Longitudinal Changes in Value-based Learning in Middle Childhood: Distinct Contributions of Hippocampus and Striatum

Johannes Falck[1], Lei Zhang[2,3,4], Laurel Raffington[5], Johannes J. Mohn[6], Jochen Triesch[7], Christine Heim[6,8] & Yee Lee Shing[1]

[1]*Department of Psychology, Goethe University Frankfurt, 60629 Frankfurt am Main, Germany*

[2]*Social, Cognitive and Affective Neuroscience Unit, Department of Cognition, Emotion, and Methods in Psychology, Faculty of Psychology, University of Vienna, 1010 Vienna, Austria*

[3]*Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK*

[4]*Institute for Mental Health, School of Psychology, University of Birmingham, Birmingham B15 2TT, UK*

[5]*Center for Lifespan Psychology, Max Planck Institute for Human Development, 14195 Berlin, Germany*

[6]*Charité – Universitätsmedizin Berlin, Institute of Medical Psychology, 10117 Berlin, Germany*

[7]*Frankfurt Institute for Advanced Studies (FIAS), 60439 Frankfurt am Main, Germany*

[8]*Center for Safe & Healthy Children, The Pennsylvania State University, State College, PA 16802, USA*

Abstract

The hippocampal-dependent memory system and striatal-dependent memory system modulate reinforcement learning depending on feedback timing in adults, but their contributions during development remain unclear. In a 2-year longitudinal study, 6-to-7-year-old children performed a reinforcement learning task in which they received feedback immediately or with a short delay following their response. Children's learning was found to be sensitive to feedback timing modulations in their reaction time and inverse temperature parameter, which quantifies value-guided decision-making. They showed longitudinal improvements towards more optimal value-based learning, and their hippocampal volume showed protracted maturation. Better delayed model-derived learning covaried with larger hippocampal volume longitudinally, in line with the adult literature. In contrast, a larger striatal volume in children was associated with both better immediate and delayed model-derived learning longitudinally. These findings show, for the first time, an early hippocampal contribution to the dynamic development of reinforcement learning in middle childhood, with neurally less differentiated and more cooperative memory systems than in adults.

Introduction

As children enter school during middle childhood, they must learn to act appropriately in new situations through feedback. For example, children receive positive feedback when raising their hand before speaking in class, which reinforces them to repeat the same action in the future. Reinforcement learning (RL)[1] provides a useful mechanistic framework to describe such feedback-driven value-based learning and decision-making. RL models allow to explicitely test for the influence of separate components during value-based learning, such as model-free and model-based learning[2], social and non-social learning[3,4], or the contribution of different memory systems[5–7].

The memory systems account is a theoretical framework that proposes that different types of memory are supported by distinct neural systems in the brain. Specifically, this account suggests that there are two memory systems: a hippocampal-dependent system and a striatal-dependent system. These systems modulate memory and value-based learning, and their interactive development has been of particular interest to developmental research[8,9]. The hippocampal-dependent memory system has been shown to contribute to episodic memory during reinforcement learning and is more engaged during feedback that is presented with a delay[6,10,11], as opposed to the striatal-dependent memory system, which is more engaged after immediate feedback and supports habitual memory[5,12–14]. Specifically, hippocampal activation was greater during delayed feedback than during immediate feedback, whereas striatal activation was greater during immediate feedback than during delayed feedback[5]. The engagement of the hippocampus during delayed feedback was further supported by enhanced episodic memory for incidentally presented objects compared to objects presented with immediate feedback. Taken together, these studies suggest that feedback timing modulates the engagement of the hippocampal and striatal memory systems during value-based learning in adults. Given the differential developmental trajectories of these systems and the impact the systems have on reinforcement learning and memory, it is important to understand whether children would show similar feedback timing modulations as previously shown in adults. In addition, whether such feedback timing modulation changes over time remains largely unexplored. To this end, in this study, we examined the contributions of hippocampal and striatal structural volumes during the longitudinal development of reinforcement learning across two years in 6-to-7-year-old children.

Reinforcement learning behavior modulated by feedback timing can be modeled computationally using at least three parameters that reflect feedback-based learning and decision-making. For feedback-based learning, a learning rate parameter determines the extent to which the reward prediction error, defined as the difference between the received reward and the expected reward, influences the update of the future choice values. A higher learning rate emphasizes recent outcomes, whereas a lower learning rate reflects learning integrated over a longer outcome history[15]. Value updates may further depend on an outcome sensitivity parameter that scales the individual magnitude of received rewards. Finally, in decision-making, the inverse temperature parameter plays a key role in determining the tendency to select the more valuable choice and quantifies choice stochasticity. A higher inverse

temperature reflects more value-guided, deterministic choice behavior compared to a lower inverse temperature reflecting more random choices. Learning rates and inverse temperature have been studied extensively across development, mainly with cross-sectional studies showing mixed findings regarding their age gradients[16]. One study reported lower learning rates in children compared to adolescents[17], while other studies found no differences[18,19] or even higher learning rates in children[8,20]. Developmental differences regarding the inverse temperature parameter are slightly more consistent, with studies reporting no differences[8,21–23] or higher inverse temperature with age that suggests that behavior is increasingly value-guided and less explorative[17–19,24]. To the best of our knowledge, outcome sensitivity has not been modeled computationally across development. However, studies that linked striatal reward activation to self-reported reward sensitivity showed increasing sensitivity from childhood to adolescence[25,26].

In general, the inconsistencies regarding developmental differences in parameters may be due to their dependency on model and task properties[27], which could be reconciled by comparing developmental changes to simulation-based optimal learning[15]. Such comparisons acknowledge that optimal parameter values vary depending on the context, and it has been suggested that humans develop towards more optimal parameter values from childhood into adulthood[16]. Importantly, to our knowledge previous reinforcement learning studies with children were cross-sectional, and only two studies investigated children under 8 years of age[17,28]. Cross-sectional studies, in which developmental change is inferred as a between-subject factor, do not capture the dynamics in middle childhood if individual differences are large, whereas longitudinal studies test development as a within-subject factor, which is crucial for uncovering change across time. Thus, longitudinal changes in reinforcement learning in middle childhood as well as their putative striatal and hippocampal associations remain unknown. To this end, learning rates, outcome sensitivity and inverse temperature are relevant computational parameters to study longitudinal changes in striatal and hippocampal systems during value-based learning.

Striatal and hippocampal contributions to reinforcement learning during middle childhood may differ as these brain regions undergo major developmental changes. Whereas earlier structural studies with relatively small sample sizes showed large developmental variability and a tendency for an earlier volume peak in the striatum than in the hippocampus[29–35], a recent cross-sectional large-scale study was able to contrast striatal and hippocampal trajectories with greater granularity[36]. These data showed striatal volume peaks in the first decade which then declined throughout later developmental periods, whereas hippocampal volume showed a more protracted inverted-U-shaped trajectory that peaked in adolescence. Based on these structural findings, striatal and hippocampal systems are expected to develop functionally at different rates[37], with habit memory depending on the earlier developing striatum and episodic memory depending on the later developing hippocampus[38]. A direct investigation of the longitudinal development of both memory systems in childhood would shed light on whether the memory systems show a differential engagement similar to that of adults[5]. Such knowledge could be

useful to structure learning processes according to the developmental status. For example, children's ability to learn from delayed feedback may depend on how well their hippocampus has developed. In the same study sample, we previously reported that children's hippocampal volume was related to their family's income level[39]. Additionally, previous research has shown that stress can reduce the effectiveness of the hippocampal-dependent memory system[11]. This suggests that environmental factors such as income and stress may play a role in shaping how well children learn from delayed feedback, particularly through their impact on hippocampal development. By identifying the specific environmental factors that impact children's learning and brain development, we can identify risk groups and tailor interventions to ameliorate adverse effects.

This study aimed to explore the development of value-based learning in children and its relationship with structural brain development over time. We hypothesized that the timing of feedback would modulate children's learning from reinforcement, and that such modulation can be captured by reinforcement learning (RL) model parameters. Additionally, we predicted that children's value-based longitudinal development would shift towards more optimal learning behavior. Regarding structural brain development, we expected the striatum to be relatively mature by middle childhood compared to the protracted hippocampal maturation. Our second objective was to investigate the relationship between value-based learning and structural brain development using longitudinal structural equation modeling. We anticipated that there would be differentiated brain-cognition links between brain volume and value-based learning. Specifically, we predicted that immediate feedback learning would be more strongly associated with striatal volume, whereas hippocampal volume would be more closely linked to delayed feedback and the facilitation of episodic memory encoding. Finally, we examined how these brain-cognition dynamics would change over time by analyzing their longitudinal changes.

Methods

Participants

Children and their parents took part in 2 waves of data collection with an interval of about 2 years (*mean* = 2.07, *SD* = 0.17, *range* = 1.69 – 2.68). The inclusion criteria for wave 1 were children attending first or second grade, no psychiatric or physical health disorders, at least one parent speaking fluent German, and born full-term (≥ 37 weeks of gestation). At wave 1, 142 children (46% female, age *mean* = 7.19, *SD* = 0.46, *Range* = 6.07 - 7.98) and their parents or caregivers participated in the study. 141 children completed the probabilistic learning task, one child was later excluded due to technical problems during the task, hence 140 were included in the analysis. A subgroup of 90 children (49% female, 100% right-handed), who was randomly selected, completed magnetic resonance imaging (MRI) scanning at wave 1, and 82 of them contributed to structural data after removing scans with excessive movement. At wave 2, 127 children (46% female, age *mean* = 9.25, *SD* = 0.45, *Range* = 8.30 -10.2) continued taking part in the study, while families of the remaining children were unable to be contacted or decided not to return

to the study. 126 children at wave 2 completed the reinforcement learning task and were included in the analysis. All children at wave 2 were invited for MRI scanning, and 104 of them completed scanning (45% female, 92% right-handed). 99 children contributed to structural data, after removing scans with excessive movement. In total, 73 children contributed to the longitudinal MRI data and 126 children contributed to the longitudinal learning data. As previously reported for this study sample, we found no systematic bias due to wave 2 dropout[39].

Procedure

The study consisted of a series of cognitive tasks tested during two behavioral sessions, including a reinforcement learning task, and one MRI session at wave 1[39,40]. Two years later, the children underwent one behavioral and one MRI session. MRI scanning was performed within three weeks of the behavioral task session. Each session lasted between 150 and 180 minutes and was scheduled either on weekdays between 2 p.m. and 6 p.m. or during weekends. Before participation, the parents provided written informed consent and children's verbal assent at both waves. All children were compensated with an honorarium of 8 euro per hour.

Measures

Reinforcement learning task. Children completed an adapted reinforcement learning task[5] in which they learned the preferred associations between four cues (cartoon characters) and two choices (round-shaped or square-shaped lolli) through probabilistic feedback (87.5 % contingent and 12.5 % non-contingent reward probability). In each trial, after an initial inter-trial interval of 0.5 s, a cue and its choice options were presented for up to 7 s until the child made a choice (Figure 1, choice phase). In the delay phase, we manipulated feedback timing. For two cues, the selected choice remained visible for 1 s (immediate feedback condition), whereas for the other two cue characters, it remained visible for 5 s before feedback was given (delayed feedback condition). A final feedback phase of 2 s indicated a reward by a green frame, and a punishment by a red frame. Inside each frame, a unique object picture was shown, which was incidentally encoded and irrelevant to the task. The children were instructed to pay attention to the feedback indicated by the frame color. In an initial practice phase of 32 trials, the ch practiced the task with a fifth cartoon character not included in the actual task to avoid practice effects. The experimenter instructem them to select the choice that was most likely to give them a reward. The Experimenter checked whether the child learned the more rewarded choice during practice and let it repeat the practice task otherwise to ensure understanding of the task. In the actual task, 128 trials were presented in four blocks and with small breaks in between. Cues were presented in a mixed, pseudo-randomized order. A total of 64 unique objects were shown in the feedback phase, each one twice within the same feedback condition. In both delay phases, contingent choice and choice location remained the same for each cue within the task, but were balanced across participants by using four different task versions. At wave 2, four new cues replaced the previous ones to rule out memory effects.

Object recognition test. At wave 1, children were additionally tested for recognition memory on the object pictures that were incidentally encoded during reinforcement learning. A total of 80 objects (48 old objects and 32 new objects) were presented in randomized order. The 48 old objects (24 for each feedback condition) were selected from the 64 old objects shown during learning based on two lists to balance the shown and omitted old objects across task versions. Each old object was shown twice during learning, but if the child failed to respond during learning, no feedback or object was shown in the trial, so some objects only appeared once. These objects were excluded at the individual level (individually missing object *mean* = 2.71). At recognition, children had 4 response options ('old sure', 'old unsure', 'new unsure', 'new sure') with up to 7 s to respond. The children answered verbally, and the experimenter entered their response. At wave 2, this test was excluded due to time constraints.

[Figure 1]

Brain volume. Structural MRI images were acquired on a Siemens Magnetom TrioTim syngo 3 Tesla scanner with a 12-channel head coil (Siemens Medical AG, Erlangen, Germany) using a 3D T1–weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence (192 slices; field of view = 256 mm; voxel size = 1 mm^3; TR = 2500 ms; TE = 3.69 ms; flip angle = 7°; TI = 1100 ms). Volumetric segmentation was performed using the Freesurfer 6.0.0 image analysis suite[41]. Previous studies suggested that software tools based on adult brain templates provide inaccurate segmentation for pediatric samples, which can be improved through the use of study–specific template brains[42,43]. Thus, we created two study-specific template brains (one for each wave) using Freesurfer's "make_average_subject" command. This pipeline utilized the default adult template brain registrations of the "recon–all–all" command to average surfaces, curvatures, and volumes from all subjects into a study–specific template brain. All subjects were then re–registered to this study-specific template brain to improve segmentation accuracy. Segmented images were manually inspected for accuracy and 8 cases at wave 1 and 5 cases at wave 2 were excluded for inaccurate or failed registration due to excessive motion.

Data analysis

Behavioral learning performance. Differences in learning accuracy, win-stay probability, lose-shift probability and reaction time with the predictors feedback timing (immediate, delayed), wave (1, 2), wave 1 age, and sex (girls, boys) were tested using generalized linear mixed models (GLMM) with the R package lme4[44]. Learning accuracy was defined as the probability to choose the more rewarding option, while win-stay and lose-shift refer to the probabilities of staying with the previously chosen option after a reward and switching to the alternative choice after not receiving a reward, respectively. All reported models included random slopes for within-subject factors feedback timing and wave (see

Supplementary Material 2 for the model structure). We systematically tested main effects and interactions between the predictors and their interaction had to statistically improve the predictive ability of the model to be included in the final reported model. All predictor variables were grand-mean-centered to interpret the interaction effects independent from other predictors.

Reinforcement learning models. We compared the learning models of basic heuristic strategies and value-based learning to determine the model that could best capture children's trial-by-trial learning behavior. For heuristic strategies, we considered models that reflected a Win-stay-lose-shift (wsls) or a Win-stay (ws) strategy. Win-stay is a heuristic strategy in which the same action is repeated if it leads to a positive outcome in the previous trial, and Win-stay-lose-shift additionally switches to a different action if the previous outcome is negative. The models quantified the learning behavior for each individual *I* for each cue *c* and trial *t*. The heuristic models consisted of a weight *w* that reflected the strategy use. In the case of reward $r = 1$, *w* was equal to 1 for the chosen option (eg. choice A), and 0 for the unchosen option (e.g. choice B), thus maximizing win-stay, i.e., choosing A at the subsquent trial $t + 1$:

$$w_{i,c,t+1,A|r\,=\,1} = 1 \text{ and } w_{i,c,t+1,B|r\,=\,1} = 0 \quad (1)$$

For trials $r = 0$ (applicable only to the wsls model), model weights were the opposite, maximizing lose-shift:

$$w_{i,c,t+1,A|r\,=\,0} = 0; \; w_{i,c,t+1,B|r\,=\,0} = 1 \quad (2)$$

The initial weights for both choices were set to $w_{i,c,t=1} = 0.5$. The weight *w* then scaled the parameter τ_wsls or τ_ws to estimate the individual strategy use during decision-making. The choice probabilities were calculated using the softmax function, eg., for the chosen option *A*:

$$p(A) = \frac{exp^{w_{i,c,t,A}*\tau\_wsls_i}}{exp^{w_{i,c,t,A}*\tau\_wsls_i} + exp^{w_{i,c,t,B}*\tau\_wsls_i}} \quad (3)$$

Thus, a higher probability of strategy use was reflected by a larger value of τ_wsls or τ_ws.

For value-based learning, we considered a Rescorla-Wagner model and several variants based on our theoretical conceptions. The baseline value-based model vbm_1 updated the value *v* of the selected choice (*A* or *B*) for the next trial *t*. This value update was determined by calculating the difference between the received reward *r* and the expected value *v* of the selected choice, which was the reward prediction error. The value update was further scaled by a learning rate α $(0 < \alpha < 1)$:

$$v_{i,c,t+1,A} = v_{i,c,t,A} + \alpha_i(r_{i,c,t} - v_{i,c,t,A}) \quad (4)$$

When the outcome sensitivity parameter ρ $(0 < \rho < 20)$ was included, the reward was additionally scaled at the value update:

$$v_{i,c,t+1,A} = v_{i,c,t,A} + \alpha_i(\rho_i * r_{i,c,t} - v_{i,c,t,A}) \quad (5)$$

The inverse temperature parameter $\tau(0 < \tau < 20)$ was included in the softmax function to compute choice probabilities:

$$p(A) = \frac{exp^{v_{i,c,t,A}*\tau_i}}{exp^{v_{i,c,t,A}*\tau_i} + exp^{v_{i,c,t,B}*\tau_i}} \quad (6)$$

Note, however, that outcome sensitivity and inverse temperature are difficult to fit simultaneously due to non-identifiability issues[45]. Therefore, models including outcome sensitivity fixed the inverse temperature at 1 (outcome sensitivity model family), and models with the inverse temperature in turn fixed outcome sensitivity at 1 (inverse temperature model family). Each model family consisted of 4 model variants vbm_{1-4} $(1\alpha 1\tau, 2\alpha 1\tau, 1\alpha 2\tau, 2\alpha 2\tau)$ and vbm_{5-8} $(1\alpha 1\rho, 2\alpha 1\rho, 1\alpha 2\rho, 2\alpha 2\rho)$, in which each parameter was either separated by feedback timing or kept as a single parameter across feedback conditions. Our baseline value-based model vbm_1 included a single learning rate and a single inverse temperature $(1\alpha 1\tau)$.

Parameter estimation. All choice data were fitted in a hierarchical Bayesian analysis using the Stan language in R[46,47] adopted from the hBayesDM package[48]. Posterior parameter distributions were estimated using Markov chain Monte Carlo (MCMC) sampling running 4 chains each with 3,000 iterations, using the first half of the chain as warmup, and group-level parameters and individual-level parameters were estimated simoultaneously. The hierarchical Bayesian approach provides more stable and reliable parameter estimates as opposed to point-estimation approaches like maximum likelihood estimation[49]. Each model fit both wave 1 and wave 2 data at once, considering the correlation structure of the same parameter across waves, to account for within-subject dependency using the Cholesky decomposition. The Cholesky decomposition used a Lewandowski-Kurowicka-Joe prior of 2, and all other group-level parameters had a prior normal distribution, Normal (0, 0.5). Non-response trials (wave 1 = 2.41%, wave 2 = 0.97% on average) were excluded in advance.

Model simulation and model-derived learning score. To appropriately interpret the parameter results with respect to the optimal parameter combination of the winning model, we simulated 5,000,000 individual datasets using 10,000 different parameter value combinations (covering the whole range of each parameter) to identify the optimal parameter combination of the winning model that was selected by model comparison. In addition, we computed the model-derived mean choice probability of the contingent, i.e., the more rewarded option, and we referred to it as the model-derived learning score. This model-derived choice probability differs from the observed empirical choice probability (i.e., the accuracy of selecting the more rewarded option), because the model-derived learning score combines the model with the data by incorporating latent information carried out by key learning parameters. Thus, the learning score captures observed behavior based on trial-by-trial latent processes predicted by value-based models. We used this as metric to interpret the fitted posterior parameters in relation to the optimal parameter combination of our probabilistic learning task.

Model selection and validation. We conducted a 2-step sequential procedure for the model development and model selection. As a first step, we compared model evidence for the baseline value-based model that does not separate learning rate and inverse temperature by feedback timing (vbm_1:$1\alpha, 1\tau$) to the

non-value-based, heuristic strategy models that reflect Win-stay or Win-stay-lose-shift strategy behavior ($ws, wsls$). As a second step, we compared model evidence for 8 value-based model variants, 4 of the model family with learning rate and inverse temperature ($1\alpha1\tau, 2\alpha1\tau, 1\alpha2\tau, 2\alpha2\tau$) and 4 of the model family with learning rate and outcome sensitivity ($1\alpha1\rho, 2\alpha1\rho, 1\alpha2\rho, 2\alpha2\rho$). This allowed us to compare whether children showed separable effects of feedback timing on one of the model parameters. We compared the model fit using Bayesian leave-one-out cross-validation and obtained the expected log pointwise predictive density ($elpd_{loo}$) using the R package loo[50]. We further computed the model weights (*Pseudo-BMA+*) using Pseudo Bayesian model averaging stabilized by Bayesian bootstrap with 100,000 iterations[51]. To validate our models, we estimated predictive accuracy by comparing one-step-ahead model predictions with the choice data[15,52]. We performed parameter recovery for the winning model and model recovery by comparing it to a set of models used during model comparison (Supplementary Material 1)[53].

Episodic memory at wave 1

We predicted the individual corrected recognition memory (hits-false alarms) by feedback condition in a linear mixed effects model using the R package lme4[44]. A total of 140 children completed the recognition memory test and 138 were included in the analysis, with two being excluded due to negative corrected recognition memory value (i.e., poor recognition memory). Age and sex were controlled for as covariates.

Longitudinal brain-cognition links

We used latent change score (LCS) models to examine the longitudinal relationships between brain and learning score measures. LCS models are longitudinal structural equation models that have been widely applied to estimate developmental changes and coupling effects across domains such as the brain and cognition[54,55]. LCS models allow the definition of specific paths between multiple variables to test explicit hypotheses and estimate latent change from the observed variables that account for measurement error and increase testing power[56]. We compiled univariate LCS models for each variable separately (learning scores and brain volumes) to examine whether there was significant individual variance and change, which could be related within a multivariate LCS model as a next step. Model fit had to be at least acceptable, with a comparative fit index (*CFI*) > 0.95, standardized root mean square residual (*SRMR*) < .08 and root mean square error of approximation (*RMSEA*) < .08[57]. Age and sex were included as covariates at wave 1, as well as the estimated total intracranial volume (eTIV) when brain volume was included in the model. Multivariate LCS models allow to estimate meaningful brain-cognition relationships: a wave 1 covariance between brain and cognition, brain predicting change onto cognition, or vice versa, and a covariance in both brain and cognition change scores (wave 1 to wave 2). Before compiling the variables into an LCS model, they were checked for outliers ± 4 *SD* around the mean. We identified one outlier for the learning rate at wave 2, which was removed for the explorative LCS model

that included model parameters. There were no further outliers in other cognitive variables or brain volumes. Continuous variables were standardized to the wave 1 measure so that wave 2 values represent the change from wave 1, sex was contrast-coded (girls = 1, boys = -1).

Results

Behavioral results

First, we were interested in whether children showed behavioral differences between waves and feedback timing. A descriptive overview is provided in Table 1 and Figure 2. The details of the reported GLMM models, including the random effects structure and the effects of age and sex, are described in the Supplementary Material 2. Since some children were poor learners who failed to reach 50 % average accuracy in their last 20 trials (13 children at wave 1 and 6 children at wave 2), we also performed behavioral analyses with a reduced dataset in which results remained unchanged (Supplementary Materials 3).

Children's learning improved between waves. With the complete dataset, we found that increased learning accuracy (i.e., the probability of choosing the more rewarding option) was predicted at wave 2 compared to wave 1, but there were no differences in accuracy by feedback timing ($\beta_{wave=2}$ = .550, *SE* = .061, z = 8.97, $p < .001$, $\beta_{feedback=delayed}$ = .013, *SE* = .024, z = 0.54, p = .590). Furthermore, win-stay probability increased and lose-shift probability decreased longitudinally, again without differences by feedback timing (WS: $\beta_{wave=2}$ = .586, *SE* = .071, z = 8.22, $p < .001$, LS: $\beta_{wave=2}$ = -.586, *SE* = .071, z = -8.22, p < .001). Reaction times were faster at wave 2 compared to wave 1, and they were faster for delayed compared to immediate feedback trials ($\beta_{wave=2}$ = -218, *SE* = 22.7, t = -9.61, $p < .001$, $\beta_{feedback=delayed}$ = -14.0, *SE* = 6.61, t = -2.12, p = .036). To summarize, children, on average, improved their accuracy over 2 years, while the win-stay probability increased and the lose-shift probability decreased between waves. Children were able to respond faster to cues paired with delayed feedback compared to cues paired with immediate feedback, and they became faster in their decision-making across waves (see mixed model effects overview in Table 1). Of note, reaction times were largely uncorrelated with accuracy and switching behavior (win-stay, lose-shift), while accuracy and switching behavior showed significant correlations at both waves (Figure 2D).

[Table 1]

[Figure 2]

Modeling results

Children's behavior was best described by value-based learning. We conducted a 2-step sequential procedure for model development and model selection. Model comparison using leave-one-out cross validation showed evidence in favor of the value-based learning model, reflected in the highest expected log pointwise predictive density and highest model weights, confirming that children's learning behavior in the longitudinal data can generally be better described by a value-based rather than by a heuristic strategy model ($elpd_{loo}$ = -15154.9, *pseudo-BMA+* = 1, Table 2). Children whose individual fit was better for a heuristic model ($wsls$) than for the value-based model (vbm_1), were at both waves more likely to be poor learners (defined as an accuracy below 50% in the last 20 trials). Taken together, children's learning behavior was best described by a value-based model, and a heuristic strategy model captured more poor learners compared to a value-based model.

[Table 2]

Feedback timing modulated choice stochasticity. Model vbm_3 ($1\alpha2\tau$) showed the largest model evidence, reflected in the highest expected log pointwise predictive density and highest model weights and suggests that feedback timing affected the inverse temperature, but not the learning rate or outcome sensitivity ($elpd_{loo}$ = -15045.3, *pseudo-BMA+* = 0.73, Table 2). Table 3 and Figure 3A provide a descriptive overview of the winning model parameters. Of note, there were only small differences in model fit ($elpd_{loo}$) to the second-best model ($vbm_7, 1\alpha2\rho$, $\Delta elpd_{loo}$ = -2.93, $elpd\_SE_{loo}$ = 2.92, *pseudo-BMA+* = 0.24), which suggests a potential separable feedback timing effect on outcome sensitivity. The average inverse temperature did not differ by feedback condition, but showed large within-person condition differences at both waves, indicating individual differences in feedback timing modulation (wave 1: $\Delta\tau_{del-ime}$ *Mean* = 0.22, *SD* = 3.80, *Range* = 21.74, wave 2: $\Delta\tau_{del-ime}$ *Mean* = 0.35, *SD* = 3.70, *Range* = 24.03). The correlations between the parameters are shown in Supplementary Material 4.

Since reaction times were predicted by feedback timing behaviorally, and inverse temperature is assumed to reflect decision-making, we were interested in whether differences in reaction time were related to inverse temperature differences. Indeed, at both waves, children who responded faster during delayed compared to immediate feedback had a higher inverse temperature at delayed compared to immediate feedback (wave 1: r = -.261, t = -3.18, p = .002, wave 2: r = -.345, t = -4.10, p < .001, Figure 3B). Taken together, children's learning behavior was best described by a value-based model, where feedback timing modulated individual differences in the choice rule during value-based learning. Interestingly, the differences in the choice rule and reaction time f were correlated. Specifically, more value-guided choice behavior (i.e., higher inverse temperature) was related to faster responses during

delayed feedback relative to immediate feedback, suggesting a link between model parameter and behavior in relation to feedback timing.

[Table 3]

[Figure 3]

Children's value-based learning became more optimal. Next, we compared the parameter space according to model simulation (Figure 4A) with the empirical posterior parameters fitted by the winning model (Table 3, Figure 4B) to determine whether children increased their value-based learning towards more optimal parameter combinations. Both fitted and simulated parameter combinations allowed us to derive a learning score that captured learning performance according to the winning value-based model. Note that the learning score was defined as the average choice probability for the more rewarded choice option. We refer to these model-derived choice probabilities as learning score, since they reflect value-based learning and combine information of learned values, that depend on the learning rate, and values translated into choice probabilities, that depend on the inverse temperature. Thus, a higher learning score reflects more optimal value-based learning. We simulated 10,000 parameter combinations and created a learning score map according to each parameter combination (Figure 4A). The optimal parameter combination was at a learning rate α = 0.29, and an inverse temperature τ = 19.8, and with an average learning score of 96.5 % (Figure 4A). Children's fitted average learning rates ranged 0.01 – 0.22 and inverse temperature 6.73 – 18.70 and were outside the parameter space above 96 % learning score (Table 3 and Figure 4A). The longitudinal average increase in learning rate and inverse temperature were mirrored by average increases in the learning scores, confirming our prediction that their parameters developed towards optimal value-based learning (arrow in Figure 4B). The one-step ahead predictions of the winning model captured children's choices overall well, predictive accuracies were 65.3 % at wave 1 and 75.7 % at wave 2 (Figure 4C).

[Figure 4]

Longitudinal brain-cognition links

Significant longitudinal change in brain and cognition. We first performed univariate LCS model analyses to estimate a latent change score of immediate and delayed learning scores as well as striatal and hippocampal volumes (see descriptive changes in Figure 5B-C). All four variables of interest showed significant positive mean changes and variances, and all univariate models provided a good fit

to the data (Supplementary Material 5). This allowed us to further relate the differences in structural brain changes to changes in learning.

Hippocampal volume exhibited more protracted development during middle childhood. We next fitted a bivariate LCS model to compare striatal and hippocampal change scores. We theorized that by middle childhood, the striatum would be relatively mature, whereas the hippocampus continues to develop. We progressively constructed multiple LCS models to test this idea. First, the bivariate LCS model provided a good data fit (χ^2 (14) = 10.09, *CFI* = 1.00, *RMSEA* (*CI*) = *0* (0-.06), *SRMR* = .04). We then further fitted two constrained models, to see whether setting the mean striatal change or the mean hippocampal change to 0 would lead to a drop in the model fit. Compared to the unrestricted model, the constrained model that assumed no striatal change did not lead to a drop in model fit ($\Delta\chi^2$ (1) = 2.74, p = .098), whereas the model that assumed hippocampal change dropped in model fit ($\Delta\chi^2$ (1) = 12.69, p < .001). Finally, we tested a more stringent assumption of equal change for striatal and hippocampal volumes, in which the model dropped in model fit compared to the unrestricted model ($\Delta\chi^2$ (1) = 18.04, p < .001) and suggests that striatal and hippocampal change differed. Together, these results support our postulation of separable maturational brain trajectories in our study sample, suggesting that the hippocampus continued to grow in middle childhood, whereas striatal volume increased less.

Hippocampal and striatal volume showed distinct associations to learning. We fitted a four-variate LCS model to test our prediction of selective brain-cognition links. Specifically, we assumed a larger contribution of striatal volume at immediate learning, and a larger contribution of hippocampal volume at delayed learning. The LCS model provided good data fit (χ^2 (27) = 15.4, *CFI* = 1.00, *RMSEA* (*CI*) = 0 (0 – .010, *SRMR* = .045), and all relevant paths are shown in Figure 5D (see Table 4 for a detailed model overview). For the striatal associations to cognition, we found that wave 1 striatal volume covaried with both immediate learning score and delayed learning score ($\phi_{STR_{w1},LS_{i,w1}}$ = 0.19, z = 2.52, *SE* = 0.07, p = .012, $\phi_{STR_{w1},LS_{d,w1}}$ = 0.18, z = 2.37, *SE* = 0.07, p = .018). Constraining the striatal association to immediate learning to 0 worsened the model fit relative to the unrestricted model ($\Delta\chi^2$ (1) = 5.66, p = .017), which was the same when constraining the striatal association to delayed learning to 0 ($\Delta\chi^2$ (1) = 5.14, p = .023). In summary, larger striatal volume was associated with better learning scores for both immediate and better delayed feedback.

Hippocampal volume, on the other hand, only covaried with delayed learning at wave 1 ($\phi_{HPC_{w1},LS_{d,w1}}$ = 0.14, z = 2.05, *SE* = 0.07, p = .041), not with immediate learning score ($\phi_{HPC_{w1},LS_{i,w1}}$ = 0.12, z = 1.68, *SE* = 0.07, p = .092). Fixing the path between hippocampal volume and delayed learning to 0 worsened the model fit relative to the unrestricted model ($\Delta\chi^2$ (1) = 4.19, p = .041), but not when its path to immediate learning was constrained to 0 ($\Delta\chi^2$ (1) = 2.94, p = .086). This suggests that larger hippocampal volume was spcifically associated with better delayed learning. As a next step, the associations between striatum and hippocampus to immediate or delayed learning was directly compared against each other.

A model equal-constraining striatal and hippocampal paths to immediate learning ($\Delta\chi^2$ (1) = 0.41, p = .521) and another model equal-constraining these paths to delayed learning ($\Delta\chi^2$ (1) = 0.14, p = .707) did not lead to a worse model fit compared to the unrestricted model, which suggests that the brain-cognition links have considerable overlap. This is in line with the high wave 1 covariance and change-change covariance within the brain and cognition domain (see Table 4). We found no longitudinal links between the brain and cognition domains, which suggests that the found brain-cognition links at wave 1 remained longitudinally stable (see Supplementary Material 5 for an exploratory LCS model that related the model parameters to striatal and hippocampal volume).

Taken together, the confirmatory LCS model results were in line with our predictions of a relatively larger involvement of the hippocampus during delayed feedback learning, but the findings on striatal volume disconfirmed a selective association with immediate feedback learning and suggest a more general role of the striatum in both learning conditions.

Weak evidence for enhanced episodic memory during delayed feedback. Finally, we investigated whether a hippocampal contribution at delayed feedback would selectively enhance episodic memory. Episodic memory, as measured by individual corrected object recognition memory (hits - false alarms), showed at trend better memory for items shown in the delayed feedback condition ($\beta_{feedback=delayed}$ = .009, SE =.005, t = 1.83, p = .069, see Figure 5A). To summarize, there was weak support for enhanced episodic memory during delayed compared to immediate feedback, in line with the idea of a selective association between hippocampal volume and delayed feedback learning.

[Figure 5]

[Table 4]

Discussion

In this study, we examined the longitudinal development of value-based learning in middle childhood and its associations with striatal and hippocampal volumes that were predicted to differ by feedback timing. Children improved their learning in the 2-year study period. Behaviorally, learning was improved by an increase in accuracy and a reduction in reaction time (i.e., faster responses). Further, children's switching behavior improved by an increase in win-stay and a decrease in lose-shift behavior. Computationally, learning was enhanced by an increase in learning rate and inverse temperature, which together constituted more optimal value-based learning. Further, feedback timing modulated specifically the inverse temperature. In terms of brain structures, we found that longitudinal changes in hippocampal volume were larger compared to striatal volume, which suggests more protracted hippocampal maturation. The brain-cognition links were longitudinally stable and partially confirmed our hypotheses.

In line with previous adult literature and our assumption, hippocampal volume was more strongly associated with delayed feedback learning, and there was weak evidence of enhanced episodic memory performance under delayed feedback compared to immediate feedback. Contrary to our expectations, striatal volume was associated with not just immediate but also delayed feedback learning, suggesting a common involvement of the striatum during value-based learning in middle childhood across timescales.

Children's learning improvement between waves was described behaviorally by increased win-stay and decreased lose-shift behavior. Our finding is in line with cross-sectional studies in the developmental literature that reported increased learning accuracy and win-stay behavior[58,59]. Our longitudinal dataset with younger children further suggests that learning change is not only accompanied by increased win-stay, but also decreased lose-shift behavior. We found lower learning performance and less optimal switching behavior in girls compared to boys, which could point to sex differences for reinforcement learning during middle childhood (Supplementary Material 2). Previous studies have found both male and female advantages depending on their age and the type of learning task[38,60,61]. Alternatively, sex differences may have been driven by confounding variables not included in the analysis. Computationally, we found longitudinally increased and more optimal learning rate and inverse temperature, as shown by simulation data, that add to the growing literature of developmental reinforcement learning[16]. Our study underscores the importance of relating empirical values to simulation-based optimal values, as task characteristics such as reward probability and learning environment stability determine the range of optimal parameter values[27].

Despite a relatively immature hippocampal structure in middle childhood, our results confirmed a longitudinally stable association between hippocampal volume and delayed feedback learning. However, episodic memory in this learning condition was not enhanced. This suggests a developmentally early hippocampal contribution to value-based learning during delayed feedback, which does not modulate episodic memory as much as compared to adults. Therefore, our study partially extends the findings from the adult literature to middle childhood[5,12–14]. The reduced effect of delayed feedback on episodic memory may be due to the protracted development of hippocampal maturation. In an aging study with a similar task, older adults failed to exhibit enhanced episodic memory for objects presented during delayed feedback trials, and they showed no enhanced hippocampal activation during delayed feedback and[14]. Therefore, the findings converge nicely at both childhood and older adulthood, during which the structural and functional integrity of hippocampus are known to be less optimal than at younger adulthood[62–64].

Our brain-cognition links were only partially confirmed, as striatal volumes exhibited associations with not just immediate learning scores, as we predicted, but also with delayed learning scores. This result suggests that the striatum may be important for value-based learning in general rather than exhibiting a

selective association with immediate feedback learning. This is also what we found in an explorative analysis that related the striatum to learning rate in general and further predicted longitudinal change in learning rate (Supplemental Material 4). This overall reduced brain-behavior specificity could reflect less differentiated memory systems during development, similar to findings from aging research. Here, older adults exhibited stronger striatal and hippocampal co-activation during both implicit and explicit learning, compared to more dissociable brain-behavior relationships in younger adults[65]. Interestingly, even in young adults, clear dissociations between memory systems such as in non-human lesion studies are uncommon, and factors like stress modulate their cooperative interaction[6,10,11,66,67]. Further, there are methodological differences to previous studies that could explain why striatal volumes were not uniquely associated with immediate learning in our study. For example, previous studies related reward prediction errors to striatal and hippocampal activation[5,13,14], whereas we examined individual differences in brain structure and the model-derived learning scores. Future functional neuroimaging studies with children could further clarify whether children's memory systems are indeed less differentiated and explain the attenuated modulation by feedback timing. Taken together, compared to the adult literature, our results with children showed that the hippocampal structure was associated with delayed feedback learning, but did not enhance episodic memory encoding, while the striatum generally supported value-based learning. These findings point towards a developmental effect of less differentiated and more cooperative memory systems in middle childhood.

Our computational modeling results revealed a separable effect of feedback timing on inverse temperature, which suggests that the memory systems modulated learning during decision-making. The reported behavioral differences in reaction time and their correlation to the inverse temperature further support the idea of a decision-related mechanism, as we found children to respond faster during delayed feedback trials and faster responding children also exhibited more value-guided choice behavior (i.e. higher inverse temperature) during delayed compared to immediate feedback. The hippocampus may contribute to a decision-related effect in the delayed feedback condition by facilitating the encoding and retrieval of learned values[68]. This is in contrast to previous event-related fMRI and EEG studies reporting feedback timing modulations at value update[5,13,14], which may be due to at least two reasons. First, we did not include a functional brain measure to examine its differential engagement during the choice and feedback phases. Second, in such a reinforcement learning task, disentangling model parameters from the choice and feedback phases can be challenging, such as for the inverse temperature and outcome sensitivity[69]. Hippocampal engagement at delayed feedback may enhance outcome sensitivity, as well as facilitate cue-choice associations and improve retrieval and choice behavior. A mechanism facilitating retrieval seems especially relevant in our paradigm, where multiple cues were learned and presented in a mixed order, thus creating a high memory load. To summarize, our study results suggest that feedback timing can modulate decision-making. However, disentangling the effects of inverse temperature and outcome sensitivity is challenging and warrants careful interpretation. Future

studies might shed new light by examining neural activations at both task phases, and by choosing a task design that allows independent manipulations on these phases and associated model parameters, e.g., by using different reward magnitudes during reinforcement learning, or by studying outcome sensitivity without decision-making.

One aim of developmental investigations is to identify the emergence of brain and cognition dynamics, such as the hippocampal-dependent and striatal-dependent memory systems, which have been shown to engage during reinforcement learning depending on the delay in feedback delivery. Our longitudinal study partially confirmed these brain-cognition links in middle childhood but with less specificity as previously found in adults.

An early existing memory system dynamic, similar to that of adults, is relevant for applying reinforcement learning principles at different timescales. For example, in a school context, learning processes can be better structured according to their development. Furthermore, probabilistic learning from delayed feedback may be a potential diagnostic tool to examine the hippocampal-dependent memory system during learning in children at risk. Environmental factors such as stress[11] and socioeconomic status[39,70] have been shown to affect hippocampal structure and function and may contribute to a heightened risk for psychopathology in the long term[71–73]. Deficits in hippocampal-dependent learning may be particularly relevant to psychopathology since dysfunctional behavior may arise from a tendency to prioritize short-term consequences over long-term ones[74,75] and from the maladaptive application of previously learned behavior in inappropriate contexts[76].

Another key question is whether developmental trajectories observed cross-sectionally are also confirmed by longitudinal results, such as for the learning rate and inverse temperature. Our results show developmental improvements in these learning parameters in only two years. This suggests that the initial two years of schooling constitute a dynamic period for feedback-based learning, in which contingent feedback is important in shaping behavior and development.

Additional Information

Funding. This study was supported by the Jacobs Foundation [grant 2014–1151] to YLS and CH. The work of YLS was also supported by the European Union (ERC-2018-StG-PIVOTAL-758898), the Deutsche Forschungsgemeinschaft (German Research Foundation, Project ID 327654276, SFB 1315,'Mechanisms and Disturbances in Memory Consolidation: From Synapses to Systems'), and the Hessisches Ministerium für Wissenschaft und Kunst (HMWK; project 'The Adaptive Mind').

Conflicts of interest. The authors declare no competing financial interests.

623 Ethics approval. This study was approved by the "Deutsche Gesellschaft für Psychologie" ethics
624 committee (YLS_012015).
625
626 Availability of data and code. https://osf.io/pju65/
627
628 Author ORCIDs.
629 JF: https://orcid.org/0000-0003-0505-0798
630 LZ: https://orcid.org/0000-0002-9586-595X
631 LR: https://orcid.org/0000-0002-0144-5605
632 JJM: https://orcid.org/0000-0002-3893-8008
633 JT: https://orcid.org/0000-0001-8166-2441
634 CH: https://orcid.org/0000-0002-6580-6326
635 YLS: https://orcid.org/0000-0001-8922-7292
636

References

1. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction*. (MIT press, 2018).
2. Gläscher, J., Daw, N., Dayan, P. & O'Doherty, J. P. States versus Rewards: Dissociable neural prediction error signals underlying model-based and model-free reinforcement learning. *Neuron* **66**, 585 (2010).
3. Bolenz, F., Reiter, A. M. F. & Eppinger, B. Developmental Changes in Learning: Computational Mechanisms and Social Influences. *Front. Psychol.* **0**, 2048 (2017).
4. Zhang, L. & Gläscher, J. A brain network supporting social influences in human decision-making. *Sci. Adv.* **6**, 1–20 (2020).
5. Foerde, K. & Shohamy, D. Feedback Timing Modulates Brain Systems for Learning in Humans. *J. Neurosci.* **31**, 13157–13167 (2011).
6. Packard, M. G. & Goodman, J. Factors that influence the relative use of multiple memory systems. *Hippocampus* **23**, 1044–1052 (2013).
7. Goodman, J. & Packard, M. G. Memory Systems of the Basal Ganglia. *Handb. Behav. Neurosci.* **24**, 725–740 (2016).
8. Davidow, J. Y., Foerde, K., Galván, A. & Shohamy, D. An Upside to Reward Sensitivity: The Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. *Neuron* **92**, 93–99 (2016).
9. Hartley, C. A., Nussenbaum, K. & Cohen, A. O. Interactive Development of Adaptive Learning and Memory. 1–27 (2021).
10. Packard, M. G., Goodman, J. & Ressler, R. L. Emotional modulation of habit memory: neural mechanisms and implications for psychopathology. *Curr. Opin. Behav. Sci.* **20**, 25–32 (2018).
11. Schwabe, L. & Wolf, O. T. Stress and multiple memory systems: from 'thinking' to 'doing'. *Trends Cogn. Sci.* **17**, 60–68 (2013).
12. Foerde, K., Race, E., Verfaellie, M. & Shohamy, D. A role for the medial temporal lobe in feedback-driven learning: Evidence from amnesia. *J. Neurosci.* **33**, 5698–5704 (2013).
13. Höltje, G. & Mecklinger, A. Feedback timing modulates interactions between feedback processing and memory encoding: Evidence from event-related potentials. *Cogn. Affect. Behav. Neurosci. 2020 202* **20**, 250–264 (2020).
14. Lighthall, N. R., Pearson, J. M., Huettel, S. A. & Cabeza, R. Feedback-Based Learning in Aging: Contributions and Trajectories of Change in Striatal and Hippocampal Systems. *J. Neurosci.* **38**, 8453–8462 (2018).
15. Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J. & Lamm, C. Using reinforcement learning models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Soc. Cogn. Affect. Neurosci.* **15**, 695–707 (2020).
16. Nussenbaum, K. & Hartley, C. A. Reinforcement learning across development: What insights can we draw from a decade of research? *Developmental Cognitive Neuroscience* **40**, (2019).

17. Decker, J. H., Lourenco, F. S., Doll, B. B. & Hartley, C. A. Experiential reward learning outweighs instruction prior to adulthood. *Cogn. Affect. Behav. Neurosci.* **15**, 310–320 (2015).
18. Javadi, A. H., Schmidt, D. H. K. & Smolka, M. N. Differential representation of feedback and decision in adolescents and adults. *Neuropsychologia* **56**, 280–288 (2014).
19. Palminteri, S., Kilford, E. J., Coricelli, G. & Blakemore, S. J. The Computational Development of Reinforcement Learning during Adolescence. *PLoS Comput. Biol.* **12**, 1–25 (2016).
20. Master, S. L. *et al.* Distentangling the systems contributing to changes in learning during adolescence. *Dev. Cogn. Neurosci.* **41**, 100732 (2020).
21. Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D. & Brem, S. Cognitive flexibility in adolescence: Neural and behavioral mechanisms of reward prediction error processing in adaptive decision making during development. *Neuroimage* **104**, 347–354 (2015).
22. Moutoussis, M. *et al.* Change, stability, and instability in the Pavlovian guidance of behaviour from adolescence to young adulthood. *PLoS Comput. Biol.* **14**, (2018).
23. Van Den Bos, W., Cohen, M. X., Kahnt, T. & Crone, E. A. Striatum-medial prefrontal cortex connectivity predicts developmental changes in reinforcement learning. *Cereb. Cortex* **22**, 1247–1255 (2012).
24. Rodriguez Buritica, J. M., Heekeren, H. R. & van den Bos, W. The computational basis of following advice in adolescents. *J. Exp. Child Psychol.* **180**, 39–54 (2019).
25. Galván, A. The Teenage Brain: Sensitivity to Rewards. *Curr. Dir. Psychol. Sci.* **22**, 88–93 (2013).
26. van Duijvenvoorde, A. C. K. *et al.* A cross-sectional and longitudinal analysis of reward-related brain activation: Effects of age, pubertal stage, and reward sensitivity. *Brain Cogn.* **89**, 3–14 (2014).
27. Eckstein, M. K., Wilbrecht, L. & Collins, A. G. E. What do RL Models Measure ? Interpreting Model Parameters in Cognition and Neuroscience. *Curr. Opin. Behav. Sci.* **41**, 128–137 (2021).
28. Cohen, A. O., Nussenbaum, K., Dorfman, H. M., Gershman, S. J. & Hartley, C. A. The rational use of causal inference to guide reinforcement learning strengthens with age. *npj Sci. Learn.* **5**, 1–9 (2020).
29. Raznahan, A. *et al.* Longitudinal four-dimensional mapping of subcortical anatomy in human development. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1592 (2014).
30. Wierenga, L. *et al.* Typical development of basal ganglia, hippocampus, amygdala and cerebellum from age 7 to 24. *Neuroimage* **96**, 67–72 (2014).
31. Giedd, J. N. Structural Magnetic Resonance Imaging of the Adolescent Brain. *Ann. N. Y. Acad. Sci.* **1021**, 77–85 (2004).
32. Uematsu, A. *et al.* Developmental Trajectories of Amygdala and Hippocampus from Infancy to Early Adulthood in Healthy Individuals. *PLoS One* **7**, e46970 (2012).
33. Giedd, J. N. *et al.* Child Psychiatry Branch of the National Institute of Mental Health

Longitudinal Structural Magnetic Resonance Imaging Study of Human Brain Development. *Neuropsychopharmacology* **40**, 43 (2015).

34. Goodman, J., Marsh, R., Peterson, B. S. & Packard, M. G. Annual research review: The neurobehavioral development of multiple memory systems--implications for childhood and adolescent psychiatric disorders. *J. Child Psychol. Psychiatry.* **55**, 582–610 (2014).
35. Goddings, A. L. *et al.* The influence of puberty on subcortical brain development. *Neuroimage* **88**, 242–251 (2014).
36. Dima, D. *et al.* Subcortical volumes across the lifespan: Data from 18,605 healthy individuals aged 3–90 years. *Hum. Brain Mapp.* 1–18 (2021). doi:10.1002/hbm.25320
37. Lavenex, P. & Banta Lavenex, P. Building hippocampal circuits to learn and remember: Insights into the development of human memory. *Behavioural Brain Research* **254**, 8–21 (2013).
38. Mandolesi, L., Petrosini, L., Menghini, D., Addona, F. & Vicari, S. Children' s radial arm maze performance as a function of age and sex. *Int. J. Dev. Neurosci.* **27**, 789–797 (2009).
39. Raffington, L. *et al.* Stable longitudinal associations of family income with children's hippocampal volume and memory persist after controlling for polygenic scores of educational attainment. *Dev. Cogn. Neurosci.* **40**, 100720 (2019).
40. Raffington, L. *et al.* Effects of stress on 6- and 7-year-old children's emotional memory differs by gender. *J. Exp. Child Psychol.* **199**, 104924 (2020).
41. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
42. Phan, T. V., Smeets, D., Talcott, J. B. & Vandermosten, M. Processing of structural neuroimaging data in young children: Bridging the gap between current practice and state-of-the-art methods. *Dev. Cogn. Neurosci.* **33**, 206–223 (2018).
43. Schoemaker, D. *et al.* Hippocampus and amygdala volumes from magnetic resonance images in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation. *Neuroimage* **129**, 1–14 (2016).
44. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
45. Brown, V. M. *et al.* Reinforcement Learning Disruptions in Individuals with Depression and Sensitivity to Symptom Change following Cognitive Behavioral Therapy. *JAMA Psychiatry* (2021). doi:10.1001/jamapsychiatry.2021.1844
46. Stan Development Team. RStan: the R interface to Stan. R package version 2.21.2. *http//mc-stan. org* (2021).
47. R Core Team. R: A Language and Environment for Statistical Computing. (2021).
48. Ahn, W.-Y., Haines, N. & Zhang, L. Revealing Neurocomputational Mechanisms of Reinforcement Learning and Decision-Making With the hBayesDM Package. *Comput. Psychiatry* **1**, 24 (2017).

748 49. Brown, V. M., Chen, J., Gillan, C. M. & Price, R. B. Improving the Reliability of
749 Computational Analyses: Model-Based Planning and Its Relationship With Compulsivity. *Biol.*
750 *Psychiatry Cogn. Neurosci. Neuroimaging* **5**, 601–609 (2020).
751 50. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out
752 cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
753 51. Yao, Y., Vehtari, A., Simpson, D. & Gelman, A. Using Stacking to Average Bayesian
754 Predictive Distributions (with Discussion). *Bayesian Anal.* **13**, 917–1007 (2018).
755 52. Crawley, D. *et al.* Modeling flexible behavior in childhood to adulthood shows age-dependent
756 learning mechanisms and less optimal learning in autism in each age group. *PLoS Biol.* **18**, 1–
757 25 (2020).
758 53. Wilson, R. C. & Collins, A. G. E. Ten simple rules for the computational modeling of
759 behavioral data. *Elife* **8**, 1–33 (2019).
760 54. Kievit, R. A. *et al.* Developmental cognitive neuroscience using latent change score models: A
761 tutorial and applications. *Dev. Cogn. Neurosci.* **33**, 99–117 (2018).
762 55. Ferrer, E. & McArdle, J. J. Longitudinal modeling of developmental changes in psychological
763 research. *Curr. Dir. Psychol. Sci.* **19**, 149–154 (2010).
764 56. Sluis, S. van der, Verhage, M., Posthuma, D. & Dolan, C. V. Phenotypic Complexity,
765 Measurement Bias, and Poor Phenotypic Resolution Contribute to the Missing Heritability
766 Problem in Genetic Association Studies. *PLoS One* **5**, e13929 (2010).
767 57. Little, T. *Longitudinal structural equation modeling*. (Guilford Press, 2013).
768 doi:10.2/JQUERY.MIN.JS
769 58. Chierchia, G. *et al.* Confirmatory reinforcement learning changes with age during adolescence.
770 *Dev. Sci.* e13330 (2021). doi:10.1111/desc.13330
771 59. Habicht, J., Bowler, A., Moses-Payne, M. E. & Hauser, T. U. Children are full of optimism, but
772 those rose-tinted glasses are fading – reduced learning from negative outcomes drives
773 hyperoptimism in children. (2021).
774 60. Overman, W. H. Sex differences in early childhood, adolescence, and adulthood on cognitive
775 tasks that rely on orbital prefrontal cortex. *Brain Cogn.* **55**, 134–147 (2004).
776 61. Evans, K. L. & Hampson, E. Sex-dependent effects on tasks assessing reinforcement learning
777 and interference inhibition. *Front. Psychol.* **6**, 1–10 (2015).
778 62. Shing, Y. L. *et al.* Episodic memory across the lifespan: The contributions of associative and
779 strategic components. *Neurosci. Biobehav. Rev.* **34**, 1080–1091 (2010).
780 63. Keresztes, A. *et al.* Hippocampal maturity promotes memory distinctiveness in childhood and
781 adolescence. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9212–9217 (2017).
782 64. Ghetti, S. & Bunge, S. A. Neural changes underlying the development of episodic memory
783 during middle childhood. *Dev. Cogn. Neurosci.* **2**, 381–395 (2012).
784 65. Dennis, N. A. & Cabeza, R. Age-related dedifferentiation of learning systems: An fMRI study

785 of implicit and explicit learning. *Neurobiol. Aging* **32**, 2318.e17-2318.e30 (2011).
786 66. Ferbinteanu, J. Contributions of Hippocampus and Striatum to Memory-Guided Behavior
787 Depend on Past Experience. *J. Neurosci.* **36**, 6459–6470 (2016).
788 67. White, N. M. & McDonald, R. J. Multiple Parallel Memory Systems in the Brain of the Rat.
789 *Neurobiol. Learn. Mem.* **77**, 125–184 (2002).
790 68. Shadlen, M. N. N. & Shohamy, D. Decision Making and Sequential Sampling from Memory.
791 *Neuron* **90**, 927–939 (2016).
792 69. Browning, M., Paulus, M. & Huys, Q. J. M. What is computational psychiatry good for? *Biol.*
793 *Psychiatry* **0**, (2022).
794 70. Hackman, D. A., Farah, M. J. & Meaney, M. J. Socioeconomic status and the brain:
795 mechanistic insights from human and animal research. *Nat. Rev. Neurosci. 2010 119* **11**, 651–
796 659 (2010).
797 71. Frodl, T. *et al.* Childhood stress, serotonin transporter Gene and Brain structures in major
798 depression. *Neuropsychopharmacology* **35**, 1383–1390 (2010).
799 72. Lucassen, P. J., Korosi, A., Krugers, H. J. & Oomen, C. A. *Early Life Stress- and Sex-*
800 *Dependent Effects on Hippocampal Neurogenesis*. *Stress: Neuroendocrinology and*
801 *Neurobiology* **2**, (Elsevier Inc., 2017).
802 73. Rahman, M. M., Callaghan, C. K., Kerskens, C. M., Chattarji, S. & O'Mara, S. M. Early
803 hippocampal volume loss as a marker of eventual memory deficits caused by repeated stress.
804 *Sci. Rep.* **6**, 1–15 (2016).
805 74. Levin, M. E., Haeger, J., Ong, C. W. & Twohig, M. P. An Examination of the Transdiagnostic
806 Role of Delay Discounting in Psychological Inflexibility and Mental Health Problems. *Psychol.*
807 *Rec.* **68**, 201–210 (2018).
808 75. Von Siebenthal, Z. *et al.* Decision-making impairments following insular and medial temporal
809 lobe resection for drug-resistant epilepsy. *Soc. Cogn. Affect. Neurosci.* **12**, 128–137 (2017).
810 76. Maren, S., Phan, K. L. & Liberzon, I. The contextual brain: Implications for fear conditioning,
811 extinction and psychopathology. *Nat. Rev. Neurosci.* **14**, 417–428 (2013).

812

813

Figures and Tables

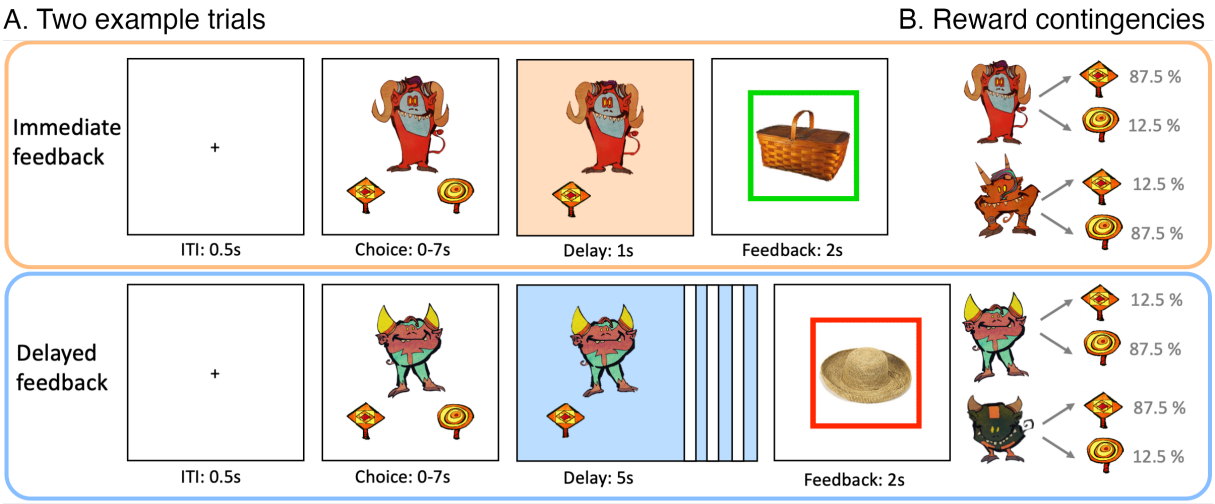

Figure 1. (A) Depiction of two example trials of immediate and delayed feedback conditions presented at wave 1. For immediate feedback (top panel), between choice response and feedback, cue and choice were presented for 1 s. At feedback, a green frame around the incidentally encoded object indicated a positive outcome, which appeared in 87.5% of the trials when selecting the squard-shaped lolli for this example cue. For delayed feedback (bottom panel), the delay phase between choice response and feedback lasted for 5 s. The red frame around the object indicated a negative outcome and appeared in 87.5% of the trials when selecting the squard-shaped lolli for this example cue. (B) For each feedback condition, two action-outcome contingencies were learned to balance a potential choice bias. With the four task versions, the cues and outcome contingencies were counterbalanced across participants.

Table 1. Descriptive behavioral results of dependent variables Accuracy (ACC, probability correct), win-stay probability (WS), lose-shift probability (LS), and reaction time (RT, in seconds), as well as mixed model fixed effects that predicted these dependent variables.

	Descriptive Results				Mixed Model Effects	
	Wave 1		Wave 2		Wave	Feedback
	Ime	Del	Ime	Del		
ACC	0.69 (0.46)	0.70 (0.46)	0.79 (0.41)	0.80 (0.40)	↑ W2	–
WS	0.81 (0.39)	0.80 (0.40)	0.88 (0.32)	0.88 (0.32)	↑ W2	–
LS	0.47 (0.50)	0.50 (0.50)	0.42 (0.49)	0.42 (0.49)	↓ W2	–
RT	2.10 (1.31)	2.07 (1.29)	1.70 (1.02)	1.67 (1.00)	↓ W2	↓ Del

Note. Mean (standard deviation) of the variables, split by wave and feedback timing, is reported in the table. Mixed model effects and their directionality (increasing ↑ or decreasing ↓) predicting the dependent variables. W2 = Wave 2, Ime = Immediate feedback, Del = Delayed feedback.

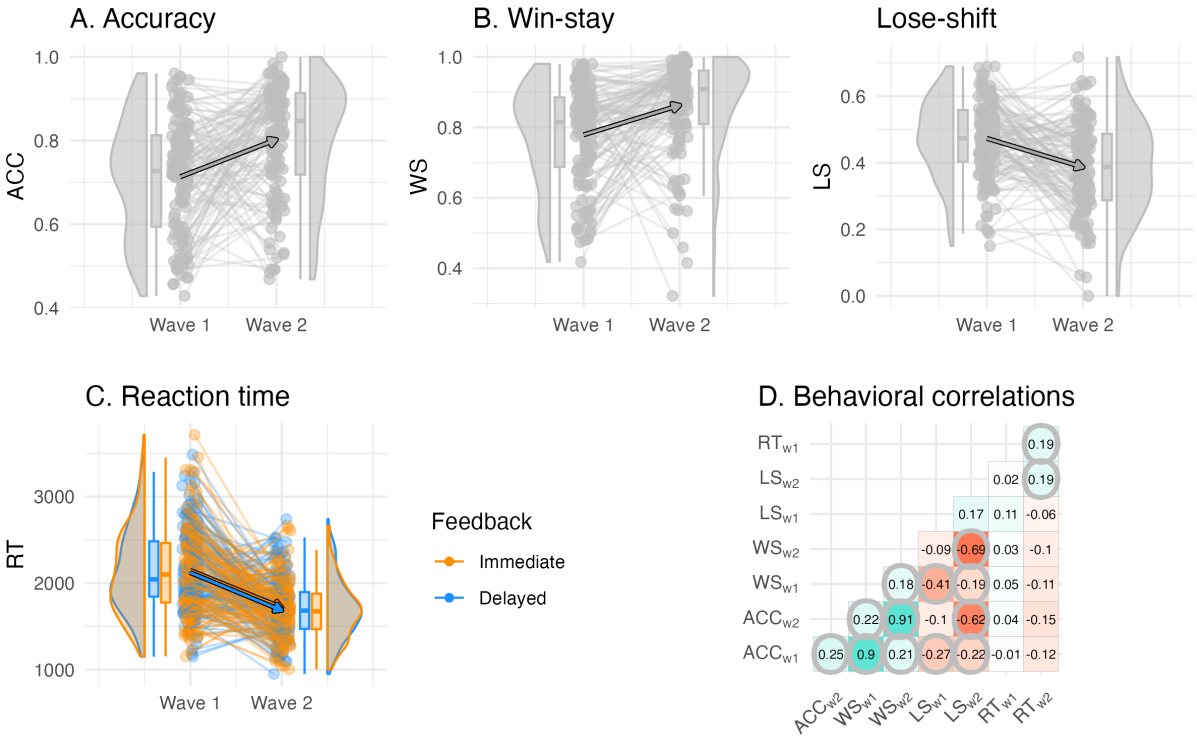

Figure 2. Individual differences in the behavioral reinforcement learning outcomes and their longitudinal change. (A) Accuracy did not differ by feedback timing and increased between waves. (B) Win-stay and lose-shift probability did not differ by feedback timing, and win-stay increased and lose-shift probability decreased between waves. (C) Reaction time differed by feedback timing, in which decisions for cues learned with delayed feedback were faster, and reaction times were faster at wave 2 compared to wave 1. (D) Correlations between behavioral outcomes reveal that learning accuracy was primarily correlated with the win-stay and lose-shift probabilities both within and between waves, but was uncorrelated to reaction time. Significant correlations are circled, *p*-values were adjusted for multiple comparisons using bonferroni correction.

Table 2. Model comparison results.

Model	Parameters	$\Delta elpd_{loo}$ [SE]	$\Sigma elpd_{loo}$ [mean]	*pseudo-BMA+*
Step 1: heuristic strategy models and value-based learning model				
vbm_1	1α, 1τ	0 [0]	-15154.9 [-0.45]	1
ws	$1\tau_{ws}$	-1327.7 [159.5]	-16482.7 [-0.49]	<0.01
$wsls$	$1\tau_{wsls}$	-4247.3 [284.8]	-19402.3 [-0.58]	0
Step 2: value-based learning models				
vbm_3	**1α, 2τ**	**0 [0]**	**-15045.3 [-0.45]**	**0.73**
vbm_7	1α, 2ρ	-2.93 [2.92]	-15048.2 [-0.45]	0.24
vbm_6	2α, 1ρ	-24.34 [8.85]	-15069.6 [-0.45]	<0.01
vbm_8	2α, 2ρ	-29.71 [15.95]	-15075.0 [-0.45]	0.02
vbm_4	2α, 2τ	-43.34[14.89]	-15088.6 [-0.45]	<0.01
vbm_2	2α, 1τ	-46.45 [13.97]	-15091.7 [-0.45]	<0.01
vbm_5	1α, 1ρ	-59.01 [7.59]	-15104.3 [-0.45]	<0.01
vbm_1	1α, 1τ	-109.63 [11.98]	-15154.9 [-0.45]	<0.01

Note. Model = heuristic (ws, $wsls$) and value-based models (vbm_{1-8}) that were compared against each other. Parameters = corresponding model parameters learning rate α, inverse temperature τ and outcome sensitivity ρ. $\Delta elpd_{loo}[SE]$ = difference in the Bayesian leave-one-out cross-validation estimate of the expected log pointwise predictive density relative to the winning model and its standard errors. $\Sigma elpd_{loo}[mean]$ = sum of expected log pointwise predictive density of all 33,460 trials, including all participants and waves, and trial mean. *Pseudo-BMA+* = model weight for relative model evidence using Bayesian model averaging stabilized by Bayesian bootstrap using 100,000 iterations.

Table 3. Description of model parameters from the winning value-based model vbm_3.

	Wave 1					Wave 2				
	α	τ_{Ime}	τ_{Del}	ls_{Ime}	ls_{Del}	α	τ_{Ime}	τ_{Del}	ls_{Ime}	ls_{Del}
Mean	0.02	14.6	14.8	0.73	0.73	0.05	16.2	16.5	0.82	0.82
SD	0.02	2.04	2.37	0.12	0.13	0.04	2.37	2.21	0.13	0.13
Min	<0.01	6.73	5.25	0.53	0.53	<0.01	4.37	6.85	0.53	0.53
Max	0.09	17.5	17.9	0.94	0.94	0.22	18.6	18.7	0.96	0.96

Note. α = *learning rate across feedback timing,* τ_{Ime}/ ls_{Ime} = inverse temperature and learning score for immediate feedback, τ_{Del}/ ls_{Del} = inverse temperature and learning score for delayed feedback.

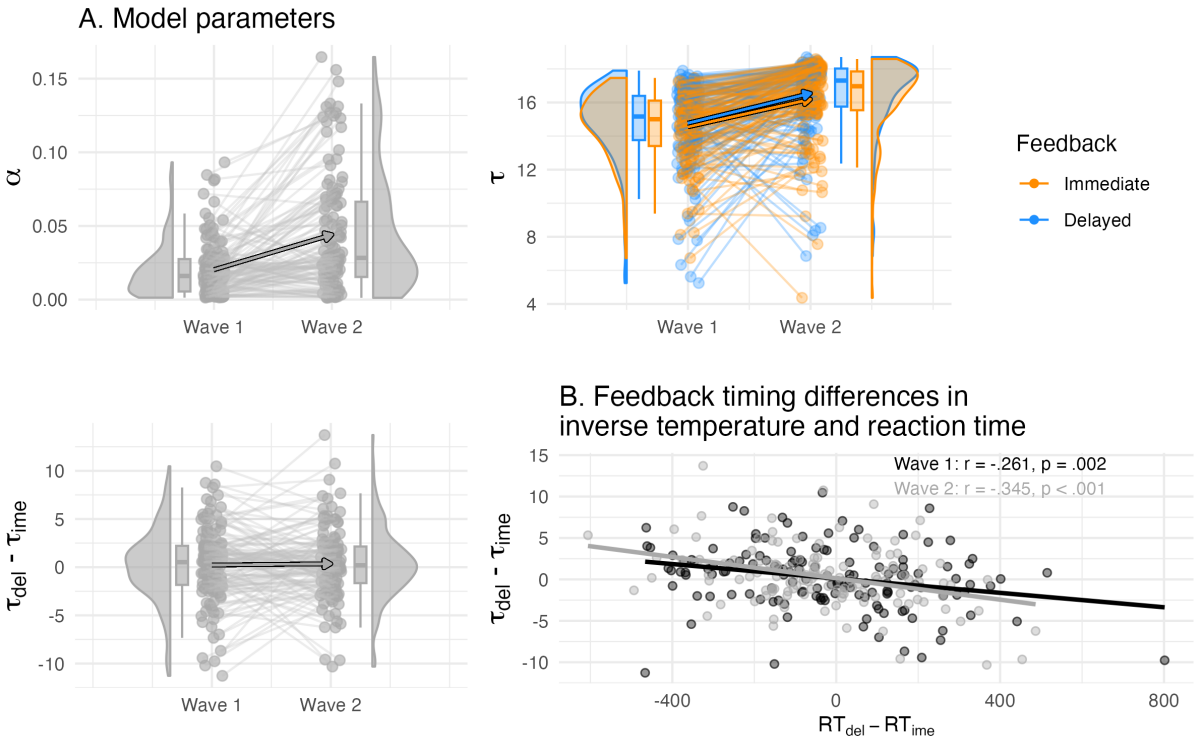

Figure 3. (A) Individual differences in the learning rate and inverse temperature of the winning model and their longitudinal change. The inverse temperature τ but not learning rate α was separated by feedback timing, and both increased between waves in their values (top panel). The condition difference in the inverse temperature did not differ on average, but showed individual differences (bottom left panel). (B) The condition differences in the inverse temperature correlated with reaction time, i.e., higher delayed compared to immediate inverse temperature was related to faster delayed compared to immediate reaction time.

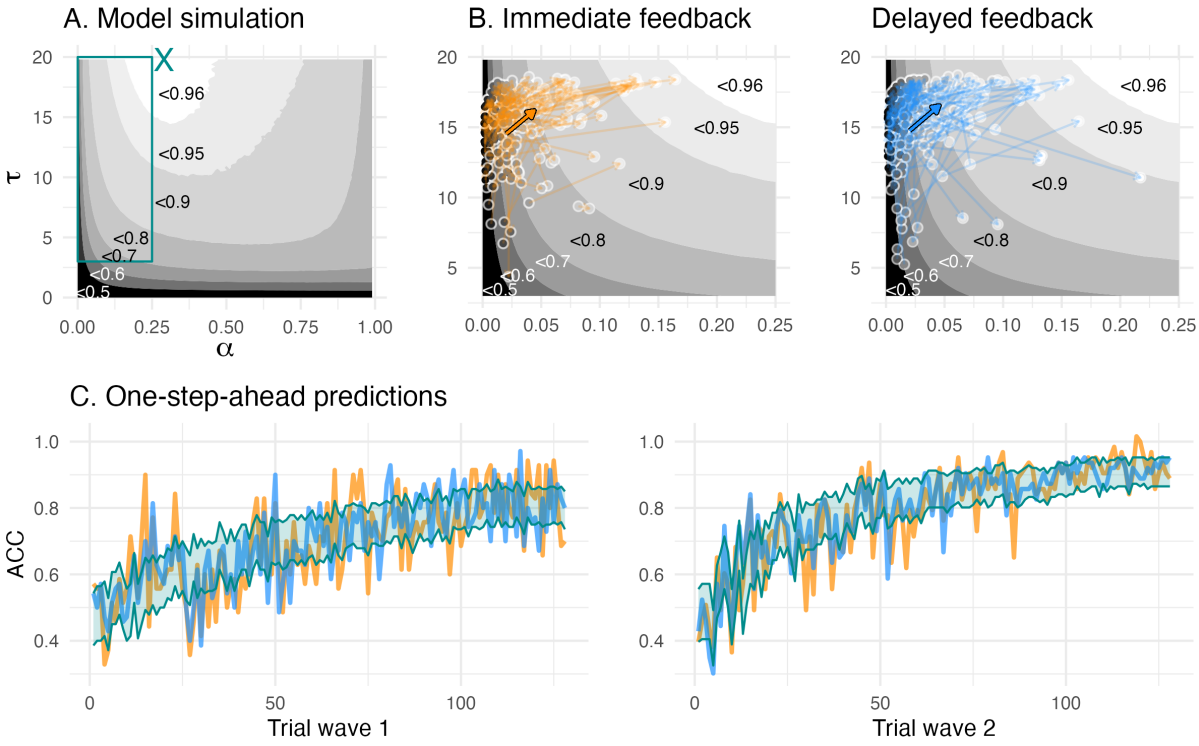

Figure 4. (A) The model simulation depicts parameter combinations and simulation-based average learning scores. The cyan "X" in the middle top depicts the optimal parameter combination where average learning scores were at 96.5 %, and the cyan rectangle depicts the space of the fitted parameter combinations, (B) Enlarged view of the space of fitted parameter combinations. The colored arrows depict mean change (bold arrow) and individual change (transparent arrows) of the fitted parameters. The greyscale gradient-filled dots, that are connected by the arrows, depict the individual learning score, while the the greyscale gradient in the background depicts the simulated average learning score. The mean change reveals an overall change towards the higher, i.e., more optimal, learning scores. (C) One-step-ahead posterior predictions of the winning model for each wave. The colored lines depict averaged trial-by-trial task behavior for each feedback condition, and a cyan ribbon indicates the 95% highest density interval of the one-step-ahead prediction using the entire posterior distribution.

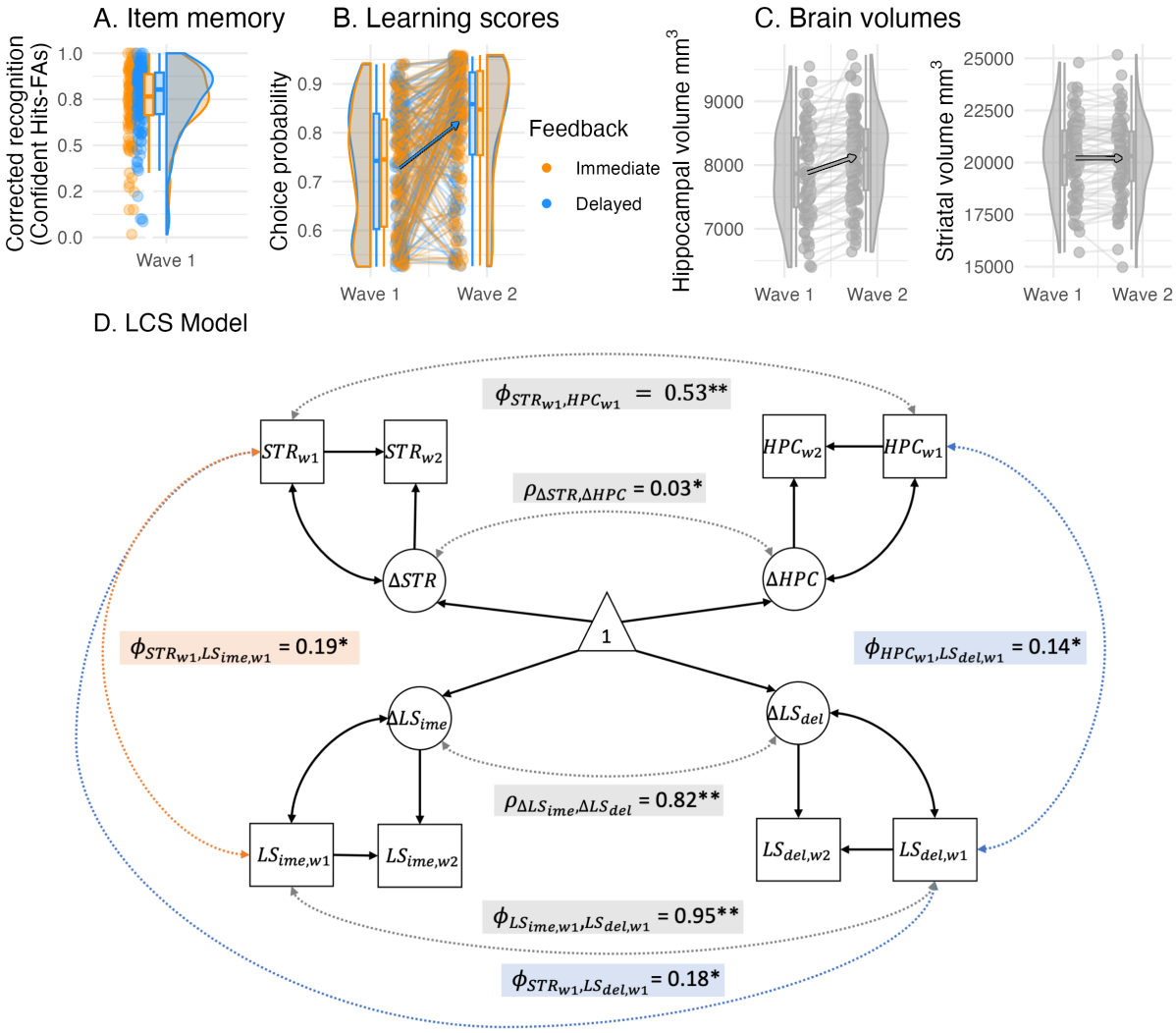

Figure 5. (A) Recognition memory (corrected recognition = hits - false alarms) for objects presented during delayed feedback was only enhanced at trend. (B) Learning scores depicted here were used in the LCS analyses. Learning scores were the model-derived choice probability of the contingent choice using fitted posterior parameters. (C) Hippocampal and striatal volumes increased between waves, while hippocampal volume increased most. (D) A four-variate latent change score (LCS) model that included striatal and hippocampal volumes as well as immediate and delayed learning scores. Depicted are significant paths cross-domain (brain-cognition, dashed lines) and within-domain (brain or cognition, solid lines), other paths are omitted for visual clarity and are summarized in Table 4. Depicted brain-cognition links included $\phi_{STR_{w1},LS_{ime,w1}}$ (covariance between striatal volume and immediate learning score at wave 1), as well as $\phi_{HPC_{w1},LS_{del,w1}}$ and $\phi_{STR_{w1},LS_{del,w1}}$ (covariances between hippocampal and striatal volumes and delayed learning score at wave 1). Brain links included $\phi_{STR_{w1},HPC_{w1}}$ and $\rho_{\Delta STR,\Delta HPC}$ (wave 1 covariance and change-change covariance), and similarly, cognition links included $\phi_{LS_{ime,w1},LS_{del,w1}}$ and $\rho_{\Delta LS_{ime},\Delta LS_{del}}$. Covariates included age, sex and estimated total intracranial volume. ** denotes significance at $\alpha < .001$, * at $\alpha < .05$.

Table 4. Parameter estimates of a four-variate latent change score model that includes brain (striatal and hippocampal volume) and cognition domains (immediate and delayed learning score)

	STR	LS_{ime}	HPC	LS_{del}
Model fit: χ^2 = 15.4, *df* = 27, *CFI* = 1, *RMSEA* (*CI*) = 0 (0-0.01), *SRMR* = 0.045				
Mean change Δ	0.06* (0.03)	0.76** (0.08)	0.38** (0.04)	0.75** (0.08)
wave 1 variance σ	fixed to 1	fixed to 1	fixed to 1	fixed to 1
change variance σ_Δ	0.07** (0.01)	0.88** (0.10)	0.18* (0.07)	0.83** (0.10)
Intercept-change regression β	-0.04 (0.04)	-0.83* (0.29)	-0.16* (0.06)	-0.73* (0.27)
Wave 1 covariates				
age onto Intercept ϕ	0.19 (0.10)	-0.05 (0.08)	0.29* (0.08)	0.08 (0.08)
sex onto Intercept ϕ	-0.42** (0.07)	-0.14 (0.07)	-0.47** (0.07)	-0.11 (0.07)
eTIV onto Intercept ϕ	0.68** (0.05)	–	0.70** (0.05)	–
Brain-cognition links (cross-domain)	$STR-LS_{ime}$	$STR-LS_{del}$	$HPC-LS_{ime}$	$HPC-LS_{del}$
wave 1 covariation ϕ	**0.19* (0.07)**	**0.18* (0.07)**	0.12 (0.07)	**0.14* (0.07)**
change-change covariance ρ	<0.01 (0.03)	<0.01 (0.03)	-0.06 (0.05)	-0.07 (0.05)
wave 1 brain onto cognition change γ	0.25 (0.13)	0.22 (0.12)	0.05 (0.11)	0.06 (0.10)
wave 1 cognition onto brain change γ	-0.19 (0.13)	0.21 (0.13)	0.05 (0.10)	<0.01 (0.10)
Brain links (within-domain)	$STR-HPC$			
wave 1 covariation ϕ	**0.53** (0.07)**			
change-change covariance ρ	**0.03* (0.01)**			
wave 1 striatum onto hippocampal change γ	0.06 (0.05)			
wave 1 hippocampus onto striatal change γ	0.02 (0.03)			
Cognition links (within-domain)	$LS_{ime}-LS_{del}$			
wave 1 covariation ϕ	**0.95** (0.10)**			
change-change covariance ρ	**0.82** (0.10)**			
wave 1 LS_{ime} into LS_{del} change γ	-0.07 (0.27)			
wave 1 LS_{del} into LS_{ime} change γ	0.06 (0.28)			

Parameter estimates in bold are the paths of interest depicted in Figure 5D. Standard errors are shown in parentheses. eTIV = estimated total intracranial volume. ** denotes significance at $\alpha < .001$, * at $\alpha < .05$. sex coded as 1 = girls, -1 = boys.