Neuron, Volume *92*

Supplemental Information

An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence

Juliet Y. Davidow, Karin Foerde, Adriana Galván, and Daphna Shohamy

Supplemental results

Testing a standard reinforcement learning model without priors on estimated parameters

When allowing the learning rate and inverse temperature parameters to take on values between negative infinity and positive infinity, the model did not arrive at a solution for 15 adolescents (37% of the group) and 11 adults (35% of the group). For the remaining sample for whom the model could solve, 26 adolescents and 20 adults, the model provided a good fit to the observed behavior across both groups (one-way t-test comparing a null model, $t_{45} = -28.99$, $p < 0.000$) and the model fits did not differ between the groups (independent samples t-test, $t_{44} = 1.8$, $p = 0.08$). The learning rate parameter was lower in the adolescents ($t_{44} = -2.06$, $p = 0.046$; adolescent mean 0.38, standard error of the mean (SEM) 0.06, adult mean 0.55, SEM 0.06) and there was no difference between groups in the inverse temperature parameter ($t_{44} = -0.17$, $p = 0.86$; adolescent mean 5.6, SEM 1.1, adult mean 6.1, SEM 3.4).

Memory accuracy for incidental images

There was a significant overall effect of memory accuracy (d-prime) in the combined sample (one-way t-test against 0, $t_{71} = 15.5$, $p < 0.000$) and in each group separately (adolescents, $t_{40} = 12.2$, $p < 0.000$; adults, $t_{30} = 9.5$, $p < 0.000$) showing that all participants had good memory for the incidental pictures that were presented at the time of reinforcement. As reported in the main manuscript there was a main effect of reinforcement in a repeated measures analysis of variance (RM-ANOVA), such that there was better memory for images that were presented with positive than negative reinforcement, in adolescents ($t_{40} = 4.4$, $p < 0.000$) and adults ($t_{30} = 2.4$, $p = 0.02$).

Testing the association of memory accuracy with trial-by-trial expected value for cues estimated from a reinforcement learning model

In an analysis parallel to the association with trial-by-trial prediction error and later memory accuracy, we investigated whether there was also such an association between expected value as estimated by the reinforcement learning model for cues and later memory accuracy. We found no significant associations (main effect of Expected value, $Z = 0.8$, $p = 0.4$; main effect of Group, $Z = -0.3$, $p = 0.8$; interaction, $Z = -1.4$, $p = 0.2$). This supports the interpretation that differences in value representation, on their own, do not explain the differences we observe between groups in the relationship between reinforcement learning and episodic memory.

Targeted analyses of responses in the ventromedial prefrontal cortex (vmPFC) during cue and reinforcement

Given the known role of vmPFC in value and reinforcement learning (Bartra et al., 2013; Clithero and Rangel, 2014), we explored whether there were differences in value representation in adolescents and adults that might explain differences in learning. We used a general linear model to detect activation at the time of reinforcement for events that were positive greater than negative. We applied a correction threshold of $Z > 2.3$ ($p < 0.01$ one-tailed) to the whole brain in a group-level general linear model, to observe average activation maps in each age group separately, and to compare activation between groups. Within the vmPFC, we found significant activation in each group (adolescents, $Z = 4.37$, $p < 0.000$, peak [-12, 44, -12]; adults, $Z = 3.77$, $p = 0.008$, peak [6, 42, -12]), but no differences between them at the time of reinforcement. We performed a similar analysis at the time of cue onset, evaluating the correlation of estimated expected value for the cue from the reinforcement learning model. This analysis again revealed no significant differences in vmPFC activation between the groups.

Supplemental Figures and Tables

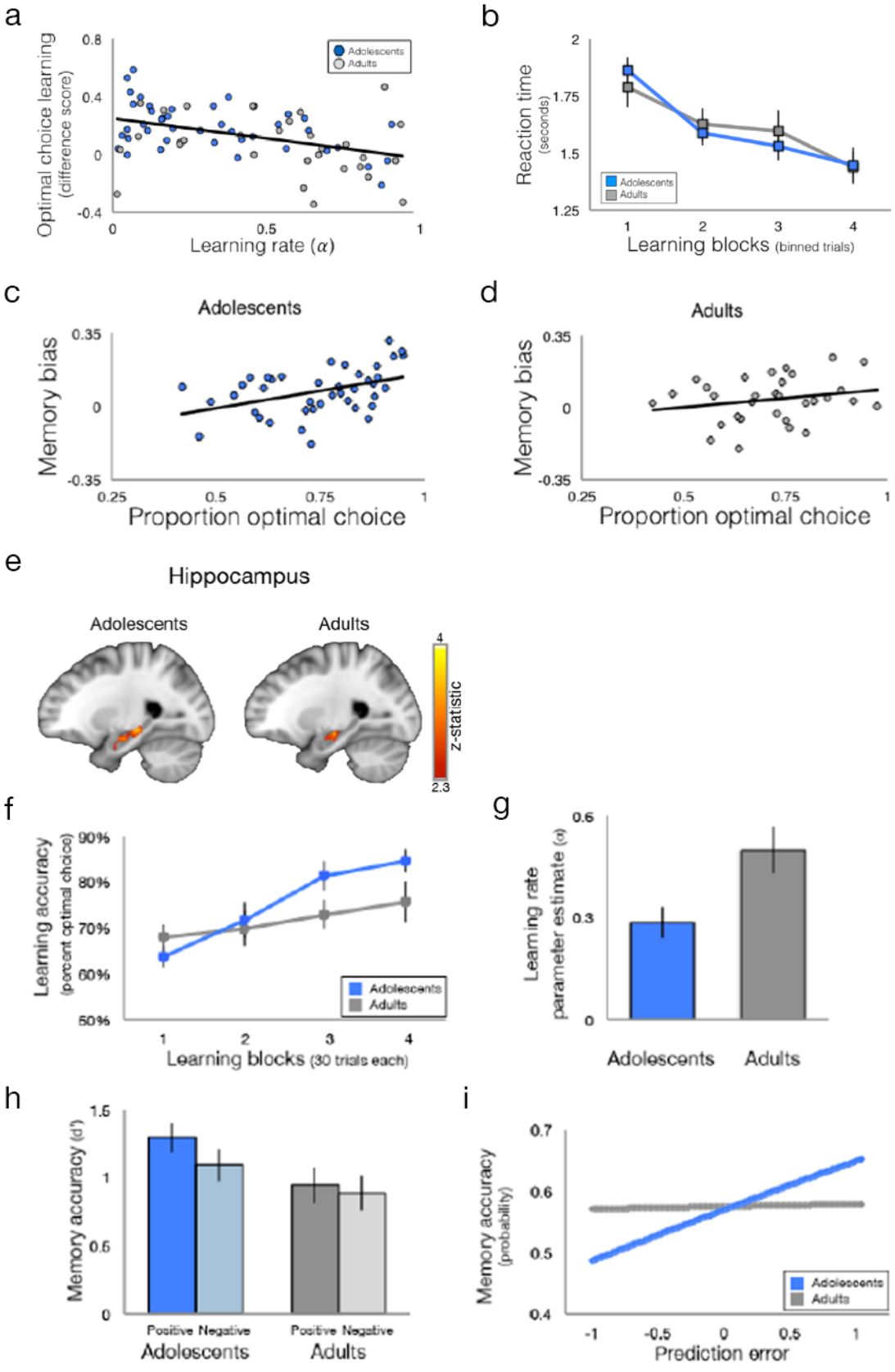

Figure S1. Related to Figure 2. (a) A negative correlation between learning rate and improvement in task performance. To investigate factors that could contribute to the differences in learning, we compared the estimated learning rate parameter (from the reinforcement learning model) and improvement in choices during the task (computed as the difference between mean optimal choice proportion in block 4 from block 1; $y = block_4 - block_1$) which reflects the increase in proportion of choosing the optimal target from the start to the end of the learning phase. Consistent with the idea that a lower learning rate is better in this sort of task, we found a significant correlation for the whole sample, $r = -0.43$, $p = 0.0002$; groups are plotted in blue and grey respectively for visualization. **(b) No difference in reaction time between the groups.** Latency to respond from onset of trial, until target selection. Both groups show a speeding of responses over the course of learning, (RM-ANOVA, Block: $F_{3,210} = 26.5$, $p = 0.000$), with no differences between them (Group: $F_{1,70} = 0.001$, $p = 0.9$; Block X Group: $F_{3,210} = 1.0$, $p = 0.4$). Points reflect mean reaction time per block; bars show ± one SEM. **(c-d) Positivity bias in memory is related to learning performance in adolescents, but not adults.** To relate behavioral measures of learning and memory across participants, we generated a single numeric index of memory sensitivity to reinforcement valence reflecting each participant's "positivity bias" in memory. This was calculated by computing a difference score for each participant between the proportion of remembered images when the outcome was positive vs. negative. We correlated this index of positivity bias with learning as the averaged proportion of optimal choice over the entire learning task. Among the adolescents, the positivity bias in memory was correlated with learning, such that adolescents who learned better also had a higher positivity bias ($n = 41$, $r = 0.43$, $p = 0.006$). Adults did not show this relationship ($n = 31$, $r = 0.22$, $p = 0.2$). A comparison of Fisher-Z transformed correlation coefficients was not significant ($Z = 0.93$, $p = 0.4$). **(e) Subsequent memory analysis of fMRI results in the hippocampus.** We investigated memory related activation in the hippocampus by comparing remembered and forgotten images from the behavioral memory test completed after the scan to brain activity at the time the image was initially presented during the learning phase. Memory regressors for fMRI analysis for each participant were derived from remembered and forgotten trials. To categorize a trial as remembered, the participant must have correctly identified the image as "old" and must have a d-prime above zero for the confidence-rating that the participant provided for that image. Similarly, forgotten images were defined as items mis-identified as "new" with a positive d-prime score for the associated confidence-rating. Any trial where a participant reported that they were "Just Guessing" was excluded. We examined BOLD correlates for remembered greater than forgotten pictures in each group separately and then directly compared between the adolescent and adult groups. These analyses were restricted to the hippocampus, our *a priori* subsequent memory ROI (see **Supplemental Table S1**). Within the adolescents this analysis revealed one significant cluster in the left hippocampus (Family-Wise Error (FWE)-p cluster < 0.001, $Z = 3.71$, 261 voxels, peak [-16, -30, -10]). For the adults, there was activation in bilateral hippocampus (Right: FWE-p cluster < 0.02, $Z = 3.63$, 96 voxels, peak [18, -14, -18]. Left: FWE-p cluster < 0.03, $Z = 3.38$, 93 voxels, peak [-22, -18, -12]), however in a direct contrast, there was no above threshold activation for either group over the other. Whole brain analysis results are in **Supplemental Table S1**. **(f-i) Behavioral results in fMRI subsample.** Behavioral results from the subset of participants who completed the probabilistic learning task while undergoing fMRI (n=25 adolescents; n=22 adults). Key behavioral findings within this sample replicate the results in the full behavioral sample reported in the main manuscript. Namely, the adolescents show (1) significantly better reinforcement learning than adults, (2) significantly lower learning rates than adults, (3) a significantly stronger association between prediction error learning signals and episodic memory for outcome events (within participants) and (4) a significant correlation between reinforcement learning and a "positivity bias" in episodic memory (within participants). **(f) Overall learning.** A RM-ANOVA (Block X Group) on the percent of trials for which participants responded with the outcome most often associated with that cue revealed that both age groups showed significant learning over time but, consistent with our prediction, adolescents' learning exceeded that of adults (main effect of Block: $F_{3,138} = 20.95$, $p < 0.0001$; Block X Group interaction $F_{3,138} = 4.49$, $p = 0.005$). Similar results were found for optimal choice by trial, showing learning for the full sample (mixed-effect regression main effect: $Z = 2.49$, $p = 0.013$), but better learning among the adolescent group (mixed-effect regression interaction: $Z = 2.65$, $p = 0.008$) with a significantly better fit of the group interaction model ($\chi^2 = 6.42$, $p = 0.01$) after penalizing model complexity using Akaike's Information Criterion (AIC). **(g) Learning rate parameter estimate from a reinforcement learning model.** Consistent with their overall better learning performance, adolescents had a lower learning rate parameter than adults ($t_{46} = 2.68$, $p = 0.01$), indicating more incremental learning. Importantly, the model provided a good fit to the observed behavior across both groups (one-way t-test comparing a null model, $t_{47} = -35.4$, $p < 0.0001$) and the model fits did not differ between them (independent samples t-test, $t_{46} = 0.77$, $p = 0.45$). **(h) Memory accuracy by reinforcement.** We compared memory for the trial-unique objects that were presented during learning, separating trials into those with positive ("correct") versus negative ("incorrect") outcomes. Adolescents had better

memory for positive than negative reinforcement events (two- tailed paired-samples test $t_{25} = 3.56$, $p = 0.002$). The adults showed significant memory overall (one-way t-test, $n = 22$, $t_{21} = 8.1$, $p = 0.0000001$; 1.1 ± 0.14), but no difference between positive and negative reinforcement ($t_{21} = 1.04$, $p = 0.3$). **(i) Prediction error and predicted probability of memory accuracy.** The level of prediction error evoked by the reinforcement on a particular trial was related to the probability of later memory accuracy, with an interaction effect for group (mixed-effect regression interaction: $Z = 2.78$, $p = 0.006$). This relationship between prediction error and memory accuracy was driven by the adolescents ($Z = 4.2$,
$p = 0.00003$), and no such relationship was found for the adults ($Z = 0.2$, $p = 0.8$).

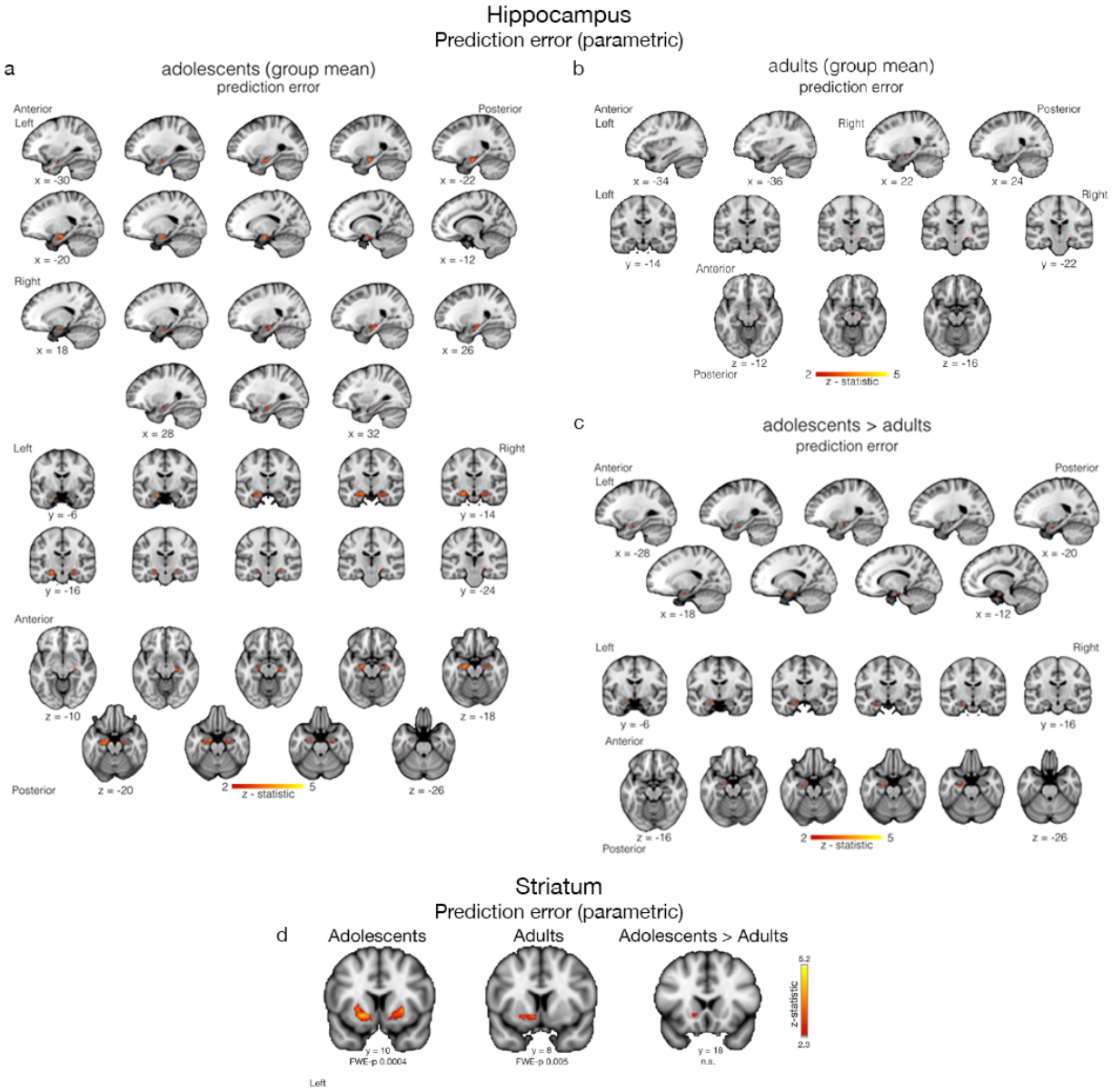

Figure S2. Related to Figure 3. (a-c) Full visualization of the extent of BOLD activity in the hippocampus correlating with prediction error. In the main manuscript (Figure 3) only one coronal plane is shown, and at the same location for 3 different maps that does not demonstrate the extent or peak of activity, so for purposes of visualization additional slices in all planes are presented. **(a) Adolescent group mean.** The adolescent group showed significant bilateral activation in the hippocampus correlated with prediction error. **(b) Adult group mean.** The adults showed bilateral activation in the hippocampus, but it did not surpass threshold for correction, (largest cluster $p > 0.1$, 13 voxels, peak [22, -20, -12], Z at peak = 2.5). **(c) Direct comparison of adolescent group greater**

than adult group. The adolescents had significantly greater activation in the left hippocampus than adults, and there were no significant findings for adults greater than adolescents. **(d) No significant differences between the groups in prediction error related activation in the striatum.** Prediction error was regressed as a parametric modulator with BOLD activity and examined within an *a priori* region of interest in the striatum (combined bilateral caudate, putamen, and accumbens, anatomically defined by the Harvard-Oxford probability atlas at a 75% probability threshold). Maps depict peak of presented contrast; see **Supplemental Table S2** for reports of local maxima. Contrasts within-group were thresholded at Z>2.3 one-tailed in FEAT. For between-group contrast correction was carried out with FSLs Cluster tool thresholded at Z > 2.3. Cluster extent correction thresholds were estimated by AFNIs 3dClustSim by providing the mask and smoothness within the mask, as calculated from the residuals using AFNIs 3dFWHMx (http://afni.nimh.nih.gov/). For adolescents, there was significant activation in the bilateral ventral striatum. For adults there was significant activation in the left ventral striatum. A comparison between groups did not meet correction threshold in either direction (depicted: adolescents > adults; largest cluster $p > 0.1$, 50 voxels, peak voxel [-14, 18, -4], Z at peak = 2.73). Whole brain results are in **Supplemental Table S2**. Contrasts within- and between-groups for whole brain analysis were thresholded at Z > 2.3 or FDR$p < 0.01$ one-tailed in FEAT. Whole brain contrasts between-groups yielded no significant differences at this threshold.

Table S1. Related to Figure 2c. Memory fMRI results (remembered > forgotten), peak and top 4 local maxima

Whole brain							
Anatomical Label (H-O atlas)	**Voxels**	**P**	**Z-score**	**MNI Coordinate**			
				X	**Y**	**Z**	
***adults**, p-voxel < 0.01*							
L. temporal occipital fusiform cortex	2167	7.16e-11	4.04	-34	-58	-8	*cluster peak*
L. middle temporal gyrus			3.94	-66	-56	6	*local max*
			3.92	-54	-50	-8	*local max*
L. lateral occipital cortex			3.92	-36	-86	24	*local max*
L. inferior temporal gyrus			3.87	-40	-58	-6	*local max*
R. lateral occipital cortex	1044	3.76e-6	4.55	46	-62	-8	*cluster peak*
			3.53	42	-72	-8	*local max*
R. temporal occipital fusiform cortex			4.45	42	-48	-14	*local max*
			3.98	38	-52	-10	*local max*
R. inferior temporal gyrus			3.72	46	-58	-12	*local max*
L. frontal pole	683	0.000268	3.95	-8	66	28	*cluster peak*
			3.58	-8	62	2	*local max*
			3.56	-2	64	16	*local max*
R. superior frontal gyrus			3.55	8	56	18	*local max*
			3.54	-8	56	22	*local max*
L. frontal orbital cortex	437	0.00769	4.45	-36	32	-14	*cluster peak*
			3.52	-44	34	-6	*local max*
			3.39	-50	28	-10	*local max*
			3.02	-28	28	-20	*local max*
L. frontal pole			2.82	-52	40	-10	*local max*
L. amygdala	421	0.00974	4.23	-20	-6	-18	*cluster peak*
L. hippocampus			3.38	-22	-18	-12	*local max*
L. frontal orbital cortex			3.35	-28	8	-16	*local max*
L. parahippocampal gyrus			3.08	-16	-22	-24	*local max*
L. insular cortex			2.54	-34	8	-8	*local max*
R. postcentral gyrus	395	0.0144	3.87	36	-24	46	*cluster peak*
			3.26	36	-26	52	*local max*
R. precentral gyrus			3.65	40	-20	52	*local max*
			3.38	38	-20	66	*local max*
			3.25	38	-6	60	*local max*
R. amygdala	325	0.0429	3.67	12	-2	-20	*cluster peak*
			3.28	20	-6	-14	*local max*
			3.18	8	0	-12	*local max*
R. hippocampus			3.63	18	-14	-18	*local max*
			3.58	22	-14	-16	*local max*
***adolescents**, p-voxel < 0.01*							
L. occipital pole	8867	5.99E-30	6.96	-32	-92	4	*cluster peak*
L. lateral occipital cortex			6.65	-42	-72	-12	*local max*
L. occipital fusiform gyrus			6.16	-42	-76	-4	*local max*
L. inferior temporal gyrus			6.3	-36	-78	-12	*local max*
			6.26	-44	-62	-8	*local max*
R. lateral occipital cortex	8229	1.99E-28	6.72	36	-88	-6	*cluster peak*
			6.47	36	-88	14	*local max*

			6.39	46	-64	-8	*local max*
			6.08	40	-64	-4	*local max*
R. occipital pole			5.87	16	-94	-2	*local max*
L. frontal operculum cortex	722	0.000164	4.15	-46	26	-2	*cluster peak*
L. frontal orbital cortex			4.06	-38	32	-10	*local max*
			3.56	-24	28	-12	*local max*
L. inferior frontal gyrus			3.81	-42	34	4	*local max*
			3.54	-46	28	14	*local max*
***adolescents < adults**, p-voxel < 0.01*							
R. precuneous cortex	694	0.000233	4.37	4	-46	40	*cluster peak*
			3.46	0	-62	32	*local max*
			3.43	2	-58	32	*local max*
L. cingulate gyrus			3.55	-4	-40	44	*local max*
L. precuneous cortex			3.45	-4	-60	38	*local max*
***adults < adolescents**, p-voxel < 0.01*							
R. lateral occipital cortex	3000	7.77E-14	4.96	36	-90	-6	*cluster peak*
			4.43	26	-90	-6	*local max*
			4.34	28	-82	18	*local max*
			4.28	34	-88	-12	*local max*
R. occipital fusiform gyrus			4.29	26	-84	-12	*local max*
L. occipital fusiform gyrus	2128	1.01E-10	4.44	-28	-84	-14	*cluster peak*
			4.27	-38	-76	-14	*local max*
			4.03	-40	-72	-12	*local max*
L. occipital pole			4.02	-26	-92	12	*local max*
L. temporal occipital fusiform cortex			3.65	-44	-52	-24	*local max*

ROI (bilateral hippocampus)

Anatomical Label (H-O atlas)	**Voxels**	**P**	**Z-score**	**MNI Coordinate**			
				X	**Y**	**Z**	
***adults**, p-voxel < 0.01*							
R. hippocampus	96	0.0239	3.63	18	-14	-18	*cluster peak*
			3.58	22	-14	-16	*local max*
			2.59	16	-8	-20	*local max*
L. hippocampus	93	0.0257	3.38	-22	-18	-12	*cluster peak*
			3.32	-18	-12	-16	*local max*
***adolescents**, p-voxel < 0.01*							
L. hippocampus	261	0.000852	3.71	-26	-30	-10	*cluster peak*
			3.69	-34	-30	-14	*local max*
			3.65	-22	-26	-12	*local max*
			3.17	-22	-10	-18	*local max*
			3.1	-24	-16	-18	*local max*
			2.95	-16	-8	-20	*local max*
***adolescents > adults**, and **adults > adolescents**, p-voxel < 0.01*							
no activation observed above threshold							

Table S2. Related to Figure 3. Prediction error fMRI results, peak and top 4 local maxima

Whole brain							
Anatomical Label (H-O atlas)	**Voxels**	**p-cluster**	**Z-score**	**MNI Coordinate**			
				X	**Y**	**Z**	
***adults**, p-voxel < 0.01*							
R. occipital pole	5439	1.43E-19	4.83	14	-94	22	*cluster peak*
L. occipital pole			4.56	-30	-94	-10	*local max*
			4.46	-34	-92	-10	*local max*
			4.35	-12	-94	-14	*local max*
			4.34	-4	-92	-8	*local max*
R. cerebellum	372	0.035	4.26	38	-70	-38	*cluster peak*
			3.93	40	-66	-38	*local max*
			3.87	36	-60	-42	*local max*
			2.88	28	-46	-42	*local max*
			2.5	38	-78	-44	*local max*
***adolescents**, p-voxel < 0.01*							
R. lateral occipital cortex	3377	6.95E-14	5.1	36	-84	4	*cluster peak*
			4.62	32	-90	12	*local max*
			4.6	34	-86	16	*local max*
			4.3	26	-84	12	*local max*
			4.05	52	-68	-8	*local max*
L. inferior temporal gyrus	3226	1.99E-13	4.83	-54	-62	-12	*cluster peak*
			4.62	-52	-52	-12	*local max*
L. lateral occipital cortex			4.54	-34	-90	4	*local max*
			4.48	-30	-88	-2	*local max*
L. middle temporal gyrus			4.49	-62	-48	-8	*local max*
L. amygdala	988	2.22E-05	4.65	-18	-8	-14	*cluster peak*
L. putamen			4.65	-18	10	-6	*local max*
			3.96	-26	0	-8	*local max*
L. accumbens (ventral striatum)			4.55	-12	6	-8	*local max*
L. hippocampus			3.64	-24	-14	-18	*local max*
L. frontal pole	785	0.000199	4.13	-48	44	6	*cluster peak*
			4.08	-48	40	12	*local max*
			3.94	-52	42	6	*local max*
			3.81	-42	36	14	*local max*
L. inferior frontal gyrus			3.62	-56	26	4	*local max*
R. precentral gyrus	578	0.00228	4.03	24	-26	72	*cluster peak*
			3.82	2	-28	64	*local max*
			3.73	2	-26	58	*local max*
			3.56	8	-26	74	*local max*
R. postcentral gyrus			3.43	26	-34	64	*local max*
adolescents* > *adults, *and* ***adults* > *adolescents***, *p-voxel < 0.01*							
no activation observed above threshold							

ROI (Striatum)							
Anatomical Label (H-O atlas)	**Voxels**	**p-cluster**	**Z-score**	**MNI Coordinate**			
				X	**Y**	**Z**	
***adults**, p-voxel < 0.01*							
L. accumbens (ventral striatum)	222	0.00469	3.59	-8	8	-8	*cluster peak*
L. putamen			3.21	-28	0	6	*local max*
			3.15	-22	6	-8	*local max*
			2.75	-28	-12	6	*local max*
			2.74	-26	-6	-6	*local max*
			2.42	-26	-2	-8	*local max*
***adolescents**, p-voxel < 0.01*							
L. putamen	388	0.000359	4.65	-18	10	-6	*cluster peak*
L. accumbens (ventral striatum)			4.55	-12	6	-8	*local max*
L. putamen			3.96	-26	0	-8	*local max*
			3.82	-24	8	0	*local max*
			3.35	-28	-14	6	*local max*
			3.27	-28	-8	-4	*local max*
R. putamen	209	0.00587	4.03	26	10	0	*cluster peak*
			3.63	16	6	-10	*local max*
adolescents* > *adults, *and* ***adults* > *adolescents***, *p-voxel < 0.01*							
no activation observed above threshold							

Supplemental Experimental Procedures

Forty-three (43) adolescents and 34 adults participated in the study. Among them, 28 adolescents and 25 adults were scanned with fMRI. Three adolescents were excluded from various analyses: two were excluded from all analyses, due to technical issues; one was excluded from brain imaging analyses due to excessive motion (absolute motion > 2.4 mm). Three adults were excluded from the final sample, two for technical issues, and one for an incidental neurological finding. This resulted in a total final sample of 41 adolescents (ages 13-17, mean age 15.9, SD = 1.4, 14 females) and 31 adults (ages 20-30, mean age 26.6, SD = 3.0, 18 females) for behavioral analyses and a subset of 25 adolescents (ages 13-17, mean age 15.9 years, SD = 1.4, 10 females) and 22 adults (ages 24-30, mean age 27.7 years, SD = 2.0, 13 females) for fMRI analyses.

All adult participants provided informed consent to participate in the study. All adolescent participants provided informed assent and had a parent provide informed consent and permission to participate. All participants were initially screened over the phone (for adolescent participants screening was completed with the adolescent and a parent). All participants were healthy, native-English speakers, with normal color vision, reported no history of psychiatric or neurological disorders, were not taking medication, and fMRI participants were all right-handed and had no contraindications for scanning. At the completion of study, participants were paid for their time; no additional monetary incentives were used as motivation in the task.

To estimate IQ, we administered the Wechsler Abbreviated Scale of Intelligence 2-part sub test, with adolescents scoring an average of 107 (SD = 10.4, range 86 to 126, n = 27) and adults scoring an average of 115 (SD = 9.5, range 99 to 138, n = 21). The subsample that completed the fMRI was similar, with adolescents scoring an average of 104 (SD = 10.3, range 90 to 126, n = 18) and adults scoring an average of 115 (SD = 10.6, range 99 to 138, n = 14). Adolescents reported estimated pubertal development using a self-report multiple choice questionnaire (Petersen et al., 1988), with a mode of stage 3, ranging from stages 2 – 5, reflecting that pubertal development in the sample was well underway but not complete. Providing racial and ethnic demographic information was optional; the fMRI sample was comprised of 24% Hispanic, 9% Asian, 19% African-American, 63% Caucasian, and 9% mixed-ethnicity participants recruited from the community.

Behavioral Task Procedures

Pre-task practice. Prior to performing the probabilistic learning task, adolescent and adult participants read instructions on a laptop and completed a practice round of the task (8 trials, different stimuli than used in the task) to become familiar with the general appearance of the task and to be certain that the instructions were correctly understood. The practice task was presented in MatLab (http://www.mathworks.com/) using the Psychophysics Toolbox (Brainard, 1997).

Learning task. Participants performed the probabilistic learning task (Foerde and Shohamy, 2011) on a computer or while undergoing fMRI. The task was presented in MatLab (http://www.mathworks.com/) using the Psychophysics Toolbox (Brainard, 1997) and presentation timing for fMRI of events and jitter durations were optimized for rapid event-related fMRI using OptSeq (Dale, 1999).

On each trial in the learning phase, participants saw one of four cues (butterfly; blue, purple, red, yellow) and had to predict which of two targets (flowers; pink, white) the butterfly was more likely to feed from. Participants had up to four seconds to make a response and were encouraged to respond as quickly as possible. Each cue was associated with one target on 80% of trials and with the other target on 20% of trials (**Figure 1A**). Thus most of the time choosing the optimal target for a cue results in "correct" reinforcement, but the other 20% of the time results in "incorrect" reinforcement, allowing the observation of learning rates over the course of the task as well as to estimate trial-by-trial expectations and prediction errors.

Participants pressed a button to choose either the left- or the right-sided target and then received reinforcement on their choice. Reinforcement was visually presented on the screen for two seconds, followed by a fixation-cross for a jittered inter-trial interval. Within participant, each target always appeared in the same location (i.e. fixed left and right) and was counterbalanced across participants. Within participant, the cue and target association was fixed over the entire task; the combinations were fully counterbalanced across participants. Learning behavior was analyzed using summary measures as well as trial-by-trial with reinforcement learning models and linear regression. For trial-by-trial regression, we fit a mixed effects linear model to each participant's optimal choice behavior using the *lmer*

package in R (Bates et al., 2015) associating accuracy with Trial X Group modeled as a fixed effect and Participants modeled as a random effect.

If a participant did not respond in time, the words "too late" were presented for the duration of the trial to preserve timing. No reinforcement was presented for such trials and thus there was no presentation of an “episodic” picture. These learning trials were discarded from behavioral analysis and modeled as a regressor of non-interest to be kept out of baseline for fMRI analysis. Such, invalid trials in the learning phase resulted in a different number of test items for the surprise memory test. There were very few of these invalid trials across all participants. For adults, the average response rate was 98% ± 0.03 SEM, with a minimum of 86%. For adolescents, the average response rate was 98% ± 0.03 SEM, with a minimum of 88%.

Post-learning test. Immediately following the learning phase was a test phase (1 fMRI run, 32 trials; behavior only sample - 1 run, 24 trials), where participants continued to make choices for the same cue-target associations, but no longer received reinforcement. Participants were instructed to continue choosing based on the associations that they had learned over the previous trials. This provided a measure for how well the associations were learned for each of the cues, in the absence of continued reinforcement.

Subsequent memory test. As described in the main text, after the learning task, participants were given a surprise test of recognition memory for the outcome images from the learning task. The pictures were orthogonal to the learning task in that they provided no information for learning the cue-target associations. Memory testing was completed on a computer, 30 minutes after the completion of the learning phase. For participants who did not undergo fMRI, the memory test was completed 10 minutes after the completion of the learning phase. Participants saw each picture that had appeared in the learning phase ("old" pictures) intermixed with an equal number of novel pictures ("new" pictures) that had not been seen in the learning phase. The order of the presentation of old and new pictures was random and unique to each participant. Confidence ratings provided an index of how sure participants felt about their memory judgments. Confidence ratings allowed the exclusion of “guess” trials from all behavioral analyses, and were used in generating behavioral memory regressors for fMRI analysis (see **Supplemental Figure S1e**).

We calculated an average memory accuracy score for each participant, excluding all memory test trials where participants indicated they were guessing. Memory accuracy was quantified by d-prime, both overall and separately for events where reinforcement was positive vs. negative. D-prime is a signal detection index and a more sensitive measure of accuracy because it takes into account the bias of any individual towards identifying items as old. We computed d-prime using normalized rates of correctly remembered trials (hits) and trials where participants mistakenly identified novel pictures as old (false alarms). For computation of d-prime for positive and negative reinforcement, normalized rate of false alarms was subtracted from normalized rates of hits when presented reinforcement was "correct" and "incorrect" respectively.

For correlating memory with functional connectivity during learning, we computed an index of valence bias in memory, by taking the difference score for each participant between the proportion of remembered images when the presented reinforcement outcome was positive vs. negative. Individuals who remembered many more images when the reinforcement was "correct" would have a positive difference score reflecting a "positivity bias" in memory, whereas those who remembered a similar proportion of images for both positive and negative reinforcements would have a value close to zero, reflecting little bias in memory.

Reinforcement Learning Model Analysis

To explore how reinforcement drives trial-by-trial updating of choices during learning, we fit a standard reinforcement learning model with two free parameters to each participant's observed choices in the learning phase (Daw, 2011; Sutton and Barto, 1998).

Limiting the prior probability distribution is appropriate for constraining possible solutions for estimated parameters, particularly when comparisons between groups is of interest (Daw, 2011). Additionally, a method for reducing noise in fitting parameters from individual participants comes from repeated estimations of LLE fits over many different initialization points for the minimization optimizer, and then checking for the smallest LLE and keeping the estimated parameters associated with that fit (Daw, 2011). This prevents solutions that fall into local minima. To address these pitfalls in reinforcement learning methods, we fit a model with a prior probability distribution limiting estimated parameters, α and β. Estimates for each free parameter were constrained with a probability density

function [β: *Gamma* (1.2, 5), α: *Beta* (1.1, 1.1)] (Daw et al., 2011). Priors were used to penalize the estimates output from the minimization optimizer. The minimization optimizer function was initialized at 20 random start points and then run 2,000 times, for 40,000 iterations per participant. The best solution, and its associated parameter estimates, among the iterations was selected. We also evaluated a model without the constraint of a prior on the free parameters (see **Supplemental Results**).

In **Equations 1** and **2**, Q is the estimated value, such that Q_t is the expected value on trial t and Q_{t+1} is the updated expected value on the trial following trial t. The difference between the expected outcome Q and the reinforcement received r on trial t, is the measure of prediction error (δ).

$$Q_{t+1} = Q_t + \alpha * \delta \qquad \textit{Equation 1}$$
$$\text{where } \delta = (r_t - Q_t)$$

One estimated parameter is a learning rate (α), which is a measure of the extent to which reinforcement on each trial is used to update choices. Values for α can range from 0 to 1. High learning rates closer to 1 indicate heavily weighing recent reinforcement for more rapid updating based on fewer trials; in a context where outcomes are probabilistic but probabilities are consistent, this can result in overweighting recent but rare reinforcement, and frequently shifting choice behavior (Sutton and Barto, 1998). A learning rate closer to 0 indicates more incremental learning in which changes in value accrue over a greater number of trials; this strategy allows for learning over several trials, such that surprising but rare reinforcement outcomes will not dramatically shift choice behavior (Sutton and Barto, 1998). In a probabilistic task like the one used here, a lower learning rate parameter is generally better, as it suggests that learning is guided by accumulating evidence over a greater number of trials rather than shifting behavior based on the outcome of any single trial. Another estimated parameter is the inverse temperature parameter (β) and follows a softmax distribution. This parameter can be thought of as a link that relates the value of an option and the actual choosing of that option (Daw, 2011). It is an index of noise in choice behavior (the extent to which a participant exploits the learned value vs. explores other options).

For each participant we also calculated the inverse log likelihood estimate (LLE; **Equation 2**) to describe how well the model fits each participant's observed choices C, using a minimization optimizer (fmincon in MatLab,

http://www.mathworks.com/) to find the global minimum for this test of model fit. The value term here is computed trial-by-trial by taking the difference of the expected value on trial *t*, Q_t, and the mean expected value for that stimulus from all trials Q_m. We used Akaike's Information Criterion (AIC; Akaike, 1974) to penalize the complexity of the model where *k* is the number of estimated parameters, and computed the chance level pseudo-*R* for two choices over the total number of trials *T* (Daw, 2011) and subtracted this from the AIC term for each participant.

$$LLE = LLE + \beta * Q_{t-m}(C_t) - \log \sum \exp(\beta * Q_{t-m}) \quad \textit{Equation 2}$$
where $Q_{t-m} = Q_t - Q_m$

$$AIC = (-2*(LLE))+(2*k) \quad \textit{Equation 3}$$

$$R = -T*LN(0.5) \quad \textit{Equation 4}$$

Parameters α and β were estimated for each participant as described above, and then respective group parameter averages were used for estimating prediction errors for individuals within the groups for behavioral and fMRI analyses (Daw, 2011).

Prediction errors were calculated by subtracting the estimated subjective value for a cue on each trial from the reinforcement received on that trial (i.e. 1 for "correct" reinforcement and 0 for "incorrect" reinforcement). Thus, if actual received outcome is larger than the expected outcome this results in a positive prediction error, and conversely received outcomes smaller than the expected outcome result in a negative prediction error. The more unexpected the reinforcement, the larger the prediction error estimate in its respective direction. Following previous work with a variant of the task (Foerde and Shohamy, 2011) we included both positive and negative prediction errors in the same parametric regressor for fMRI analyses. For trial-by-trial regression associating prediction error magnitude with later memory accuracy, we fit a mixed effects linear model to each participant's memory accuracy with Prediction error X Group as a fixed effect and Participants modeled as a random effect.

FMRI Data Acquisition and Preprocessing

Scanning was performed on a 3-Tesla Siemens Trio MRI scanner with a 12-channel head coil at the Ahmanson-Lovelace Brain Mapping Center at the University of California – Los Angeles. For each functional run during the learning phase of the behavioral task, after discarding 4 initial TRs (8 seconds) as the scanner stabilized, 200

volumes of T2*-weighted interleaved echo-planar (EPI) images were acquired in the transverse plane (slice thickness = 4 mm, 34 slices, voxel size = 3x3x4 mm^3, TR = 2000 ms, TE = 30 ms, FA = 90º, matrix = 64 x 64, FOV = 192 mm) for a total of 800 volumes. For the test phase all parameters were the same, except 175 volumes were collected. A high-resolution, T1-weighted magnetization-prepared rapid-acquisition gradient echo (MPRAGE) anatomical scan was acquired in the sagittal plane for each participant at the end of functional scanning (slice thickness = 1mm, voxel size = 1x1x1 mm^3, 160 slices, TR = 2170 ms, TE = 4.33 ms, FA = 7º, matrix = 256 x 256, FOV = 256 mm). Functional images were motion corrected, slice-time corrected using Fourier-space time-series phase-shifting, skull stripped using the FMRIB Software Library's (FSL) Brain Extraction Tool (Smith, 2002), spatially smoothed with a 5mm full width at half maximum (FWHM) Gaussian kernel, grand-mean intensity normalized, highpass filtered at 0.02 Hz, and registered to standard Montreal Neurological Institute (MNI) template. High-resolution MPRAGE images were skull-stripped with BET, segmented into probability maps for gray matter, white matter and CSF, and linearly registered to standard Montreal Neurological Institute (MNI) template with 12 degrees of freedom. All preprocessed images were visually inspected.

Motion

Functional images were motion corrected using MCFLIRT (Jenkinson et al., 2002) deriving 24 nuisance regressors related to motion (6-rigid body transformations – 3-translational, 3-rotational – their derivatives, their squares, and the square of the derivatives). One adolescent participant was excluded from all brain analyses for having motion greater than 2.5 standard deviations away from the whole sample (adolescents and adults) average motion. After removing this participant, the relative motion (t_{45} = -0.11, p = 0.9) and absolute motion (t_{45} = 0.29, p = 0.8) did not differ between adolescents (n = 25, mean absolute motion 0.25 ± 0.027 SEM, mean relative motion 0.07 ± 0.006 SEM) and adults (n = 22, mean absolute motion 0.24 ± 0.019 SEM, mean relative motion 0.07 ± 0.004 SEM).

FMRI Statistical Analysis

Imaging data were processed and analyzed using the FMRI Expert Analysis Tool (FEAT) from FSL v6.00 toolbox (www.fmrib.ox.ac.uk/fsl). Subject-level and group-level general linear models were analyzed in FEAT, using a general linear model (GLM) in FLAME1. For detecting activations correlated with prediction errors, the model included the onset of each reinforcement event weighted by its calculated parametric prediction error value,

orthogonalized to the reinforcement event. Onset of the choice for each trial and missed responses were also included in the GLM as regressors of non-interest. We used a within participant fixed-effects analysis to average regressed activation over the 4 functional runs (http://fsl.fmrib.ox.ac.uk/fsl/fslwiki/FEAT). Outputs from participant-level analyses were then analyzed in a group-level general linear model, to observe average activation maps in each age group separately, and to compare activation between groups.

Small Volume Correction

Given our hypotheses, centered on our *a priori* regions of interest in the striatum and hippocampus, group level whole brain results were masked for small volume correction. Correction was carried out in FEAT as part of post-stats, so that the mask was applied at a threshold of $Z > 2.3$ ($p < 0.01$ one-tailed) before cluster extraction for group means. The striatum and hippocampus ROIs were anatomically defined using the Harvard-Oxford probability atlas, thresholded at 75% probability. The striatum ROI was comprised of bilateral caudate, putamen, and nucleus accumbens (Desikan et al., 2006). The hippocampus ROI included the entire head, body, and tail, and visual inspection of the hippocampal ROI confirmed that the head was posterior to the amygdala (see **Supplemental Figure S2a-c**). Small volume correction for between-group contrasts was carried out by identifying clusters within masked regions using FSLs Cluster tool thresholded at $Z > 1.65$ ($p < 0.05$ one-tailed), and cluster extent correction thresholds were estimated by AFNIs 3dClustSim by providing the mask and smoothness as calculated from the masked residuals using AFNIs 3dFWHMx (http://afni.nimh.nih.gov/).

Functional Connectivity Analysis

To examine functional connectivity between the striatum and the hippocampus, the two *a priori* regions of interests where we found prediction error related activation (see **Figure 3** and **Supplemental Figure S2**), we conducted a psycho-physiological interaction (PPI) analysis (O'Reilly et al., 2012). For the physiological regressor, we extracted timeseries in a region of the anterior hippocampus that consisted of a 3mm sphere drawn around the peak voxel of the adolescent group's mean response to prediction error. To be certain that this seed drawn around the peak voxel did not contain activation from regions outside of the hippocampus (e.g. the amygdala), we masked the sphere with the same anatomical ROI of hippocampus, described above. For the psychological regressor we used the contrast of correct > incorrect presented reinforcement. We included all other trial events in the GLM as regressors of non-

interest, as well as the physiological and psychological regressors, in addition to the interaction term of interest. The resulting map of activation of the interaction between the physiological and psychological regressors was masked with the anatomical ROI of the bilateral striatum, described above.

Supplemental References

Bates, D., Mächler, M., Bolker, B., Walker, S., 2015. Fitting linear mixed-effects models using lme4. J. Stat. Softw. 67, 1–48. doi:10.18637/jss.v067.i01

Brainard, D.H., 1997. The Psychophysics Toolbox. Spat. Vis. 10, 433–436.

Dale, A.M., 1999. Optimal experimental design for event-related fMRI. Hum. Brain Mapp. 8, 109–114. doi:10.1002/(SICI)1097-0193(1999)8:2/3<109::AID-HBM7>3.0.CO;2-W

Desikan, R.S., Ségonne, F., Fischl, B., Quinn, B.T., Dickerson, B.C., Blacker, D., Buckner, R.L., Dale, A.M., Maguire, R.P., Hyman, B.T., Albert, M.S., Killiany, R.J., 2006. An automated labeling system for subdividing the human cerebral cortex on MRI scans into gyral based regions of interest. NeuroImage 31, 968–980. doi:10.1016/j.neuroimage.2006.01.021

Foerde, K., Shohamy, D., 2011. Feedback timing modulates brain systems for learning in humans. J. Neurosci. 31, 13157–13167. doi:10.1523/JNEUROSCI.2701-11.2011

Jenkinson, M., Bannister, P., Brady, M., Smith, S., 2002. Improved Optimization for the Robust and Accurate Linear Registration and Motion Correction of Brain Images. NeuroImage 17, 825–841.

O'Reilly, J.X., Woolrich, M.W., Behrens, T.E.J., Smith, S.M., Johansen-Berg, H., 2012. Tools of the trade: psychophysiological interactions and functional connectivity. Soc. Cogn. Affect. Neurosci. 7, 604–609. doi:10.1093/scan/nss055

Petersen, A.C., Crockett, L., Richards, M., Boxer, A., 1988. A self-report measure of pubertal status: Reliability, validity, and initial norms. J. Youth Adolesc. 17, 117–133. doi:10.1007/BF01537962

Smith, S.M., 2002. Fast robust automated brain extraction. Hum. Brain Mapp. 17, 143–155.

Sutton, R.S., Barto, A.G., 1998. Reinforcement Learning : An Introduction, Adaptive Computation and Machine Learning. MIT Press, Cambridge, Mass.