measured from 275 to 375ms in the fronto-central ROI between good and bad learners. Dark red colored electrode sites in the frontal ROI shows significantly greater reward-related positivity amplitudes for bad learners (i.e., those with low learning rates) compared to good learners (i.e., those with high learning rates).

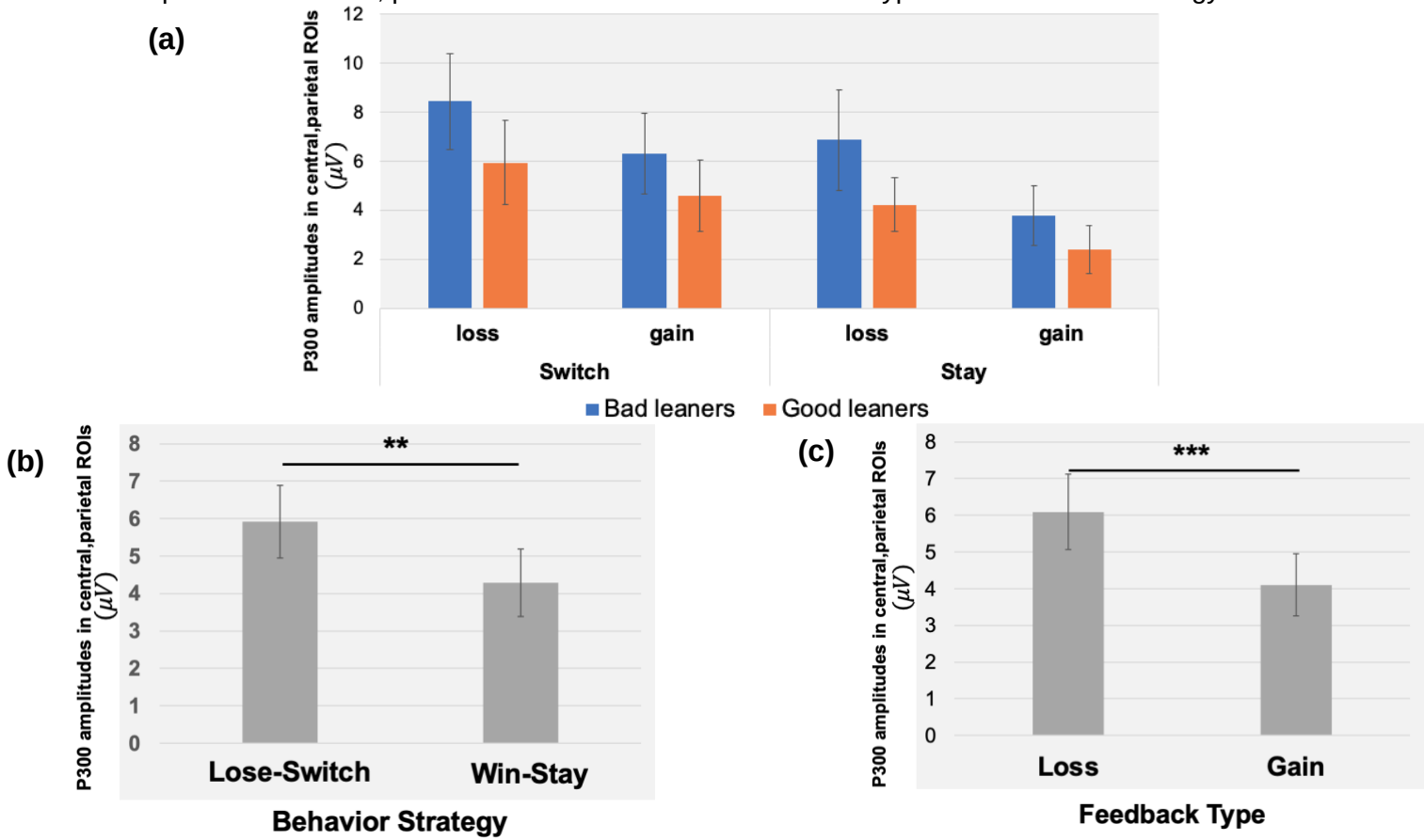
Figure 4.1. Topography of P300 during gain vs. loss feedback as a function of behavioral strategy




*Note.* P300 amplitude differences in central and parietal ROIs when participants switched their responses on the next trial after gain feedback vs. after loss feedback. The red dot in the topography at the bottom shows the location of the central-parietal ROI.
The top row: warm colored electrode sites indicate bigger P300 amplitudes on stay trials for good learners with high learning rates.
The bottom row: warm colored electrode sites indicate bigger P300 on stay trials for bad learners with low learning rates.

Figure 4.2 P300 amplitudes in central, parietal ROIs as a function of feedback type and behavioral strategy



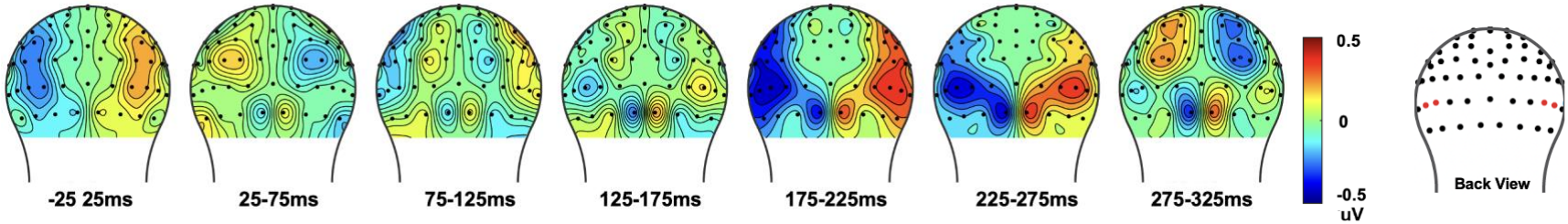

*Note.* The bar indicates standard error; ***p* = 0.01, ****p* =.005.

(a) Using P300 amplitudes in the central-parietal ROIs between 400 and 600 ms for good and bad learners, we found that bad learners (i.e., those with low learning rates), relative to good learners, showed higher P300 amplitude differences in central, parietal ROIs when they switched their responses on the next trial after gain feedback vs. after loss feedback.

(b) For all children, we found greater P300 amplitudes in the central-parietal ROI when participants switched their choice on the next trial after loss feedback (i.e., lose-switch strategy) compared to when they chose the same choice after gain feedback (i.e., win-stay strategy).

(c) The P300 amplitudes in the same central-parietal ROIs were higher when participants received loss feedback compared to gain feedback.

Figure 5. The cue-evoked attentional bias of N2pc towards the later chosen side




*Note.* To increase signal-to-noise ratio, we collapsed over whether the chosen image was that of a face or a house and analyzed the effect of choice behavior on the N2pc over the course of the 18-trial set. The cue-evoked N2pc amplitudes measured after onset of the cue image pair, from corresponding left and right occipital ROIs was found from 175 to 225ms (see red dots in the topography for the location of the ROI electrodes).