

# Striatum–Medial Prefrontal Cortex Connectivity Predicts Developmental Changes in Reinforcement Learning

Wouter van den Bos<sup>1,2,3</sup>, Michael X. Cohen<sup>4</sup>, Thorsten Kahnt<sup>5</sup> and Eveline A. Crone<sup>1,2,4,6</sup>

<sup>1</sup>Institute of Psychology, Leiden University, 2333 AK Leiden, the Netherlands, <sup>2</sup>Leiden Institute for Brain and Cognition, 2300 RC Leiden, the Netherlands, <sup>3</sup>Department of Psychology, Stanford University, Stanford, CA 94305, USA, <sup>4</sup>Department of Psychology, University of Amsterdam, 1018 WB Amsterdam, the Netherlands, <sup>5</sup>Bernstein Center for Computational Neuroscience, Charité—Universitätsmedizin Berlin, 10115 Berlin, Germany and <sup>6</sup>Department of Radiology, Leiden University Medical Center, 2300 RC Leiden, the Netherlands

Address correspondence to Wouter van den Bos, Department of Psychology, Stanford University, Jordan Hall, 450 Serra Mall, Stanford, CA 94305, USA. Email: wvdmbos@stanford.edu.

**During development, children improve in learning from feedback to adapt their behavior. However, it is still unclear which neural mechanisms might underlie these developmental changes. In the current study, we used a reinforcement learning model to investigate neurodevelopmental changes in the representation and processing of learning signals. Sixty-seven healthy volunteers between ages 8 and 22 (children: 8–11 years, adolescents: 13–16 years, and adults: 18–22 years) performed a probabilistic learning task while in a magnetic resonance imaging scanner. The behavioral data demonstrated age differences in learning parameters with a stronger impact of negative feedback on expected value in children. Imaging data revealed that the neural representation of prediction errors was similar across age groups, but functional connectivity between the ventral striatum and the medial prefrontal cortex changed as a function of age. Furthermore, the connectivity strength predicted the tendency to alter expectations after receiving negative feedback. These findings suggest that the underlying mechanisms of developmental changes in learning are not related to differences in the neural representation of learning signals per se but rather in how learning signals are used to guide behavior and expectations.**

**Keywords:** brain maturation, development, fMRI, functional connectivity, reinforcement learning

## Introduction

The ability to learn contingencies between actions and positive or negative outcomes in a dynamic environment forms the foundation of adaptive behavior (Rushworth and Behrens 2008). Learning from feedback in probabilistic environments is sensitive to developmental changes, given the developmental improvements in learning from positive and negative feedback which are observed until early adulthood (Crone and van der Molen 2004; Hooper et al. 2004; Huizinga et al. 2006). Intriguingly, prior neuroimaging studies have demonstrated developmental differences in neural circuits associated with learning from feedback in a fixed static learning environment (Crone et al. 2008; van Duijvenvoorde et al. 2008). These studies show that dorsolateral prefrontal cortex (DLPFC) and parietal cortex are increasingly engaged when receiving negative feedback. However, in a probabilistic learning environment, learning takes place gradually over trials, and both positive and negative feedback informs future behavior. Therefore, an important question concerns the neural mechanisms that underlie developmental differences in probabilistic learning.

A crucial aspect of adaptive learning is using feedback to estimate the expected value of the available options. The first step in estimating the expected value is the computation of prediction errors, that is, calculating the difference between expected and experienced outcomes. Prediction errors can be positive, indicating that outcomes are better than expected or negative, indicating that outcomes are worse than expected (Sutton and Barto 1998). Next, these prediction errors are used to update the expected value associated with the chosen option: The expected value increases when the prediction error is positive and decreases when the prediction error is negative.

Prior neuroimaging studies have shown that activity in the ventral striatum, a target area of dopaminergic midbrain neurons, correlates with positive and negative prediction errors (Knutson et al. 2000; Pagnoni et al. 2002; e.g., McClure et al. 2003, 2004; O'Doherty et al. 2003). The relation between prediction errors and subsequent learning is confirmed by studies demonstrating an association between the representation of prediction errors in the striatum and individual differences in performance on probabilistic learning tasks (PLTs) (Pessiglione et al. 2006; Schönberg et al. 2007). Furthermore, several studies have reported increased sensitivity of the striatum in adolescence after receiving monetary rewards or following other emotional stimuli (Galvan et al. 2006; McClure-Tone et al. 2008; Van Leijenhorst et al. 2009). This suggests that developmental differences in striatal sensitivity to rewards might contribute to the observed developmental differences in adaptive behavior. This hypothesis is supported by a recent developmental study that revealed heightened sensitivity in the striatum to positive prediction errors in adolescents relative to children and adults (Cohen et al. 2010).

In contrast, there are also several studies using less salient rewards that have reported differences in adaptive behavior but suggest that there is a stable striatal activation pattern across adolescence (Casey et al. 2004; van Duijvenvoorde et al. 2008; Velanova et al. 2008). However, none of these developmental studies investigated the neural representation of prediction errors directly. Therefore, it is possible that developmental differences in the representation of prediction errors are contributing to developmental changes in adaptive behavior.

Several neuroimaging studies have shown that activity in the medial prefrontal cortex (mPFC) correlates with the expected value of stimuli or actions (for review, see Rangel et al. 2008). Representations of expected values in the mPFC are thought to be updated by means of frontostriatal connections, relating striatal prediction errors to medial prefrontal representations (Pasupathy

and Miller 2005; Frank and Claus 2006; Camara et al. 2009). In support of this hypothesis, recent studies have shown increased functional connectivity between the ventral striatum and mPFC during feedback processing (Camara et al. 2008; Munte et al. 2008). Furthermore, group differences in learning may be related to the connectivity strength between the striatum and the PFC during feedback processing. For example, substance-dependent individuals have an intact striatal representation of prediction errors but are impaired in subsequently using these signals for learning (Park et al. 2010). This study showed that there is a positive relation between the learning speed and the strength of functional connectivity between the striatum and PFC (see also Klein et al. 2007). Therefore, a second possible mechanism that may contribute to developmental changes in adaptive behavior is an increase in striatal-mPFC connectivity. Indeed, there are also still substantial changes in anatomical connectivity between the subcortical structures and the PFC during adolescence (Supekar et al. 2009; Schmithorst and Yuan 2010).

To test these 2 hypotheses, a computational reinforcement learning model was applied to investigate developmental differences in 1) the neural representation of prediction errors and 2) changes in frontostriatal connectivity. Participants of 3 age groups (children ages 8–11, adolescents ages 13–16, and young adults ages 18–22) performed a PLT (Frank et al. 2004) in a magnetic resonance imaging (MRI) scanner. We expect that with age, there is an improvement in learning from probabilistic feedback (Crone and van der Molen 2004; van den Bos et al. 2009). In order to capture age-related changes in learning from positive and negative feedback separately, we use a reinforcement learning model with separate learning rates for positive and negative feedback (Kahnt et al. 2009). The individually estimated trial-by-trial prediction errors generated by this reinforcement model were subsequently used to test whether developmental differences in learning reflect functional differences in the representation of prediction errors and/or developmental changes in the propagation of prediction errors as measured by functional frontostriatal connectivity (Park et al. 2010).

## Materials and Methods

### Participants

Sixty-seven healthy right-handed paid volunteers ages 8–22 participated in the functional MRI (fMRI) experiment. Age groups were based on adolescent development stage, resulting in 3 age groups: children (8–11 years old,  $n = 18$ ; 9 female), mid-adolescents (13–16 years old,  $n = 27$ ; 13 female), and young adults (18–22 years old,  $n = 22$ ; 13 female). A chi-square analysis indicated that gender distribution did not differ between age groups,  $\chi^2(2) = 0.79$ ,  $P = 0.67$ . All participants reported normal or corrected-to-normal vision, and participants or their caregivers indicated an absence of neurological or psychiatric impairments. Participants gave informed consent for the study, and all procedures were approved by the medical ethical committee of the Leiden University Medical Center.

Participants completed 2 subscales (similarities and block design) of either the Wechsler Adult Intelligence Scale or the Wechsler Intelligence Scale for Children in order to obtain an estimate of their intelligence quotient (Wechsler 1991, 1997). There were no significant differences in estimated IQ scores between the different age groups,  $F_{2,66} = 1.63$ ,  $P = 0.20$  (see Table 1).

### Task Procedure

The procedure for the PLT (Frank et al. 2004; van den Bos et al. 2009) was as follows: The task consisted of 2 stimulus pairs (called AB and CD). The stimulus pairs consisted of pictures of everyday objects (e.g., a chair and a clock). Each trial started with the presentation of 1 of the

**Table 1**

Brain regions revealed by whole-brain contrasts

Anatomical region	L/R	BA	Z	MNI coordinates		
				x	y	z
Prediction error						
Ventral striatum	L/R		6.33	−19	13	−8
Right parahippocampal gyrus	R		5.61	37	−13	−37
Medial PFC	L/R	10/11	5.92	2	51	0
PPI (positive > negative)						
Medial prefrontal cortex	L/R	10	6.02	3	44	2
Ventral striatum (caudate and putamen)	L/R		7.50	9	9	3
PPI (positive > negative) × age						
Medial PFC	L	10	5.32	−9	49	−2

Note: Montreal Neurological Institute (MNI) coordinates, peak voxels reported.

2 stimulus pairs, and subsequently, the participant had to choose one (e.g., A or B). Stimuli were presented randomly on the left or the right side of the screen. Participants were instructed to choose either the left or the right stimulus by pressing a button with the index or middle finger of the right hand. Responses had to be given within a 2500-ms window, which was followed by a 1000-ms feedback display (see Fig. 1A). If no response was given within 2500 ms, the text “too slow” was presented on the screen.

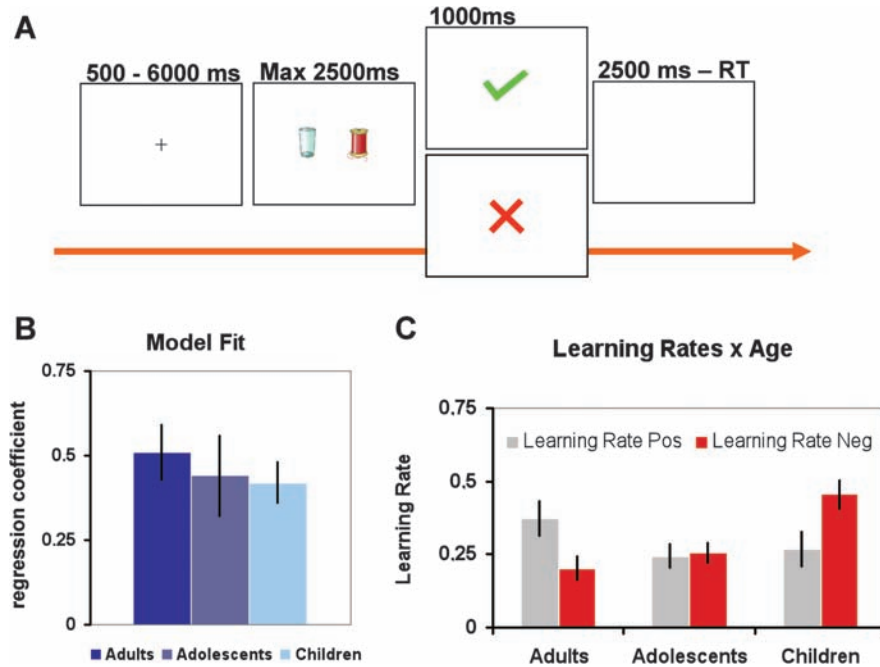
Feedback was probabilistic; choosing stimulus A led to positive feedback on 80% of AB trials, whereas choosing stimulus B led to positive feedback on 20% of these trials. The CD pair procedure was similar, but probability for reward was different; choosing stimulus C led to positive feedback on 70% of CD trials, whereas choosing stimulus D led to positive feedback on 30% in these trials.

Participants were instructed to earn as many points as possible (as indicated by receiving a positive feedback signal) but were also informed that it was not possible to receive positive feedback on every trial. After the instructions and before the scanning session, the participants played 40 practice rounds on a computer in a quiet laboratory to ensure that they understood the task.

In total, the task in the scanner consisted of 2 blocks of 100 trials each: 50 AB trials and 50 CD trials per block. The first and the second block consisted of different sets of pictures, and therefore, participants had to learn a new mapping in both task blocks. The data from the last 60 trials of each block were also reported in another study using a rule-based analysis (van den Bos et al. 2009). The duration of each block was approximately 8.5 min. The stimuli were presented in pseudorandom order with a jittered interstimulus interval (min = 1000 ms, max = 6000 ms) optimized with OptSeq2 (Dale 1999).

### Reinforcement Learning Model

A standard reinforcement learning model (Sutton and Barto 1998) was used to analyze behavioral and neural data (McClure et al. 2003; Cohen and Ranganath 2005; Haruno and Kawato 2006; Frank and Kong 2008; Kahnt et al. 2009). The reinforcement learning model uses the prediction error ( $\delta$ ) to update the decisions weights ( $w$ ) associated with each stimulus (in this case A, B, C, or D). Thus, whenever feedback is better than expected, the model will generate a positive prediction error which is used to “increase” the decision weight of the chosen stimulus (e.g., stimulus A). However, when feedback is worse than expected, the model will generate a negative prediction error, which is used to “decrease” the decision weight of the chosen stimulus (e.g., stimulus B). The impact of the prediction error is usually scaled by the learning rate ( $\alpha$ ). We extended the standard reinforcement learning model by using separate learning rates for positive feedback ( $\alpha_{\text{pos}}$ ) and negative feedback ( $\alpha_{\text{neg}}$ ) (e.g., Kahnt et al. 2009). Thus, positive and negative feedback might have a different impact of the decisions weights. To model trial-by-trial choices, we used the soft-max mechanism to compute the probability ( $P$ ) of choosing a high probability target (A or C) on trial  $t$  as the difference in the decision weights in each trial ( $w_t$ ) associated with each stimulus, passed through a sigmoid function (Montague et al. 2004; Kahnt et al. 2009). For example, when stimulus pair AB is presented, the probability of choosing A is determined by:



**Figure 1.** (A) Participants chose one stimulus by pressing the left or right button and received positive or negative feedback according to probabilistic rules. Two pairs of stimuli were presented to the participants: (1) the AB pair with 80% positive feedback for A and 20% for B and (2) the CD pair with 70% positive feedback for C and 30% for D. (B) Estimated model fits per age group. (C) Estimated learning rates for positive and negative feedback per age group. Error bars represent standard errors in all graphs.

$$P(A)_t = \frac{1}{1 + e^{-\beta \cdot (w(A)_t - w(B)_t)}} \quad (1)$$

where  $\beta$  is the inverse temperature accounting for the stochasticity of the choices.

After each decision, the prediction error ( $\delta$ ) is calculated as the difference between the outcome received ( $r = 1$  for positive feedback and 0 for negative feedback) and the decision weight ( $w_t$ ) for the chosen stimulus:

$$\delta_t = r_t - w(\text{chosen\_stimulus}) \quad (2)$$

Subsequently, the decision weights are updated according to:

$$w_{t+1} = w_t + \lambda \times \alpha(\text{outcome})_t \times \delta_t \quad (3)$$

where  $\lambda$  is 1 for the chosen and 0 for the unchosen stimulus,  $\alpha(\text{outcome})$  is a set of learning rates for positive ( $\alpha_{\text{pos}}$ ) and negative feedback ( $\alpha_{\text{neg}}$ ), which scale the effect of the prediction error on the future decision weights and thus subsequent decisions. For example, a high learning rate for positive feedback but a low learning rate for negative feedback indicates that positive feedback has a high impact on future behavior, whereas negative feedback will hardly change future behavior. These 2 learning rates were individually estimated by fitting the model predictions ( $P(\text{high probability stimulus})$ ) to participants' actual decisions. We used the multivariate constrained minimization function (fmincon) of the optimization toolbox implemented in MATLAB 6.5 for this fitting procedure. Initial values for learning rates were  $\alpha_{\text{pos}} = \alpha_{\text{neg}} = 0.5$  and for action values,  $u(\text{left}) = u(\text{right}) = 0$ .

Finally, we performed behavioral analyses with an alternative model with just one-learning parameter in order to benchmark the performance of the two-learning parameter model. Model comparisons revealed that the 2 parameter had a superior fit to the behavioral data, according to both the Bayesian and Akaike information criterion (BIC and AIC, see Supplementary Table 2). Because the two-learning rate model provides a better fit, this is used in all subsequent analyses.

### Behavioral Analyses

To examine the correspondence between model predictions and participants' behavior, model predictions were compared with the actual behavior on a trial-by-trial basis. Model predictions based on

estimated learning rates were regressed against the vector of participants' actual choices, and individual regression coefficients were used to compare group differences in model fits. Only when there are no differences in model fit between groups, one can confidently compare model parameters.

Next, we defined 2 dependent variables of behavioral performance to further investigate the relation between model parameters and choice behavior:  $p(\text{lose/shift})$  and  $p(\text{win/stay})$ . "Win-stay" was computed by calculating the proportion of choice repetitions following positive feedback and the total number of positive feedback events. Likewise, "lose-shift" was computed by calculating the proportion of choice shifts following negative feedback and the total number of negative feedback events. To test whether the individually estimated learning rates  $\alpha(\text{win})$  and  $\alpha(\text{loss})$  predict different aspects of participants' behavior, both learning rates were simultaneously regressed against  $p(\text{lose-shift})$  and  $p(\text{win-stay})$ , respectively, using multiple regression.

### Data Acquisition

Participants were familiarized with the scanner environment on the day of the fMRI session through the use of a mock scanner, which simulated the sounds and environment of a real MRI scanner. Data were acquired using a 3.0T Philips Achieva scanner at the Leiden University Medical Center. Stimuli were projected onto a screen located at the head of the scanner bore and viewed by participants by means of a mirror mounted to the head coil assembly. First, a localizer scan was obtained for each participant. Subsequently,  $T_2^*$ -weighted Echo-Planar Images (EPI) (time repetition [TR] = 2.2 s, time echo = 30 ms,  $80 \times 80$  matrix, FOV =  $220, 35, 2.75$  mm transverse slices with 0.28 mm gap) were obtained during 2 functional runs of 232 volumes each. A high-resolution  $T_1$ -weighted anatomical scan and a high-resolution  $T_2$ -weighted matched-bandwidth anatomical scan, with the same slice prescription as the EPIs, were obtained from each participant after the functional runs. Stimulus presentation and the timing of all stimuli and response events were acquired using E-Prime software. Head motion was restricted by using a pillow and foam inserts that surrounded the head.

### fMRI Data Analysis

Data were preprocessed using SPM5 (Wellcome Department of Cognitive Neurology, London). The functional time series were realigned



to compensate for small head movements. Translational movement parameters never exceeded 1 voxel ( $<3$  mm) in any direction for any subject or scan. There were no significant differences in movement parameters between age groups  $F_{2,65} = 0.15$ ,  $P = 0.85$ , (see Supplementary Table S1). Functional volumes were spatially normalized to EPI templates. The normalization algorithm used a 12 parameter affine transformation together with a nonlinear transformation involving cosine basis functions and resampled the volumes to 3-mm cubic voxels. Functional volumes were spatially smoothed using a 8 mm full-width half-maximum Gaussian kernel. The MNI305 template was used for visualization, and all results are reported in the MNI305 stereotaxic space (Cosoco et al. 1997).

Statistical analyses were performed on individual participants' data using the general linear model (GLM) in SPM5. The fMRI time series data were modeled by a series of events convolved with a canonical hemodynamic response function (HRF). The presentation of the feedback screen was modeled as 0 duration events. The stimuli and responses were not modeled separately as these occurred in one prior or overlapping EPI images as feedback presentation.

To investigate the neural responses to feedback valence and prediction errors, we set up a GLM with the onsets of each feedback type (positive and negative) as regressors. In this model, the stimulus functions for feedback were parametrically modulated by the trial-wise prediction errors derived from the reinforcement learning model. The modulated stick functions were convolved with the canonical HRF. These regressors were then orthogonalized with respect to the onset regressors of positive and negative feedback trials and regressed against the blood oxygen level-dependent (BOLD) signal.

Finally, to investigate linear and quadratic age trends, we applied polynomial expansion analysis (Büchel et al. 1996) with age as continuous variable, using the forward model selection as described by Büchel et al. (1998). Thresholds were set to  $P < 0.05$  Family Wise Error with an extend threshold of 10 continuous voxels for the whole group analyses. Analyses of age trends were set to  $P < 0.001$  uncorrected with an extend threshold of 20 continuous voxels, reporting the SPM5 implemented small volume correction (SVC) FWE corrected  $P$  values, using the whole group psychophysiological interaction (PPI) mPFC as a volume of interest 2.

### Region of Interest Analyses

We used the Marsbar toolbox for use with SPM5 (<http://marsbar.sourceforge.net>, Brett et al. 2002) to perform Region of Interest (ROI) analyses to further characterize patterns of activation and estimate individual differences in connectivity measures.

### Functional Connectivity Analyses

To explore the interplay between the ventral striatum and other brain regions during reinforcement-guided decision-making, functional connectivity was assessed using PPI analysis (Friston 1994; Cohen et al. 2005, 2008). The functional whole-brain mask, in which activity correlated significantly with prediction errors for the whole group, was masked with an anatomical striatum ROI of the Marsbar toolbox that included the bilateral caudate, putamen, and nucleus accumbens, to create the seed ROI. The method used here relies on correlations in the observed BOLD time series data and makes no assumptions about the nature of the neural event that contributed to the BOLD signal (Cohen et al. 2008). For each model, the entire time series over the experiment was extracted from each subject in the clusters of the (left and right) ventral striatum. Regressors were then created by multiplying the normalized time series of each ROI with condition vectors that contained ones for 4 TRs after positive or negative prediction errors and zeros otherwise (see also Cohen and Ranganath 2005; Kahnt et al. 2009; Park et al. 2010). Thus, the 2 condition vectors of positive and negative prediction errors (containing ones and zeros) were each multiplied with the time course of each ROI. These regressors were then used as covariates in subsequent analyses.

The time series between the left and right hemispheres for the ventral striatum were highly correlated ( $r = 0.89$ ). Therefore, parameter estimates of left- and right structures were collapsed, and thus,

represent the extent to which feedback-related activity in each voxel correlates with feedback-related activity in the bilateral ventral striatum.

Individual contrast images for positive versus negative feedback were computed and entered into second-level one-sample  $t$ -tests. In order to find age-related differences in the whole-brain analyses of functional connectivity with the ventral striatum, we performed a second-level regression analysis with a regressor for age.

## Results

### Behavioral Data

#### Reinforcement Learning

First, we assessed how the model parameters differed between age groups. First of all, there was a good fit of the model to participants' behavior; the average regression coefficient was significantly above zero for all age groups (all  $P$ 's  $< 0.001$ , Fig. 1B). Importantly, the model fit did not differ significantly between groups ( $F_{2,64} = 0.96$ ,  $P = 0.38$ ), reassuring that parameters estimations could be compared between groups. Importantly, we also found no significant relation between age and the value of the stochasticity parameter  $\beta$  ( $r = 0.05$ ,  $P = 0.74$ ). This indicates that behavioral differences are not due to age differences in choice stochasticity. Furthermore, previous behavioral analyses suggest that there are no significant difference in learning speed and that participants of all ages reach a stable behavioral pattern after about 60 trials, showing matching behavior (see van den Bos et al. 2009).

Next, a 2 (learning parameters)  $\times$  3 (age groups) analysis of variance tested for age differences in learning from positive and negative feedback. This analysis showed a significant group by parameter interaction ( $F_{2,64} = 12.34$ ,  $P < 0.001$ , see Fig. 1C), and post hoc tests revealed that there was an age-related decrease in  $\alpha_{\text{neg}}$ ,  $F_{2,67} = 9.87$ ,  $P < 0.001$  and a marginal age-related increase in  $\alpha_{\text{pos}}$ ,  $F_{2,67} = 2.73$ ,  $P = 0.06$ .

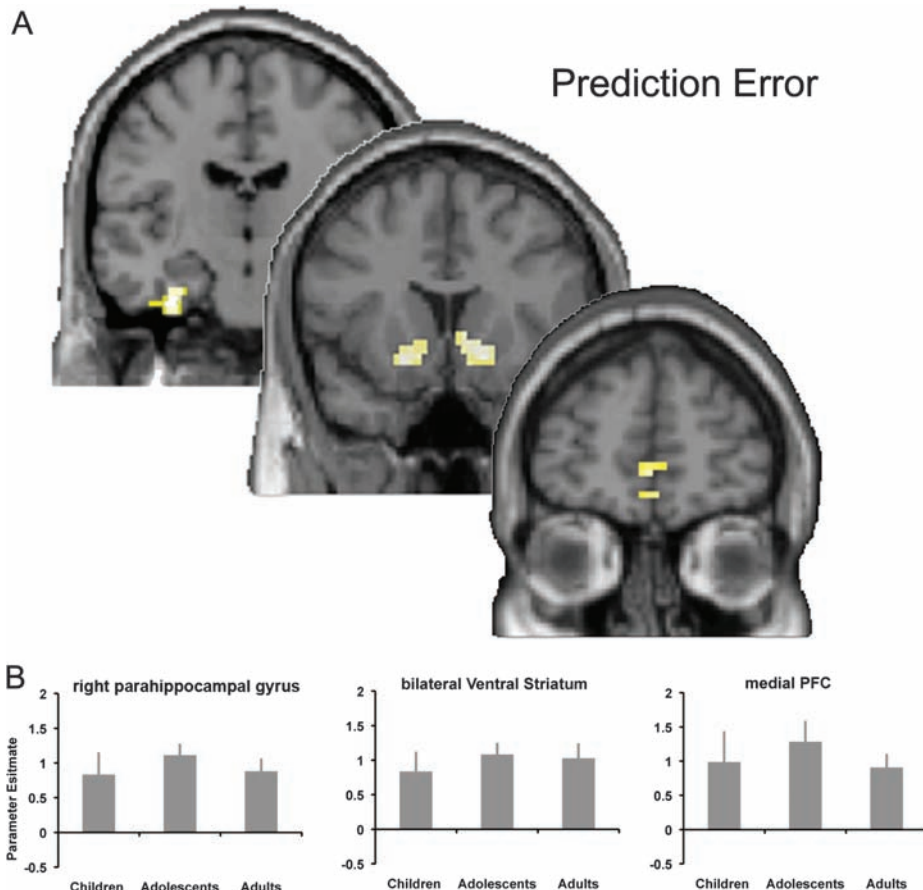
Finally, to assess whether different learning rates captured different aspects of behavior,  $\alpha_{\text{win}}$  and  $\alpha_{\text{loss}}$  were simultaneously regressed against the 2 dependent variables of this study [ $p(\text{win/stay})$  and  $p(\text{lose/switch})$ ]. A multiple regression of both learning rates on  $p(\text{win/stay})$  fitted significantly ( $r = 0.51$ ,  $F_{2,64} = 11.05$ ,  $P < 0.001$ ), but only  $\alpha_{\text{win}}$  ( $b_{\alpha(\text{win})} = 0.49$ ,  $t_{64} = 4.46$ ,  $P < 0.001$ ) and not  $\alpha_{\text{loss}}$  ( $b_{\alpha(\text{loss})} = -0.27$ ,  $t_{64} = -2.04$ ,  $P = 0.08$ ) contributed significantly to the regression. In contrast, in the regression against  $p(\text{lose/switch})$  ( $r = 0.33$ ,  $F_{2,64} = 6.85$ ,  $P < 0.01$ ),  $\alpha_{\text{loss}}$  ( $b_{\alpha(\text{loss})} = 0.32$ ,  $t_{64} = 2.55$ ,  $P < 0.01$ ) but not  $\alpha_{\text{win}}$  ( $b_{\alpha(\text{win})} = -0.218$ ,  $t_{64} = -1.83$ ,  $P = 0.08$ ) contributed significantly.

Taken together, these results show that the learning rates captured different behavioral aspects of reinforcement-guided decision-making. The results further show that mainly the age-related decrease in the influence of negative feedback on expected values underlies developmental differences in adaptive behavior.

### fMRI Results

#### Model-Based fMRI

Across all participants, individually generated trial-wise prediction errors (positive and negative combined) correlated significantly with BOLD responses in bilateral ventral striatum, mPFC, and the right parahippocampal gyrus (Fig. 2A and Table 1). Activity in the ventral striatum was localized at an area



**Figure 2.** (A) Regions in the mPFC, ventral striatum, and parahippocampal gyrus in which BOLD signal was significantly correlated with prediction errors. Thresholded at  $P < 0.05$ , FWE,  $k > 10$ . (B) Parameter estimates of the prediction errors per age group in the functionally defined ROIs for the mPFC, ventral striatum, and parahippocampal gyrus.

comprising the ventral intersection between the putamen and the head of the caudate. Tests for positive and negative prediction errors separately revealed comparable results.

Whole-brain regression analyses for age differences revealed no linear or nonlinear age group differences (Fig. 2B). This analysis was repeated for positive and negative prediction errors separately, and these analyses also revealed no linear or nonlinear age effects. These findings demonstrate that prediction errors (positive or negative) are not represented differently between the 3 age groups.

#### Functional Connectivity

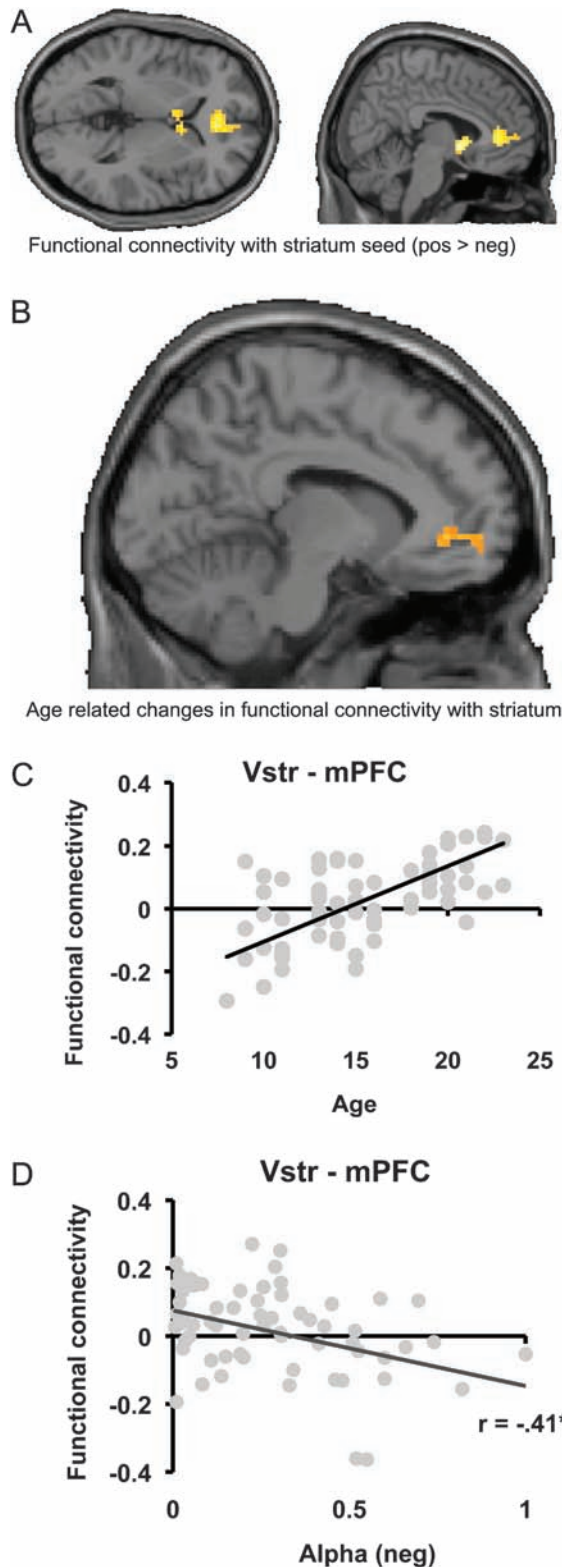
Functional connectivity between the striatum and other brain regions was assessed during processing of negative and positive feedback using PPI. The contrast used for testing functional connectivity was positive > negative feedback. Note that the vectors for positive feedback events contain all positive prediction error events, and the vectors for negative feedback events contain all negative prediction error events. Significantly enhanced functional connectivity was found during positive > negative feedback between the bilateral ventral striatum seed and the mPFC (Fig. 3A). The opposite contrast (negative > positive feedback) did not reveal any significant changes in functional connectivity.

Next, we examined age differences in ventral striatum connectivity by adding age as a regressor to the second-level PPI analysis. These analyses revealed age-related increases in

functional connectivity of the ventral striatum seed with the mPFC (BA32/10) for positive > negative feedback (Fig. 3B) at an uncorrected threshold of  $P < 0.001$  and  $k > 20$  voxels (SVC: FWE,  $P < 0.02$ ). No other areas were found when testing for nonlinear age effects in functional connectivity.

To further illustrate the age-related changes in frontostriatal connectivity, we extracted the strength of functional connectivity between ventral striatum and mPFC for each participant and plotted it against age as a continuous variable (Fig. 3C). This plot reveals that the connectivity pattern shifts from a stronger connection after negative feedback for the youngest participants toward a stronger connection after positive prediction errors for the oldest participants.

Finally, we performed ROI analyses to investigate whether striatum-mPFC connectivity was related to the individual learning parameters. The differential connectivity strength (positive > negative) between the ventral striatum and mPFC ROI was used to predict the individual differences in learning rates for positive and negative feedback. The relative connectivity measure correlated negatively with the learning rate for negative feedback ( $r = -0.41$ ,  $P < 0.001$ , Fig. 3D) and, moderately, positively with the learning rate for positive feedback ( $r = 0.26$ ,  $P = 0.06$ ). Thus, there was stronger striatum-mPFC coupling during negative > positive feedback in participants for whom negative feedback had a relatively large impact on future expected value, whereas the reverse was true (i.e., stronger coupling during positive > negative feedback) in



**Figure 3.** (A) Regions that showed increased functional connectivity with the striatal seed region after positive compared with negative feedback. Thresholded at  $P < 0.05$ , FWE,  $k > 10$ . (B) Region in the mPFC that revealed age-related changes in functional connectivity with the striatal seed region. Thresholded at  $P < 0.001$ , uncorrected,  $k > 20$ . (C) Scatterplot depicting the relationship between the functional connectivity measure of the striatum–mPFC (positive > negative feedback) and age. (D) Scatterplot depicting the relationship between the functional connectivity measure of the striatum–mPFC (positive > negative feedback) and learning rate ( $\alpha_{\text{neg}}$ ).

participants for whom positive feedback had a relatively large impact on future expected value.

To summarize, increased functional connectivity between the ventral striatum and mPFC was observed during processing of positive feedback compared with negative feedback. Furthermore, this analysis revealed that the relative strength of the striatum–mPFC connectivity is correlated positively with age but negatively with the learning rate for negative feedback.

## Discussion

The goal of this study was to examine developmental changes in the neural mechanisms of probabilistic learning. The reinforcement model showed that with increasing age, negative feedback had decreasing effects on future expected values. Imaging analyses revealed that neural activation to prediction errors did not differ between age groups; however, age differences in the learning rates were associated with an age-related increase in functional connectivity between the ventral striatum and the mPFC.

### Developmental Changes in Learning Rates

Using a reinforcement learning model, we were able to disentangle differences in sensitivity to positive and negative feedback by estimating learning rates for positive and negative feedback separately. These estimated learning rates reflect the degree to which the future expected value of a stimulus will be changed after positive or negative prediction errors. Importantly, the model revealed that developmental differences in adaptive behavior were not related to differences in stochasticity in choice behavior. However, the analyses showed that with age, there is a decrease in the learning rate for negative prediction errors ( $\alpha_{\text{neg}}$ ). This finding indicates that with increasing age, particularly, the impact of negative prediction errors on the future expected value decreases. Furthermore, as expected, the individual differences in learning rates were related to shifting behavior, showing a relation between updating of expected value and decision strategies. These results are consistent with developmental studies that have shown with increasing age, participants are less influenced by irrelevant negative feedback (Crone et al. 2004; Eppinger et al. 2009).

Taken together, the results show that an extended reinforcement model is 1) able to identify different computational processes involved in adaptive behavior and 2) reveal an important (single) parameter underlying age-related changes in adaptive learning, the learning rate for negative learning signals. Additionally, given that the model fits the behavior of all ages equally well, it provides a solid basis for exploring the neurodevelopment changes in representing and the processing of learning signals.

### Neural Representation of Prediction Errors

Consistent with previous studies, trial-by-trial prediction errors generated by the reinforcement learning model correlated with activity of a network of areas including the ventral striatum and the mPFC (Pagnoni et al. 2002; McClure et al. 2003; O'Doherty et al. 2003; Cohen and Ranganath 2005). This result indicates that these areas are sensitive to differences in expected versus received feedback, showing increased activation when feedback is better than expected and decreased activation when the feedback is worse than expected. Interestingly, our



analyses did not reveal any (linear or nonlinear) age-related differences in (positive or negative) prediction error-related activity in the striatum.

These findings are consistent with prior studies using cognitive learning tasks, which have also reported stable striatal activation patterns across adolescence (Casey et al. 2004; van Duijvenvoorde et al. 2008; Velanova et al. 2008). However, the results of the current study provide different findings in comparison with affective decision-making paradigms. These studies have reported a peak in sensitivity of the striatum in adolescence after receiving monetary rewards or highly emotional stimuli (Galvan et al. 2006; McClure-Tone et al. 2008; Van Leijenhorst et al. 2009), which may be related to adolescent typical changes in the dopamine system (for a review, see Galvan 2010). Importantly, a recent developmental study of reward-based learning using a comparable reinforcement model with a single learning rate (for both negative and positive feedback), has also shown heightened sensitivity to positive prediction errors in adolescents compared with children and adults (Cohen et al. 2010) (It should be noted, however, that Cohen and colleagues compared different age groups, as adolescence in this study was defined as the age range 14–19 years and adulthood as 25–30 years. In this respect, the findings of the current study and the findings of Cohen et al. are not directly comparable). Interestingly, Cohen et al. (2010) observed adolescent-specific increases in reaction times for large relative to small rewards. This suggests that particularly, in the presence of salient rewards, adolescents show increased striatal sensitivity, which in turn might bias decision-making processes. One possibility is that during adolescence, the presence of salient rewards increases the baseline level of striatal dopamine, which in turn increases sensitivity to positive prediction errors and may even decrease the sensitivity to negative prediction errors (Frank et al. 2004; Frank and Claus 2006). In future studies, it will be important to further examine how the prediction error representation can be modulated by the use of specific reward magnitude manipulations, and how these manipulations affect decision-making parameters.

### ***Developmental Changes in Striatum–mPFC Connectivity***

Connectivity analyses revealed that during feedback processing, the seed region in the ventral striatum sensitive to prediction errors showed increased functional connectivity with the mPFC during positive compared with negative feedback. This pattern of connectivity is consistent with several studies that have shown feedback-related changes in functional connectivity of the striatum (for a review, see Camara et al. 2009). In contrast to the neural representation of prediction errors, subsequent analyses revealed age-related changes in striatum–mPFC functional connectivity. The pattern shifted toward stronger connectivity after positive feedback with increasing age. Importantly, the striatum–mPFC connectivity strength was negatively correlated with the negative learning rate. Taken together, these results suggest that the age-related increase in striatum–mPFC connectivity underlies changes in adaptive behavior. In other words, developmental changes in learning are not related to differences in the computation of learning signals per se, but rather to differences in how learning signals are used to update future expectations and subsequent behavior.

Given that during adolescent development, there are still substantial changes in structural connectivity within the PFC (Schmithorst and Yuan 2010) and it could be hypothesized that the developmental differences in striatum–mPFC functional connectivity are related to changes in structural connectivity between these 2 structures (Cohen et al. 2008). In future developmental studies, it will be of interest to combine measures of structural and functional connectivity in order to further explore this hypothesis.

A final question concerns how these results relate to previous developmental studies on feedback processing in deterministic environments (Crone et al. 2008; van Duijvenvoorde et al. 2008). Learning theories have suggested 2 separate learning strategies (Daw et al. 2005; Maia 2009); a model-based strategy that operates on explicit task representations, such as rules describing the reward contingencies given the current state, and a model-free strategy that uses feedback directly to compute action values without any explicit model of the environment. Furthermore, research has suggested that the relative contribution of each learning strategy might be dependent on their respective certainties (Doya et al. 2002; Daw et al. 2005).

Thus, given the deterministic or rule-based structure of previous experimental paradigms, it is likely that reported developmental changes in the DLPFC-parietal network represent differences in the learning system that operates on task representations, whereas the current study shows developmental differences in neural systems that subserve the model-free computational strategy (see also Galvan et al. 2006; Cohen et al. 2010). This interpretation is supported by a recent study that showed that updating model-based task representations relies on the DLPFC-parietal network, whereas model-free feedback updating was associated with the striatal activity (Gläscher et al. 2010).

The challenge for future developmental studies will be to disentangle the relative contributions of these learning strategies dependent on the learning context (Daw et al. 2005) and to understand how these 2 strategies, and related neural systems, contribute to developmental changes in feedback learning. An interesting hypothesis is that in a context where learning mainly relies on a model-based strategy, adolescents may be less susceptible to the presence of salient rewards than when learning is mainly based on a model-free strategy.

### **Conclusion**

In the current study, we used a reinforcement learning model to investigate neurodevelopmental changes in the representation and processing of learning signals in a probabilistic environment. The results of this study advance our understanding of the mechanisms underlying developmental changes related to learning in a probabilistic environment.

First, behavioral analyses singled out a specific computational process, updating based on negative prediction errors, which showed developmental differences. Importantly, the age-related differences in updating were also related to shifting behavior after negative feedback. Second, we provide evidence that developmental differences in adaptive learning may not be due to differences in the computation of learning signals, but rather to developmental differences in how learning signals are used to guide behavior and expectations. The imaging results

suggest that the latter process is reflected in the strength of functional connectivity between the striatum and the mPFC.

## Supplementary Material

Supplementary material can be found at: <http://www.cercor.oxfordjournals.org/>

## Funding

Vidi grant from the Netherlands Organisation for Scientific Research to E.A.C.

## Notes

We thank Berna Guroglu and Bianca van den Bulk for help with data collection. *Conflict of Interest*: None declared.

## References

- Brett MC, Anton J-L, Valabregue R, Poline J-B. 2002. Region of interest analysis using an spm toolbox. *Neuroimage*. 16:497.
- Büchel C, Holmes AP, Rees G, Friston KJ. 1998. Characterizing stimulus-response functions using nonlinear regressors in parametric fMRI experiments. *Neuroimage*. 8:140–148.
- Büchel C, Wise RJS, Mummery CJ, Poline JB, Friston KJ. 1996. Nonlinear regression in parametric activation studies. *Neuroimage*. 4:60–66.
- Camara E, Rodriguez-Fornells A, Munte TF. 2008. Functional connectivity of reward processing in the brain. *Front Hum Neurosci*. 2:19–19.
- Camara E, Rodriguez-Fornells A, Ye Z, Munte TF. 2009. Reward networks in the brain as captured by connectivity measures. *Front Neurosci*. 3:350–362.
- Casey BJ, Davidson MC, Hara Y, Thomas KM. 2004. Early development of subcortical regions involved in non-cued attention switching. *Dev Sci*. 7:534–542.
- Cohen JR, Asarnow RF, Sabb FW, Bilder RM, Bookheimer SY, Knowlton BJ, Poldrack RA. 2010. A unique adolescent response to reward prediction errors. *Nat Neurosci*. 13:669–671.
- Cohen MX, Elger CE, Weber B. 2008. Amygdala tractography predicts functional connectivity and learning during feedback-guided decision-making. *Neuroimage*. 39:1396–1407.
- Cohen MX, Heller AS, Ranganath C. 2005. Functional connectivity with anterior cingulate and orbitofrontal cortices during decision-making. *Brain Res Cogn Brain Res*. 23:61–70.
- Cohen MX, Ranganath C. 2005. Behavioral and neural predictors of upcoming decisions. *Cogn Affect Behav Neurosci*. 5:117–126.
- Cosoco CA, Kollokian V, Kwan RKS, Evans AC. 1997. Brainweb: online interface of a 3-d mri simulated brain database. *Neuroimage*. 5:425.
- Crone EA, Jennings JR, Van der Molen MW. 2004. Developmental change in feedback processing as reflected by phasic heart rate changes. *Dev Psychol*. 40:1228–1238.
- Crone EA, van der Molen MW. 2004. Developmental changes in real life decision making: performance on a gambling task previously shown to depend on the ventromedial prefrontal cortex. *Dev Neuropsychol*. 25:251–279.
- Crone EA, Zanolie K, van Leijenhorst L, Westenberg P, Rombouts SA. 2008. Neural mechanisms supporting flexible performance adjustment during development. *Cogn Affect Behav Neurosci*. 8:165–177.
- Dale AM. 1999. Optimal experimental design for event-related fmri. *Hum Brain Mapp*. 8:109–114.
- Daw ND, Niv Y, Dayan P. 2005. Uncertainty-based competition between prefrontal and dorsolateral striatal systems for behavioral control. *Nat Neurosci*. 8:1704–1711.
- Doya K. 2008. Modulators of decision making. *Nat Neurosci*. 11:410–416.
- Eppinger B, Mock B, Kray J. 2009. Developmental differences in learning and error processing: evidence from ERPs. *Psychophysiology*. 46:1043–1053.
- Frank MJ, Claus E. 2006. Anatomy of a decision: striato-orbitofrontal interactions in reinforcement learning, decision making, and reversal. *Psychol Rev*. 113:300–326.
- Frank MJ, Kong L. 2008. Learning to avoid in older age. *Psychol Aging*. 23:392–398.
- Frank MJ, Seeberger LC, O'Reilly RC. 2004. By carrot or by stick: cognitive reinforcement learning in parkinsonism. *Science*. 306:1940–1943.
- Friston KJ. 1994. Functional and effective connectivity in neuroimaging: a synthesis. *Hum Brain Mapp*. 2:56–78.
- Galvan A. 2010. Adolescent development of the reward system. *Front Hum Neurosci*. 4:6.
- Galvan A, Hare TA, Parra CE, Penn J, Voss H, Glover G, Casey BJ. 2006. Earlier development of the accumbens relative to orbitofrontal cortex might underlie risk-taking behavior in adolescents. *J Neurosci*. 26:6885–6892.
- Gläscher J, Daw N, Dayan P, O'Doherty JP. 2010. States versus rewards: dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron*. 66:585–595.
- Haruno M, Kawato M. 2006. Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning. *J Neurophysiol*. 95:948–959.
- Holroyd CB, Coles MG. 2002. The neural basis of human error processing: reinforcement learning, dopamine, and the error-related negativity. *Psychol Rev*. 109:679–709.
- Hooper CJ, Luciana M, Conklin HM, Yarger RS. 2004. Adolescents' performance on the Iowa Gambling Task: implications for the development of decision making and ventromedial prefrontal cortex. *Dev Psychol*. 40:1148–1158.
- Huizinga M, Dolan CV, van der Molen MW. 2006. Age-related change in executive function: developmental trends and a latent variable analysis. *Neuropsychologia*. 44:2017–2036.
- Kahnt T, Park SQ, Cohen MX, Beck A, Heinz A, Wrase J. 2009. Dorsal striatal-midbrain connectivity in humans predicts how reinforcements are used to guide decisions. *J Cogn Neurosci*. 21:1332–1345.
- Klein TA, Neumann J, Reuter M, Hennig J, von Cramon DY, Ullsperger M. 2007. Genetically determined differences in learning from errors. *Science*. 318:1642–1645.
- Knutson B, Westdorp A, Kaiser E, Hommer D. 2000. FMRI visualization of brain activity during a monetary incentive delay task. *Neuroimage*. 12:20–27.
- Maia TV. 2009. Reinforcement learning, conditioning, and the brain: successes and challenges. *Cogn Affect Behav Neurosci*. 9:343–364.
- McClure SM, Berns GS, Montague P. 2003. Temporal prediction errors in a passive learning task activate human striatum. *Neuron*. 38:339–346.
- McClure SM, Laibson DI, Loewenstein G, Cohen JD. 2004. Separate neural systems value immediate and delayed monetary rewards. *Science*. 306:503–507.
- Monk CS, Klein RG, Telzer EH, Schroth EA, Mannuzza S, Moulton JL 3rd, Guardino M, Masten CL, McClure-Tone EB, Fromm S, et al. 2008. Amygdala and nucleus accumbens activation to emotional facial expressions in children and adolescents at risk for major depression. *Am J Psychiatry*. 165:90–98.
- Montague P, Hyman SE, Cohen JD. 2004. Computational roles for dopamine in behavioural control. *Nature*. 431:760–767.
- Munte TF, Heldmann M, Hinrichs H, Marco-Pallares J, Krämer UM, Sturm V, Heinze HJ. 2008. Nucleus accumbens is involved in human action monitoring: evidence from invasive electrophysiological recordings. *Front Hum Neurosci*. 4:12.
- O'Doherty JP, Dayan P, Friston K, Critchley H, Dolan RJ. 2003. Temporal difference models and reward-related learning in the human brain. *Neuron*. 38:329–337.
- Pagnoni G, Zink CF, Montague PR, Berns GS. 2002. Activity in human ventral striatum locked to errors of reward prediction. *Nat Neurosci*. 5:97–98.
- Park SQ, Kahnt T, Beck A, Cohen MX, Dolan RJ, Wrase J, Heinz A. 2010. Prefrontal cortex fails to learn from reward prediction errors in alcohol dependence. *J Neurosci*. 30:7749–7753.
- Pasupathy A, Miller EK. 2005. Different time courses of learning-related activity in the prefrontal cortex and striatum. *Nature*. 433:873–876.
- Pessiglione M, Seymour B, Flandin G, Dolan RJ, Frith CD. 2006. Dopamine-dependent prediction errors underpin reward-seeking behaviour in humans. *Nature*. 442:1042–1045.



- Rangel A, Camerer C, Montague P. 2008. A framework for studying the neurobiology of value-based decision making. *Nat Rev Neurosci*. 9:545–556.
- Rushworth MFS, Behrens TEJ. 2008. Choice, uncertainty and value in prefrontal and cingulate cortex. *Nat Neurosci*. 11:389–397.
- Schmithorst VJ, Yuan WH. 2010. White matter development during adolescence as shown by diffusion MRI. *Brain Cogn*. 72: 16–25.
- Schönberg T, Daw ND, Joel D, O'Doherty JP. 2007. Reinforcement learning signals in the human striatum distinguish learners from nonlearners during reward-based decision making. *J Neurosci*. 27:12860–12867.
- Supekar K, Musen M, Menon V. 2009. Development of Large-Scale Functional Brain Networks in Children. *PLoS Biol*. 7:e1000157.
- Sutton RS, Barto AG. 1998. Reinforcement learning. Cambridge (MA): MIT Press.
- van den Bos W, Guroglu B, van den Bulk BG, Rombouts S, Crone EA. 2009. Better than expected or as bad as you thought? The neurocognitive development of probabilistic feedback processing. *Front Hum Neurosci*. 3:52.
- van Duijvenvoorde ACK, Zanolie K, Rombouts S, Raijmakers MEJ, Crone EA. 2008. Evaluating the negative or valuing the positive? Neural mechanisms supporting feedback-based learning across development. *J Neurosci*. 28:9495–9503.
- Van Leijenhorst L, Zanolie K, Van Meel CS, Westenberg PM, Rombouts SA, Crone EA. 2009. What motivates the adolescent? Brain regions mediating reward sensitivity across adolescence. *Cereb Cortex*. 20:61–69.
- Velanova K, Wheeler ME, Luna B. 2008. Maturation changes in anterior cingulate and frontoparietal recruitment support the development of error processing and inhibitory control. *Cereb Cortex*. 18:2505–2522.
- Wechsler D. 1991. Wechsler intelligence scale for children-third edition. Manual. San Antonio (TX): The Psychological Corporation.
- Wechsler D. 1997. Wechsler adult intelligence scale—third edition. Administration and scoring manual. San Antonio (TX): The Psychological Corporation.