

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

3 Johannes Falck¹, Lei Zhang^{2,3,4}, Laurel Raffington⁵, Johannes J. Mohn^{6,7}, Jochen Triesch⁸, Christine
4 Heim^{6,9} & Yee Lee Shing¹

5
6 ¹*Department of Psychology, Goethe University Frankfurt, 60629 Frankfurt am Main, Germany*

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

9 ³*Centre for Human Brain Health, School of Psychology, University of Birmingham, Birmingham B15
10 2TT, UK*

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

13 ⁵*Center for Lifespan Psychology, Max Planck Institute for Human Development, 14195 Berlin,
14 Germany*

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

16 ⁷*Max Planck School of Cognition, Max Planck Institute for Human Cognitive and Brain Sciences, 04103
17 Leipzig, Germany*

18 ⁸*Frankfurt Institute for Advanced Studies (FIAS), 60439 Frankfurt am Main, Germany*

19 ⁹*Center for Safe & Healthy Children, The Pennsylvania State University, State College, PA 16802, USA*

20

21

22

Abstract

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

36

Introduction

37

38 As children enter school during middle childhood, they must learn to act appropriately in new situations
39 through feedback. For example, children receive positive feedback when raising their hand before
40 speaking in class, which reinforces them to repeat the same action in the future. Reinforcement learning
41 (RL)¹ provides a useful mechanistic framework to describe such feedback-driven value-based learning
42 and decision-making. RL models allow to explicitly test for the influence of separate components
43 during value-based learning, such as model-free and model-based learning², social and non-social
44 learning^{3,4}, or the contribution of different memory systems⁵⁻⁷.

45 The memory systems account is a theoretical framework that proposes that different types of
46 memory are supported by distinct neural systems in the brain. Specifically, this account suggests that
47 there are two memory systems: a hippocampal-dependent system and a striatal-dependent system. These
48 systems modulate memory and value-based learning, and their interactive development has been of
49 particular interest to developmental research^{8,9}. The hippocampal-dependent memory system has been
50 shown to contribute to episodic memory during reinforcement learning and is more engaged during
51 feedback that is presented with a delay^{6,10,11}, as opposed to the striatal-dependent memory system, which
52 is more engaged after immediate feedback and supports habitual memory^{5,12-14}. Specifically,
53 hippocampal activation was greater during delayed feedback than during immediate feedback, whereas
54 striatal activation was greater during immediate feedback than during delayed feedback⁵. The
55 engagement of the hippocampus during delayed feedback was further supported by enhanced episodic
56 memory for incidentally presented objects compared to objects presented with immediate feedback.
57 Taken together, these studies suggest that feedback timing modulates the engagement of the
58 hippocampal and striatal memory systems during value-based learning in adults. Given the differential
59 developmental trajectories of these systems and the impact the systems have on reinforcement learning
60 and memory, it is important to understand whether children would show similar feedback timing
61 modulations as previously shown in adults. In addition, whether such feedback timing modulation
62 changes over time remains largely unexplored. To this end, in this study, we examined the contributions
63 of hippocampal and striatal structural volumes during the longitudinal development of reinforcement
64 learning across two years in 6-to-7-year-old children.

65 Reinforcement learning behavior modulated by feedback timing can be modeled
66 computationally using at least three parameters that reflect feedback-based learning and decision-
67 making. For feedback-based learning, a learning rate parameter determines the extent to which the
68 reward prediction error, defined as the difference between the received reward and the expected reward,
69 influences the update of the future choice values. A higher learning rate emphasizes recent outcomes,
70 whereas a lower learning rate reflects learning integrated over a longer outcome history¹⁵. Value updates
71 may further depend on an outcome sensitivity parameter that scales the individual magnitude of received
72 rewards. Finally, in decision-making, the inverse temperature parameter plays a key role in determining
73 the tendency to select the more valuable choice and quantifies choice stochasticity. A higher inverse

74 temperature reflects more value-guided, deterministic choice behavior compared to a lower inverse
75 temperature reflecting more random choices. Learning rates and inverse temperature have been studied
76 extensively across development, mainly with cross-sectional studies showing mixed findings regarding
77 their age gradients¹⁶. One study reported lower learning rates in children compared to adolescents¹⁷,
78 while other studies found no differences^{18,19} or even higher learning rates in children^{8,20}. Developmental
79 differences regarding the inverse temperature parameter are slightly more consistent, with studies
80 reporting no differences^{8,21-23} or higher inverse temperature with age that suggests that behavior is
81 increasingly value-guided and less explorative^{17-19,24}. To the best of our knowledge, outcome sensitivity
82 has not been modeled computationally across development. However, studies that linked striatal reward
83 activation to self-reported reward sensitivity showed increasing sensitivity from childhood to
84 adolescence^{25,26}.

85 In general, the inconsistencies regarding developmental differences in parameters may be due
86 to their dependency on model and task properties²⁷, which could be reconciled by comparing
87 developmental changes to simulation-based optimal learning¹⁵. Such comparisons acknowledge that
88 optimal parameter values vary depending on the context, and it has been suggested that humans develop
89 towards more optimal parameter values from childhood into adulthood¹⁶. Importantly, to our knowledge
90 previous reinforcement learning studies with children were cross-sectional, and only two studies
91 investigated children under 8 years of age^{17,28}. Cross-sectional studies, in which developmental change
92 is inferred as a between-subject factor, do not capture the dynamics in middle childhood if individual
93 differences are large, whereas longitudinal studies test development as a within-subject factor, which is
94 crucial for uncovering change across time. Thus, longitudinal changes in reinforcement learning in
95 middle childhood as well as their putative striatal and hippocampal associations remain unknown. To
96 this end, learning rates, outcome sensitivity and inverse temperature are relevant computational
97 parameters to study longitudinal changes in striatal and hippocampal systems during value-based
98 learning.

99 Striatal and hippocampal contributions to reinforcement learning during middle childhood may
100 differ as these brain regions undergo major developmental changes. Whereas earlier structural studies
101 with relatively small sample sizes showed large developmental variability and a tendency for an earlier
102 volume peak in the striatum than in the hippocampus²⁹⁻³⁵, a recent cross-sectional large-scale study was
103 able to contrast striatal and hippocampal trajectories with greater granularity³⁶. These data showed
104 striatal volume peaks in the first decade which then declined throughout later developmental periods,
105 whereas hippocampal volume showed a more protracted inverted-U-shaped trajectory that peaked in
106 adolescence. Based on these structural findings, striatal and hippocampal systems are expected to
107 develop functionally at different rates³⁷, with habit memory depending on the earlier developing striatum
108 and episodic memory depending on the later developing hippocampus³⁸. A direct investigation of the
109 longitudinal development of both memory systems in childhood would shed light on whether the
110 memory systems show a differential engagement similar to that of adults⁵. Such knowledge could be

111 useful to structure learning processes according to the developmental status. For example, children's
112 ability to learn from delayed feedback may depend on how well their hippocampus has developed. In
113 the same study sample, we previously reported that children's hippocampal volume was related to their
114 family's income level³⁹. Additionally, previous research has shown that stress can reduce the
115 effectiveness of the hippocampal-dependent memory system¹¹. This suggests that environmental factors
116 such as income and stress may play a role in shaping how well children learn from delayed feedback,
117 particularly through their impact on hippocampal development. By identifying the specific
118 environmental factors that impact children's learning and brain development, we can identify risk groups
119 and tailor interventions to ameliorate adverse effects.

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

133

134

Methods

135

136 Participants

137 Children and their parents took part in 2 waves of data collection with an interval of about 2 years (*mean*
138 *= 2.07*, *SD = 0.17*, *range = 1.69 – 2.68*). The inclusion criteria for wave 1 were children attending first
139 or second grade, no psychiatric or physical health disorders, at least one parent speaking fluent German,
140 and born full-term (≥ 37 weeks of gestation). At wave 1, 142 children (46% female, age *mean = 7.19*,
141 *SD = 0.46*, *Range = 6.07 - 7.98*) and their parents or caregivers participated in the study. 141 children
142 completed the probabilistic learning task, one child was later excluded due to technical problems during
143 the task, hence 140 were included in the analysis. A subgroup of 90 children (49% female, 100% right-
144 handed), who was randomly selected, completed magnetic resonance imaging (MRI) scanning at wave
145 1, and 82 of them contributed to structural data after removing scans with excessive movement. At wave
146 2, 127 children (46% female, age *mean = 9.25*, *SD = 0.45*, *Range = 8.30 - 10.2*) continued taking part in
147 the study, while families of the remaining children were unable to be contacted or decided not to return

148 to the study. 126 children at wave 2 completed the reinforcement learning task and were included in the
149 analysis. All children at wave 2 were invited for MRI scanning, and 104 of them completed scanning
150 (45% female, 92% right-handed). 99 children contributed to structural data, after removing scans with
151 excessive movement. In total, 73 children contributed to the longitudinal MRI data and 126 children
152 contributed to the longitudinal learning data. As previously reported for this study sample, we found no
153 systematic bias due to wave 2 dropout³⁹.

154

155 Procedure

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

163

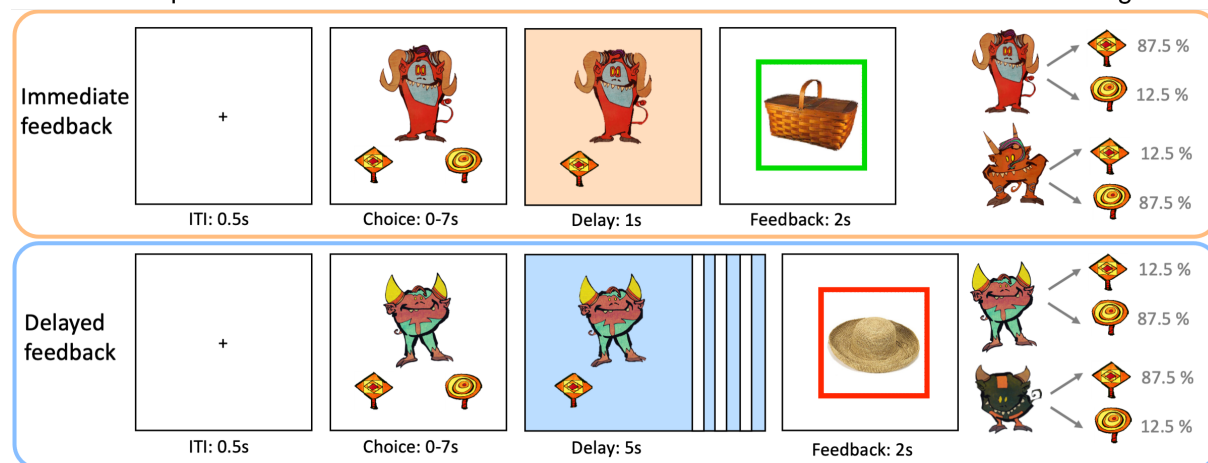
164 Measures

165 *Reinforcement learning task.* Children completed an adapted reinforcement learning task⁵ in which they
166 learned the preferred associations between four cues (cartoon characters) and two choices (round-shaped
167 or square-shaped lolli) through probabilistic feedback (87.5 % contingent and 12.5 % non-contingent
168 reward probability). In each trial, after an initial inter-trial interval of 0.5 s, a cue and its choice options
169 were presented for up to 7 s until the child made a choice (Figure 1, choice phase). In the delay phase,
170 we manipulated feedback timing. For two cues, the selected choice remained visible for 1 s (immediate
171 feedback condition), whereas for the other two cue characters, it remained visible for 5 s before feedback
172 was given (delayed feedback condition). A final feedback phase of 2 s indicated a reward by a green
173 frame, and a punishment by a red frame. Inside each frame, a unique object picture was shown, which
174 was incidentally encoded and irrelevant to the task. The children were instructed to pay attention to the
175 feedback indicated by the frame color. In an initial practice phase of 32 trials, the child practiced the task
176 with a fifth cartoon character not included in the actual task to avoid practice effects. The experimenter
177 instructed them to select the choice that was most likely to give them a reward. The Experimenter
178 checked whether the child learned the more rewarded choice during practice and let it repeat the practice
179 task otherwise to ensure understanding of the task. In the actual task, 128 trials were presented in four
180 blocks and with small breaks in between. Cues were presented in a mixed, pseudo-randomized order. A
181 total of 64 unique objects were shown in the feedback phase, each one twice within the same feedback
182 condition. In both delay phases, contingent choice and choice location remained the same for each cue
183 within the task, but were balanced across participants by using four different task versions. At wave 2,
184 four new cues replaced the previous ones to rule out memory effects.

185 *Object recognition test.* At wave 1, children were additionally tested for recognition memory on the
186 object pictures that were incidentally encoded during reinforcement learning. A total of 80 objects (48
187 old objects and 32 new objects) were presented in randomized order. The 48 old objects (24 for each
188 feedback condition) were selected from the 64 old objects shown during learning based on two lists to
189 balance the shown and omitted old objects across task versions. Each old object was shown twice during
190 learning, but if the child failed to respond during learning, no feedback or object was shown in the trial,
191 so some objects only appeared once. These objects were excluded at the individual level (individually
192 missing object *mean* = 2.71). At recognition, children had 4 response options ('old sure', 'old unsure',
193 'new unsure', 'new sure') with up to 7 s to respond. The children answered verbally, and the
194 experimenter entered their response. At wave 2, this test was excluded due to time constraints.
195

A. Two example trials

B. Reward contingencies



196
197 Figure 1. (A) Depiction of two example trials of immediate and delayed feedback conditions presented
198 at wave 1. For immediate feedback (top panel), between choice response and feedback, cue and choice
199 were presented for 1 s. At feedback, a green frame around the incidentally encoded object indicated a
200 positive outcome, which appeared in 87.5% of the trials when selecting the sward-shaped lolly for this
201 example cue. For delayed feedback (bottom panel), the delay phase between choice response and
202 feedback lasted for 5 s. The red frame around the object indicated a negative outcome and appeared in
203 87.5% of the trials when selecting the sward-shaped lolly for this example cue. (B) For each feedback
204 condition, two action-outcome contingencies were learned to balance a potential choice bias. With the
205 four task versions, the cues and outcome contingencies were counterbalanced across participants.
206

207 *Brain volume.* Structural MRI images were acquired on a Siemens Magnetom TrioTim syngo 3 Tesla
208 scanner with a 12-channel head coil (Siemens Medical AG, Erlangen, Germany) using a 3D T1-
209 weighted Magnetization Prepared Rapid Gradient Echo (MPRAGE) sequence (192 slices; field of view
210 = 256 mm; voxel size = 1 mm³; TR = 2500 ms; TE = 3.69 ms; flip angle = 7°; TI = 1100 ms).
211 Volumetric segmentation was performed using the Freesurfer 6.0.0 image analysis suite⁴¹. Previous
212 studies suggested that software tools based on adult brain templates provide inaccurate segmentation for

213 pediatric samples, which can be improved through the use of study-specific template brains^{42,43}. Thus,
214 we created two study-specific template brains (one for each wave) using Freesurfer's
215 "make_average_subject" command. This pipeline utilized the default adult template brain registrations
216 of the "recon-all-all" command to average surfaces, curvatures, and volumes from all subjects into a
217 study-specific template brain. All subjects were then re-registered to this study-specific template brain
218 to improve segmentation accuracy. Segmented images were manually inspected for accuracy and 8 cases
219 at wave 1 and 5 cases at wave 2 were excluded for inaccurate or failed registration due to excessive
220 motion.

221

222 Data analysis

223

224 *Behavioral learning performance.* Differences in learning accuracy, win-stay probability, lose-shift
225 probability and reaction time with the predictors feedback timing (immediate, delayed), wave (1, 2),
226 wave 1 age, and sex (girls, boys) were tested using generalized linear mixed models (GLMM) with the
227 R package lme4⁴⁴. Learning accuracy was defined as the probability to choose the more rewarding
228 option, while win-stay and lose-shift refer to the probabilities of staying with the previously chosen
229 option after a reward and switching to the alternative choice after not receiving a reward, respectively.
230 All reported models included random slopes for within-subject factors feedback timing and wave (see
231 Supplementary Material 2 for the model structure). We systematically tested main effects and
232 interactions between the predictors and their interaction had to statistically improve the predictive ability
233 of the model to be included in the final reported model. All predictor variables were grand-mean-
234 centered to interpret the interaction effects independent from other predictors.

235

236 *Reinforcement learning models.* We compared the learning models of basic heuristic strategies and
237 value-based learning to determine the model that could best capture children's trial-by-trial learning
238 behavior. For heuristic strategies, we considered models that reflected a Win-stay-lose-shift (wsls) or a
239 Win-stay (ws) strategy. Win-stay is a heuristic strategy in which the same action is repeated if it leads
240 to a positive outcome in the previous trial, and Win-stay-lose-shift additionally switches to a different
241 action if the previous outcome is negative. The models quantified the learning behavior for each
242 individual I for each cue c and trial t . The heuristic models consisted of a weight w that reflected the
243 strategy use. In the case of reward $r = 1$, w was equal to 1 for the chosen option (eg. choice A), and 0
244 for the unchosen option (e.g. choice B), thus maximizing win-stay, i.e., choosing A at the subsequent trial
245 $t + 1$:

$$246 \quad w_{i,c,t+1,A|r=1} = 1 \text{ and } w_{i,c,t+1,B|r=1} = 0 \quad (1)$$

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

$$249 \quad w_{i,c,t+1,A|r=0} = 0; w_{i,c,t+1,B|r=0} = 1 \quad (2)$$

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

$$253 \quad p(A) = \frac{\exp^{w_{i,c,t,A} \tau_{wsls_i}}}{\exp^{w_{i,c,t,A} \tau_{wsls_i}} + \exp^{w_{i,c,t,B} \tau_{wsls_i}}} \quad (3)$$

254 Thus, a higher probability of strategy use was reflected by a larger value of τ_{wsls} or τ_{ws} .

255 For value-based learning, we considered a Rescorla-Wagner model and several variants based on our
256 theoretical conceptions. The baseline value-based model vbm_1 updated the value v of the selected choice
257 (A or B) for the next trial t . This value update was determined by calculating the difference between the
258 received reward r and the expected value v of the selected choice, which was the reward prediction error.

259 The value update was further scaled by a learning rate α ($0 < \alpha < 1$):

$$260 \quad v_{i,c,t+1,A} = v_{i,c,t,A} + \alpha_i (r_{i,c,t} - v_{i,c,t,A}) \quad (4)$$

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

$$263 \quad v_{i,c,t+1,A} = v_{i,c,t,A} + \alpha_i (\rho_i * r_{i,c,t} - v_{i,c,t,A}) \quad (5)$$

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

$$266 \quad p(A) = \frac{\exp^{v_{i,c,t,A} \tau_i}}{\exp^{v_{i,c,t,A} \tau_i} + \exp^{v_{i,c,t,B} \tau_i}} \quad (6)$$

267 Note, however, that outcome sensitivity and inverse temperature are difficult to fit simultaneously due
268 to non-identifiability issues⁴⁵. Therefore, models including outcome sensitivity fixed the inverse
269 temperature at 1 (outcome sensitivity model family), and models with the inverse temperature in turn
270 fixed outcome sensitivity at 1 (inverse temperature model family). Each model family consisted of 4
271 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
272 each parameter was either separated by feedback timing or kept as a single parameter across feedback
273 conditions. Our baseline value-based model vbm_1 included a single learning rate and a single inverse
274 temperature ($1\alpha 1\tau$).

275

276 *Parameter estimation.* All choice data were fitted in a hierarchical Bayesian analysis using the Stan
277 language in R^{46,47} adopted from the hBayesDM package⁴⁸. Posterior parameter distributions were
278 estimated using Markov chain Monte Carlo (MCMC) sampling running 4 chains each with 3,000
279 iterations, using the first half of the chain as warmup, and group-level parameters and individual-level
280 parameters were estimated simultaneously. The hierarchical Bayesian approach provides more stable
281 and reliable parameter estimates as opposed to point-estimation approaches like maximum likelihood
282 estimation⁴⁹. Each model fit both wave 1 and wave 2 data at once, considering the correlation structure
283 of the same parameter across waves, to account for within-subject dependency using the Cholesky
284 decomposition. The Cholesky decomposition used a Lewandowski-Kurowicka-Joe prior of 2, and all

285 other group-level parameters had a prior normal distribution, Normal (0, 0.5). Non-response trials (wave
286 1 = 2.41%, wave 2 = 0.97% on average) were excluded in advance.

287

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

300

301 *Model selection and validation.* We conducted a 2-step sequential procedure for the model development
302 and model selection. As a first step, we compared model evidence for the baseline value-based model
303 that does not separate learning rate and inverse temperature by feedback timing ($vbm_1:1\alpha, 1\tau$) to the
304 non-value-based, heuristic strategy models that reflect Win-stay or Win-stay-lose-shift strategy behavior
305 ($ws, wsls$). As a second step, we compared model evidence for 8 value-based model variants, 4 of the
306 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
307 family with learning rate and outcome sensitivity ($1\alpha1\rho, 2\alpha1\rho, 1\alpha2\rho, 2\alpha2\rho$). This allowed us to
308 compare whether children showed separable effects of feedback timing on one of the model parameters.
309 We compared the model fit using Bayesian leave-one-out cross-validation and obtained the expected
310 log pointwise predictive density ($elpd_{loo}$) using the R package `loo`⁵⁰. We further computed the model
311 weights (*Pseudo-BMA+*) using Pseudo Bayesian model averaging stabilized by Bayesian bootstrap with
312 100,000 iterations⁵¹. To validate our models, we estimated predictive accuracy by comparing one-step-
313 ahead model predictions with the choice data^{15,52}. We performed parameter recovery for the winning
314 model and model recovery by comparing it to a set of models used during model comparison
315 (Supplementary Material 1)⁵³.

316

317 *Episodic memory at wave 1*

318 We predicted the individual corrected recognition memory (hits-false alarms) by feedback condition in
319 a linear mixed effects model using the R package `lme4`⁴⁴. A total of 140 children completed the
320 recognition memory test and 138 were included in the analysis, with two being excluded due to negative

321 corrected recognition memory value (i.e., poor recognition memory). Age and sex were controlled for
322 as covariates.

323

324 *Longitudinal brain-cognition links*

325 We used latent change score (LCS) models to examine the longitudinal relationships between brain and
326 learning score measures. LCS models are longitudinal structural equation models that have been widely
327 applied to estimate developmental changes and coupling effects across domains such as the brain and
328 cognition^{54,55}. LCS models allow the definition of specific paths between multiple variables to test
329 explicit hypotheses and estimate latent change from the observed variables that account for measurement
330 error and increase testing power⁵⁶. We compiled univariate LCS models for each variable separately
331 (learning scores and brain volumes) to examine whether there was significant individual variance and
332 change, which could be related within a multivariate LCS model as a next step. Model fit had to be at
333 least acceptable, with a comparative fit index (*CFI*) > 0.95, standardized root mean square residual
334 (*SRMR*) < .08 and root mean square error of approximation (*RMSEA*) < .08⁵⁷. Age and sex were included
335 as covariates at wave 1, as well as the estimated total intracranial volume (eTIV) when brain volume
336 was included in the model. Multivariate LCS models allow to estimate meaningful brain-cognition
337 relationships: a wave 1 covariance between brain and cognition, brain predicting change onto cognition,
338 or vice versa, and a covariance in both brain and cognition change scores (wave 1 to wave 2). Before
339 compiling the variables into an LCS model, they were checked for outliers ± 4 *SD* around the mean. We
340 identified one outlier for the learning rate at wave 2, which was removed for the explorative LCS model
341 that included model parameters. There were no further outliers in other cognitive variables or brain
342 volumes. Continuous variables were standardized to the wave 1 measure so that wave 2 values represent
343 the change from wave 1, sex was contrast-coded (girls = 1, boys = -1).

344

345

Results

346

347 Behavioral results

348

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

356

357 *Children's learning improved between waves.* With the complete dataset, we found that increased
 358 learning accuracy (i.e., the probability of choosing the more rewarding option) was predicted at wave 2
 359 compared to wave 1, but there were no differences in accuracy by feedback timing ($\beta_{wave=2} = .550$, SE
 360 $= .061$, $z = 8.97$, $p < .001$, $\beta_{feedback=delayed} = .013$, $SE = .024$, $z = 0.54$, $p = .590$). Furthermore, win-
 361 stay probability increased and lose-shift probability decreased longitudinally, again without differences
 362 by feedback timing (WS: $\beta_{wave=2} = .586$, $SE = .071$, $z = 8.22$, $p < .001$, LS: $\beta_{wave=2} = -.586$, $SE = .071$,
 363 $z = -8.22$, $p < .001$). Reaction times were faster at wave 2 compared to wave 1, and they were faster for
 364 delayed compared to immediate feedback trials ($\beta_{wave=2} = -218$, $SE = 22.7$, $t = -9.61$, $p < .001$,
 365 $\beta_{feedback=delayed} = -14.0$, $SE = 6.61$, $t = -2.12$, $p = .036$). To summarize, children, on average,
 366 improved their accuracy over 2 years, while the win-stay probability increased and the lose-shift
 367 probability decreased between waves. Children were able to respond faster to cues paired with delayed
 368 feedback compared to cues paired with immediate feedback, and they became faster in their decision-
 369 making across waves (see mixed model effects overview in Table 1). Of note, reaction times were
 370 largely uncorrelated with accuracy and switching behavior (win-stay, lose-shift), while accuracy and
 371 switching behavior showed significant correlations at both waves (Figure 2D).

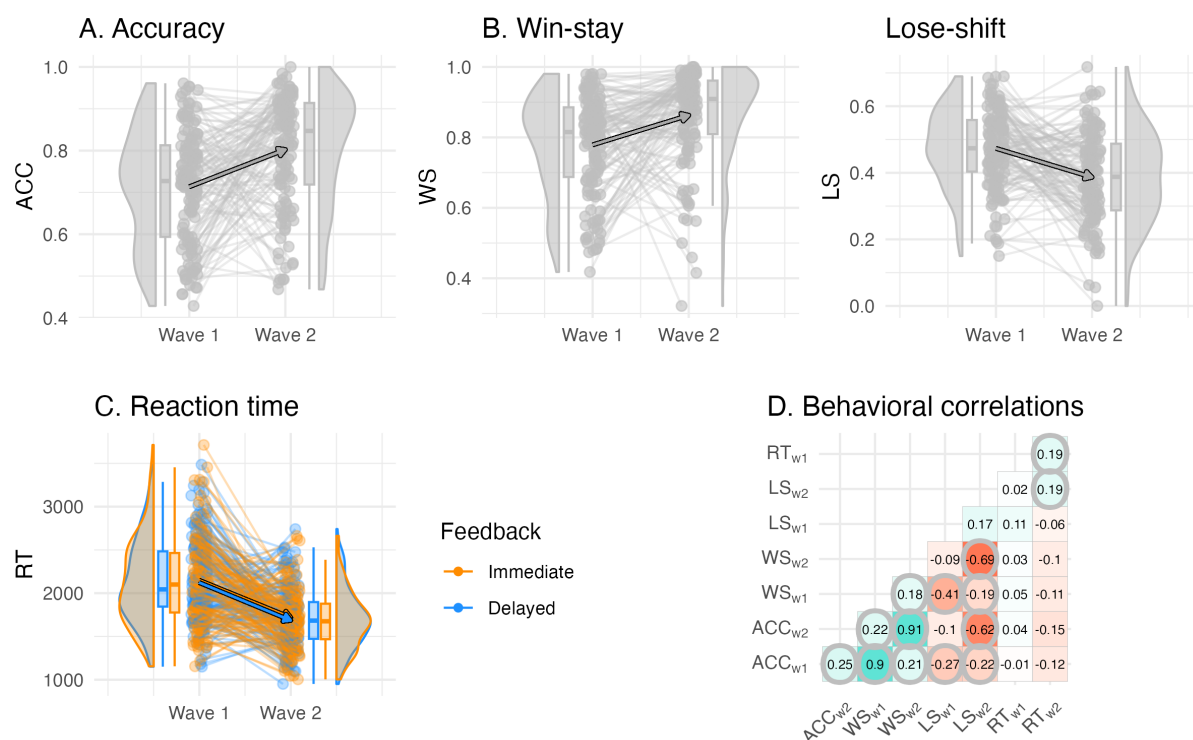
372

373 Table 1. Descriptive behavioral results of dependent variables Accuracy (ACC, probability correct),
 374 win-stay probability (WS), lose-shift probability (LS), and reaction time (RT, in seconds), as well as
 375 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

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

379



380
 381 Figure 2. Individual differences in the behavioral reinforcement learning outcomes and their longitudinal
 382 change. (A) Accuracy did not differ by feedback timing and increased between waves. (B) Win-stay and
 383 lose-shift probability did not differ by feedback timing, and win-stay increased and lose-shift probability
 384 decreased between waves. (C) Reaction time differed by feedback timing, in which decisions for cues
 385 learned with delayed feedback were faster, and reaction times were faster at wave 2 compared to wave
 386 1. (D) Correlations between behavioral outcomes reveal that learning accuracy was primarily correlated
 387 with the win-stay and lose-shift probabilities both within and between waves, but was uncorrelated to
 388 reaction time. Significant correlations are circled, p -values were adjusted for multiple comparisons using
 389 bonferroni correction.

390

391 Modeling results

392

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

403

404 Table 2. Model comparison results.

Model	Parameters	$\Delta elpd_{loo}$ [SE]	$\Sigma elpd_{loo}$ [mean]	<i>pseudo-BMA+</i>
Step 1: heuristic strategy models and value-based learning model				
<i>vbm</i> ₁	$1\alpha, 1\tau$	0 [0]	-15154.9 [-0.45]	1
<i>ws</i>	$1\tau_{ws}$	-1327.7 [159.5]	-16482.7 [-0.49]	<0.01
<i>wsls</i>	$1\tau_{wsls}$	-4247.3 [284.8]	-19402.3 [-0.58]	0
Step 2: value-based learning models				
<i>vbm</i>₃	$1\alpha, 2\tau$	0 [0]	-15045.3 [-0.45]	0.73
<i>vbm</i> ₇	$1\alpha, 2\rho$	-2.93 [2.92]	-15048.2 [-0.45]	0.24
<i>vbm</i> ₆	$2\alpha, 1\rho$	-24.34 [8.85]	-15069.6 [-0.45]	<0.01
<i>vbm</i> ₈	$2\alpha, 2\rho$	-29.71 [15.95]	-15075.0 [-0.45]	0.02
<i>vbm</i> ₄	$2\alpha, 2\tau$	-43.34[14.89]	-15088.6 [-0.45]	<0.01
<i>vbm</i> ₂	$2\alpha, 1\tau$	-46.45 [13.97]	-15091.7 [-0.45]	<0.01
<i>vbm</i> ₅	$1\alpha, 1\rho$	-59.01 [7.59]	-15104.3 [-0.45]	<0.01
<i>vbm</i> ₁	$1\alpha, 1\tau$	-109.63 [11.98]	-15154.9 [-0.45]	<0.01

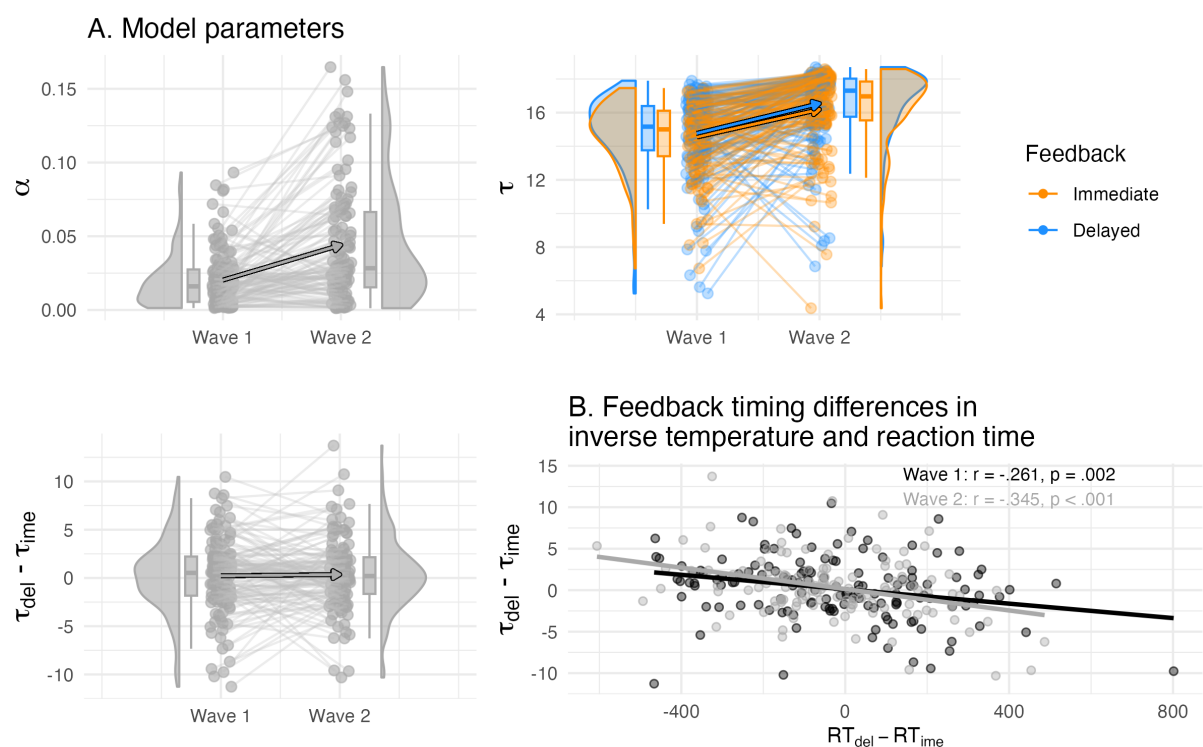
405 *Note.* Model = heuristic (*ws*, *wsls*) and value-based models (*vbm*₁₋₈) that were compared against each
 406 other. Parameters = corresponding model parameters learning rate α , inverse temperature τ and
 407 outcome sensitivity ρ . $\Delta elpd_{loo}$ [SE] = difference in the Bayesian leave-one-out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

412

413 *Feedback timing modulated choice stochasticity.* Model *vbm*₃ ($1\alpha 2\tau$) showed the largest model
 414 evidence, reflected in the highest expected log pointwise predictive density and highest model weights
 415 and suggests that feedback timing affected the inverse temperature, but not the learning rate or outcome
 416 sensitivity ($elpd_{loo} = -15045.3$, *pseudo-BMA+* = 0.73, Table 2). Table 3 and Figure 3A provide a
 417 descriptive overview of the winning model parameters. Of note, there were only small differences in
 418 model fit ($elpd_{loo}$) to the second-best model (*vbm*₇, $1\alpha 2\rho$, $\Delta elpd_{loo} = -2.93$, $elpd_{SE_{loo}} = 2.92$,
 419 *pseudo-BMA+* = 0.24), which suggests a potential separable feedback timing effect on outcome
 420 sensitivity. The average inverse temperature did not differ by feedback condition, but showed large
 421 within-person condition differences at both waves, indicating individual differences in feedback timing
 422 modulation (wave 1: $\Delta\tau_{del-ime}$ Mean = 0.22, SD = 3.80, Range = 21.74, wave 2: $\Delta\tau_{del-ime}$ Mean =
 423 0.35, SD = 3.70, Range = 24.03). The correlations between the parameters are shown in Supplementary
 424 Material 4.

425 Since reaction times were predicted by feedback timing behaviorally, and inverse temperature is
426 assumed to reflect decision-making, we were interested in whether differences in reaction time were
427 related to inverse temperature differences. Indeed, at both waves, children who responded faster during
428 delayed compared to immediate feedback had a higher inverse temperature at delayed compared to
429 immediate feedback (wave 1: $r = -.261$, $t = -3.18$, $p = .002$, wave 2: $r = -.345$, $t = -4.10$, $p < .001$, Figure
430 3B). Taken together, children's learning behavior was best described by a value-based model, where
431 feedback timing modulated individual differences in the choice rule during value-based learning.
432 Interestingly, the differences in the choice rule and reaction time f were correlated. Specifically, more
433 value-guided choice behavior (i.e., higher inverse temperature) was related to faster responses during
434 delayed feedback relative to immediate feedback, suggesting a link between model parameter and
435 behavior in relation to feedback timing.

436



437

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

445

446

447

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

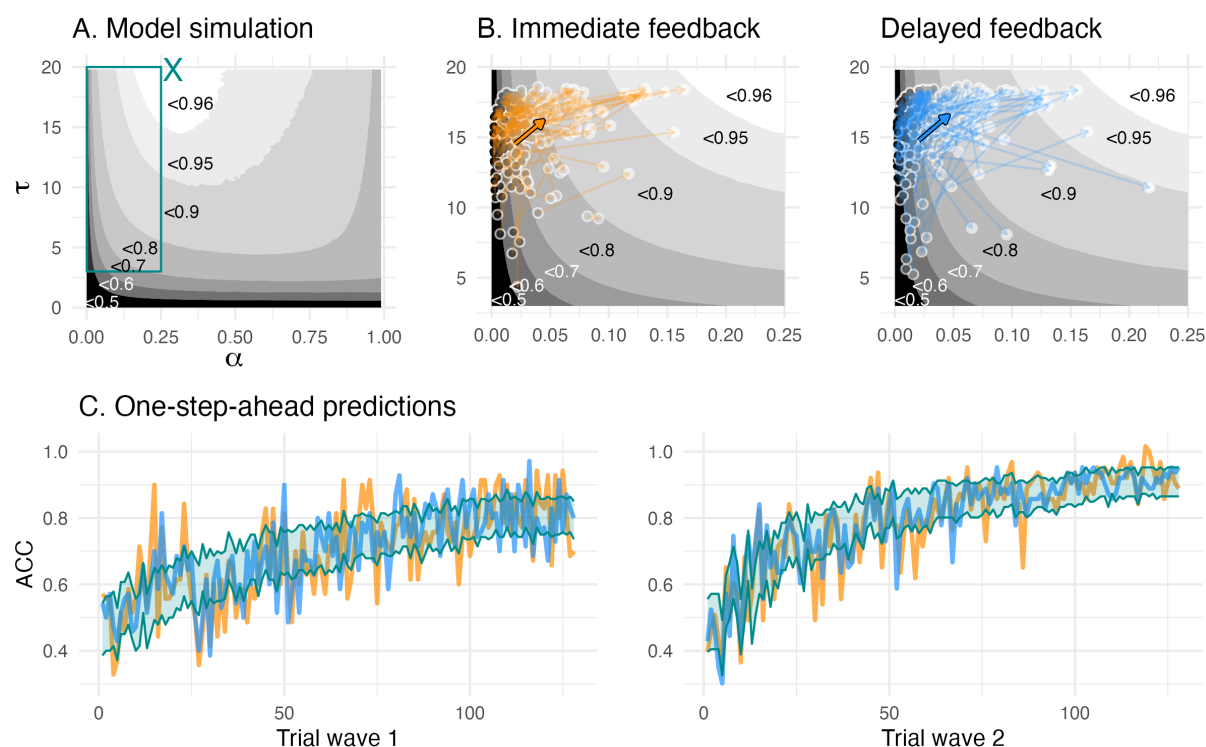
	Wave 1					Wave 2				
	α	τ_{Ime}	τ_{Del}	ls_{Ime}	ls_{Del}	α	τ_{Ime}	τ_{Del}	ls_{Ime}	ls_{Del}
<i>Mean</i>	0.02	14.6	14.8	0.73	0.73	0.05	16.2	16.5	0.82	0.82
<i>SD</i>	0.02	2.04	2.37	0.12	0.13	0.04	2.37	2.21	0.13	0.13
<i>Min</i>	<0.01	6.73	5.25	0.53	0.53	<0.01	4.37	6.85	0.53	0.53
<i>Max</i>	0.09	17.5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

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

451

452 *Children's value-based learning became more optimal.* Next, we compared the parameter space
 453 according to model simulation (Figure 4A) with the empirical posterior parameters fitted by the
 454 winning model (Table 3, Figure 4B) to determine whether children increased their value-based
 455 learning towards more optimal parameter combinations. Both fitted and simulated parameter
 456 combinations allowed us to derive a learning score that captured learning performance according to
 457 the winning value-based model. Note that the learning score was defined as the average choice
 458 probability for the more rewarded choice option. We refer to these model-derived choice probabilities
 459 as learning score, since they reflect value-based learning and combine information of learned values,
 460 that depend on the learning rate, and values translated into choice probabilities, that depend on the
 461 inverse temperature. Thus, a higher learning score reflects more optimal value-based learning. We
 462 simulated 10,000 parameter combinations and created a learning score map according to each
 463 parameter combination (Figure 4A). The optimal parameter combination was at a learning rate α =
 464 0.29, and an inverse temperature τ = 19.8, and with an average learning score of 96.5 % (Figure 4A).
 465 Children's fitted average learning rates ranged 0.01 – 0.22 and inverse temperature 6.73 – 18.70 and
 466 were outside the parameter space above 96 % learning score (Table 3 and Figure 4A). The
 467 longitudinal average increase in learning rate and inverse temperature were mirrored by average
 468 increases in the learning scores, confirming our prediction that their parameters developed towards
 469 optimal value-based learning (arrow in Figure 4B). The one-step ahead predictions of the winning
 470 model captured children's choices overall well, predictive accuracies were 65.3 % at wave 1 and
 471 75.7 % at wave 2 (Figure 4C).

472



473

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

485

486 Longitudinal brain-cognition links

487

488 *Significant longitudinal change in brain and cognition.* We first performed univariate LCS model
489 analyses to estimate a latent change score of immediate and delayed learning scores as well as striatal
490 and hippocampal volumes (see descriptive changes in Figure 5B-C). All four variables of interest
491 showed significant positive mean changes and variances, and all univariate models provided a good fit
492 to the data (Supplementary Material 5). This allowed us to further relate the differences in structural
493 brain changes to changes in learning.

494

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

509

510 *Hippocampal and striatal volume showed distinct associations to learning.* We fitted a four-variate LCS
511 model to test our prediction of selective brain-cognition links. Specifically, we assumed a larger
512 contribution of striatal volume at immediate learning, and a larger contribution of hippocampal volume
513 at delayed learning. The LCS model provided good data fit ($\chi^2(27) = 15.4$, $CFI = 1.00$, $RMSEA(CI) =$
514 $0(0 - .010)$, $SRMR = .045$), and all relevant paths are shown in Figure 5D (see Table 4 for a detailed
515 model overview). For the striatal associations to cognition, we found that wave 1 striatal volume
516 covaried with both immediate learning score and delayed learning score ($\phi_{STR_{w1}, LS_{i,w1}} = 0.19$, $z = 2.52$,
517 $SE = 0.07$, $p = .012$, $\phi_{STR_{w1}, LS_{d,w1}} = 0.18$, $z = 2.37$, $SE = 0.07$, $p = .018$). Constraining the striatal
518 association to immediate learning to 0 worsened the model fit relative to the unrestricted model ($\Delta\chi^2(1)$
519 $= 5.66$, $p = .017$), which was the same when constraining the striatal association to delayed learning to
520 0 ($\Delta\chi^2(1) = 5.14$, $p = .023$). In summary, larger striatal volume was associated with better learning
521 scores for both immediate and better delayed feedback.

522 Hippocampal volume, on the other hand, only covaried with delayed learning at wave 1 ($\phi_{HPC_{w1}, LS_{d,w1}} =$
523 0.14 , $z = 2.05$, $SE = 0.07$, $p = .041$), not with immediate learning score ($\phi_{HPC_{w1}, LS_{i,w1}} = 0.12$, $z = 1.68$,
524 $SE = 0.07$, $p = .092$). Fixing the path between hippocampal volume and delayed learning to 0 worsened
525 the model fit relative to the unrestricted model ($\Delta\chi^2(1) = 4.19$, $p = .041$), but not when its path to
526 immediate learning was constrained to 0 ($\Delta\chi^2(1) = 2.94$, $p = .086$). This suggests that larger hippocampal
527 volume was specifically associated with better delayed learning. As a next step, the associations between
528 striatum and hippocampus to immediate or delayed learning was directly compared against each other.
529 A model equal-constraining striatal and hippocampal paths to immediate learning ($\Delta\chi^2(1) = 0.41$, p
530 $= .521$) and another model equal-constraining these paths to delayed learning ($\Delta\chi^2(1) = 0.14$, $p = .707$)
531 did not lead to a worse model fit compared to the unrestricted model, which suggests that the brain-

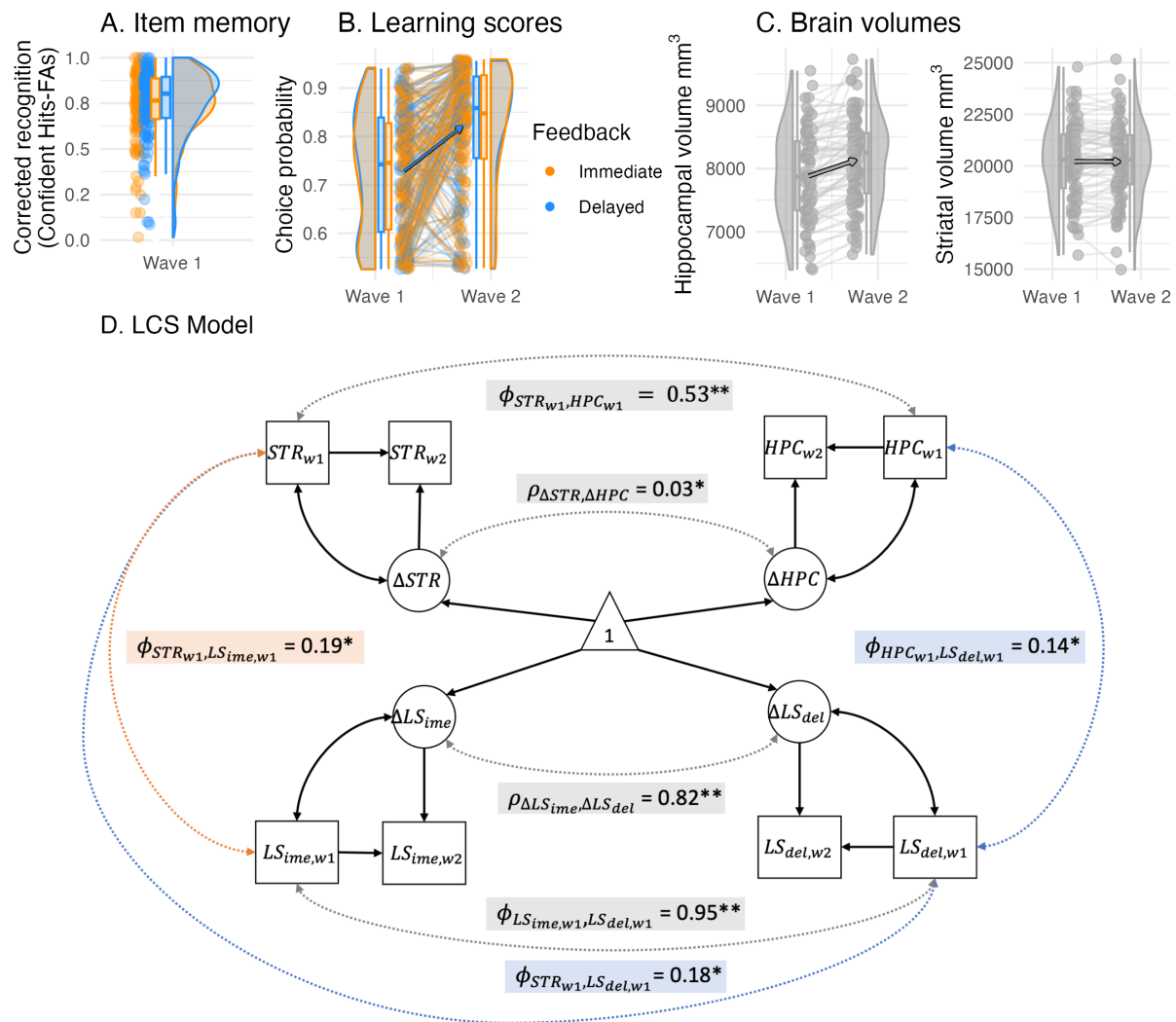
532 cognition links have considerable overlap. This is in line with the high wave 1 covariance and change-
533 change covariance within the brain and cognition domain (see Table 4). We found no longitudinal links
534 between the brain and cognition domains, which suggests that the found brain-cognition links at wave
535 1 remained longitudinally stable (see Supplementary Material 5 for an exploratory LCS model that
536 related the model parameters to striatal and hippocampal volume).

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

541

542 *Weak evidence for enhanced episodic memory during delayed feedback.* Finally, we investigated
543 whether a hippocampal contribution at delayed feedback would selectively enhance episodic memory.
544 Episodic memory, as measured by individual corrected object recognition memory (hits - false alarms),
545 showed at trend better memory for items shown in the delayed feedback condition ($\beta_{feedback=delayed}$
546 = .009, $SE = .005$, $t = 1.83$, $p = .069$, see Figure 5A). To summarize, there was weak support for enhanced
547 episodic memory during delayed compared to immediate feedback, in line with the idea of a selective
548 association between hippocampal volume and delayed feedback learning.

549



550 Figure 5. (A) Recognition memory (corrected recognition = hits - false alarms) for objects presented
 551 during delayed feedback was only enhanced at trend. (B) Learning scores depicted here were used in
 552 the LCS analyses. Learning scores were the model-derived choice probability of the contingent choice
 553 using fitted posterior parameters. (C) Hippocampal and striatal volumes increased between waves, while
 554 hippocampal volume increased most. (D) A four-variate latent change score (LCS) model that included
 555 striatal and hippocampal volumes as well as immediate and delayed learning scores. Depicted are
 556 significant paths cross-domain (brain-cognition, dashed lines) and within-domain (brain or cognition,
 557 solid lines), other paths are omitted for visual clarity and are summarized in Table 4. Depicted brain-
 558 cognition links included $\phi_{STR_{w1}, LS_{ime, w1}}$ (covariance between striatal volume and immediate learning
 559 score at wave 1), as well as $\phi_{HPC_{w1}, LS_{del, w1}}$ and $\phi_{STR_{w1}, LS_{del, w1}}$ (covariances between hippocampal and
 560 striatal volumes and delayed learning score at wave 1). Brain links included $\phi_{STR_{w1}, HPC_{w1}}$ and
 561 $\rho_{\Delta STR, \Delta HPC}$ (wave 1 covariance and change-change covariance), and similarly, cognition links included
 562 $\phi_{LS_{ime, w1}, LS_{del, w1}}$ and $\rho_{\Delta LS_{ime}, \Delta LS_{del}}$. Covariates included age, sex and estimated total intracranial
 563 volume. ****** denotes significance at $\alpha < .001$, ***** at $\alpha < .05$.

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

	<i>STR</i>	<i>LS_{ime}</i>	<i>HPC</i>	<i>LS_{del}</i>
Model fit: $\chi^2 = 15.4$, $df = 27$, $CFI = 1$, $RMSEA (CI) = 0 (0-0.01)$, $SRMR = 0.045$				
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)				
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)				
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)				
wave 1 covariation ϕ	0.95** (0.10)			
change-change covariance ρ	0.82** (0.10)			
wave 1 <i>LS_{ime}</i> into <i>LS_{del}</i> change γ	-0.07 (0.27)			
wave 1 <i>LS_{del}</i> into <i>LS_{ime}</i> change γ	0.06 (0.28)			

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

571

572

Discussion

573

574 In this study, we examined the longitudinal development of value-based learning in middle childhood
575 and its associations with striatal and hippocampal volumes that were predicted to differ by feedback
576 timing. Children improved their learning in the 2-year study period. Behaviorally, learning was
577 improved by an increase in accuracy and a reduction in reaction time (i.e., faster responses). Further,
578 children's switching behavior improved by an increase in win-stay and a decrease in lose-shift behavior.

579 Computationally, learning was enhanced by an increase in learning rate and inverse temperature, which
580 together constituted more optimal value-based learning. Further, feedback timing modulated specifically
581 the inverse temperature. In terms of brain structures, we found that longitudinal changes in hippocampal
582 volume were larger compared to striatal volume, which suggests more protracted hippocampal
583 maturation. The brain-cognition links were longitudinally stable and partially confirmed our hypotheses.
584 In line with previous adult literature and our assumption, hippocampal volume was more strongly
585 associated with delayed feedback learning, and there was weak evidence of enhanced episodic memory
586 performance under delayed feedback compared to immediate feedback. Contrary to our expectations,
587 striatal volume was associated with not just immediate but also delayed feedback learning, suggesting
588 a common involvement of the striatum during value-based learning in middle childhood across
589 timescales.

590

591 Children's learning improvement between waves was described behaviorally by increased win-stay and
592 decreased lose-shift behavior. Our finding is in line with cross-sectional studies in the developmental
593 literature that reported increased learning accuracy and win-stay behavior^{58,59}. Our longitudinal dataset
594 with younger children further suggests that learning change is not only accompanied by increased win-
595 stay, but also decreased lose-shift behavior. We found lower learning performance and less optimal
596 switching behavior in girls compared to boys, which could point to sex differences for reinforcement
597 learning during middle childhood (Supplementary Material 2). Previous studies have found both male
598 and female advantages depending on their age and the type of learning task^{38,60,61}. Alternatively, sex
599 differences may have been driven by confounding variables not included in the analysis.
600 Computationally, we found longitudinally increased and more optimal learning rate and inverse
601 temperature, as shown by simulation data, that add to the growing literature of developmental
602 reinforcement learning¹⁶. Our study underscores the importance of relating empirical values to
603 simulation-based optimal values, as task characteristics such as reward probability and learning
604 environment stability determine the range of optimal parameter values²⁷.

605

606 Despite a relatively immature hippocampal structure in middle childhood, our results confirmed a
607 longitudinally stable association between hippocampal volume and delayed feedback learning. However,
608 episodic memory in this learning condition was not enhanced. This suggests a developmentally early
609 hippocampal contribution to value-based learning during delayed feedback, which does not modulate
610 episodic memory as much as compared to adults. Therefore, our study partially extends the findings
611 from the adult literature to middle childhood^{5,12-14}. The reduced effect of delayed feedback on episodic
612 memory may be due to the protracted development of hippocampal maturation. In an aging study with
613 a similar task, older adults failed to exhibit enhanced episodic memory for objects presented during
614 delayed feedback trials, and they showed no enhanced hippocampal activation during delayed feedback
615 and¹⁴. Therefore, the findings converge nicely at both childhood and older adulthood, during which the

616 structural and functional integrity of hippocampus are known to be less optimal than at younger
617 adulthood⁶²⁻⁶⁴.

618 Our brain-cognition links were only partially confirmed, as striatal volumes exhibited associations with
619 not just immediate learning scores, as we predicted, but also with delayed learning scores. This result
620 suggests that the striatum may be important for value-based learning in general rather than exhibiting a
621 selective association with immediate feedback learning. This is also what we found in an explorative
622 analysis that related the striatum to learning rate in general and further predicted longitudinal change in
623 learning rate (Supplemental Material 4). This overall reduced brain-behavior specificity could reflect
624 less differentiated memory systems during development, similar to findings from aging research. Here,
625 older adults exhibited stronger striatal and hippocampal co-activation during both implicit and explicit
626 learning, compared to more dissociable brain-behavior relationships in younger adults⁶⁵. Interestingly,
627 even in young adults, clear dissociations between memory systems such as in non-human lesion studies
628 are uncommon, and factors like stress modulate their cooperative interaction^{6,10,11,66,67}. Further, there are
629 methodological differences to previous studies that could explain why striatal volumes were not
630 uniquely associated with immediate learning in our study. For example, previous studies related reward
631 prediction errors to striatal and hippocampal activation^{5,13,14}, whereas we examined individual
632 differences in brain structure and the model-derived learning scores. Future functional neuroimaging
633 studies with children could further clarify whether children's memory systems are indeed less
634 differentiated and explain the attenuated modulation by feedback timing. Taken together, compared to
635 the adult literature, our results with children showed that the hippocampal structure was associated with
636 delayed feedback learning, but did not enhance episodic memory encoding, while the striatum generally
637 supported value-based learning. These findings point towards a developmental effect of less
638 differentiated and more cooperative memory systems in middle childhood.

639
640 Our computational modeling results revealed a separable effect of feedback timing on inverse
641 temperature, which suggests that the memory systems modulated learning during decision-making. The
642 reported behavioral differences in reaction time and their correlation to the inverse temperature further
643 support the idea of a decision-related mechanism, as we found children to respond faster during delayed
644 feedback trials and faster responding children also exhibited more value-guided choice behavior (i.e.
645 higher inverse temperature) during delayed compared to immediate feedback. The hippocampus may
646 contribute to a decision-related effect in the delayed feedback condition by facilitating the encoding and
647 retrieval of learned values⁶⁸. This is in contrast to previous event-related fMRI and EEG studies
648 reporting feedback timing modulations at value update^{5,13,14}, which may be due to at least two reasons.
649 First, we did not include a functional brain measure to examine its differential engagement during the
650 choice and feedback phases. Second, in such a reinforcement learning task, disentangling model
651 parameters from the choice and feedback phases can be challenging, such as for the inverse temperature
652 and outcome sensitivity⁶⁹. Hippocampal engagement at delayed feedback may enhance outcome

653 sensitivity, as well as facilitate cue-choice associations and improve retrieval and choice behavior. A
654 mechanism facilitating retrieval seems especially relevant in our paradigm, where multiple cues were
655 learned and presented in a mixed order, thus creating a high memory load. To summarize, our study
656 results suggest that feedback timing can modulate decision-making. However, disentangling the effects
657 of inverse temperature and outcome sensitivity is challenging and warrants careful interpretation. Future
658 studies might shed new light by examining neural activations at both task phases, and by choosing a
659 task design that allows independent manipulations on these phases and associated model parameters,
660 e.g., by using different reward magnitudes during reinforcement learning, or by studying outcome
661 sensitivity without decision-making.

662
663 One aim of developmental investigations is to identify the emergence of brain and cognition dynamics,
664 such as the hippocampal-dependent and striatal-dependent memory systems, which have been shown to
665 engage during reinforcement learning depending on the delay in feedback delivery. Our longitudinal
666 study partially confirmed these brain-cognition links in middle childhood but with less specificity as
667 previously found in adults.

668 An early existing memory system dynamic, similar to that of adults, is relevant for applying
669 reinforcement learning principles at different timescales. For example, in a school context, learning
670 processes can be better structured according to their development. Furthermore, probabilistic learning
671 from delayed feedback may be a potential diagnostic tool to examine the hippocampal-dependent
672 memory system during learning in children at risk. Environmental factors such as stress¹¹ and
673 socioeconomic status^{39,70} have been shown to affect hippocampal structure and function and may
674 contribute to a heightened risk for psychopathology in the long term⁷¹⁻⁷³. Deficits in hippocampal-
675 dependent learning may be particularly relevant to psychopathology since dysfunctional behavior may
676 arise from a tendency to prioritize short-term consequences over long-term ones^{74,75} and from the
677 maladaptive application of previously learned behavior in inappropriate contexts⁷⁶.

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

683

684

685

686

687

688

Additional Information

689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716

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').

Acknowledgments. We thank the Max Planck Institute for Human Development and all members of the Jacobs study team for their vital contribution, and all participants and family members for taking part in the study.

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

Ethics approval. This study was approved by the “Deutsche Gesellschaft für Psychologie” ethics committee (YLS_012015).

Availability of data and code. <https://osf.io/pju65/>

Author ORCIDs.

JF: <https://orcid.org/0000-0003-0505-0798>

LZ: <https://orcid.org/0000-0002-9586-595X>

LR: <https://orcid.org/0000-0002-0144-5605>

JJM: <https://orcid.org/0000-0002-3893-8008>

JT: <https://orcid.org/0000-0001-8166-2441>

CH: <https://orcid.org/0000-0002-6580-6326>

YLS: <https://orcid.org/0000-0001-8922-7292>

References

- 717
- 718 1. Sutton, R. S. & Barto, A. G. *Reinforcement learning: An introduction*. (MIT press, 2018).
- 719 2. Gläscher, J., Daw, N., Dayan, P. & O’Doherty, J. P. States versus Rewards: Dissociable neural
720 prediction error signals underlying model-based and model-free reinforcement learning.
721 *Neuron* **66**, 585 (2010).
- 722 3. Bolenz, F., Reiter, A. M. F. & Eppinger, B. Developmental Changes in Learning:
723 Computational Mechanisms and Social Influences. *Front. Psychol.* **0**, 2048 (2017).
- 724 4. Zhang, L. & Gläscher, J. A brain network supporting social influences in human decision-
725 making. *Sci. Adv.* **6**, 1–20 (2020).
- 726 5. Foerde, K. & Shohamy, D. Feedback Timing Modulates Brain Systems for Learning in
727 Humans. *J. Neurosci.* **31**, 13157–13167 (2011).
- 728 6. Packard, M. G. & Goodman, J. Factors that influence the relative use of multiple memory
729 systems. *Hippocampus* **23**, 1044–1052 (2013).
- 730 7. Goodman, J. & Packard, M. G. Memory Systems of the Basal Ganglia. *Handb. Behav.*
731 *Neurosci.* **24**, 725–740 (2016).
- 732 8. Davidow, J. Y., Foerde, K., Galván, A. & Shohamy, D. An Upside to Reward Sensitivity: The
733 Hippocampus Supports Enhanced Reinforcement Learning in Adolescence. *Neuron* **92**, 93–99
734 (2016).
- 735 9. Hartley, C. A., Nussenbaum, K. & Cohen, A. O. Interactive Development of Adaptive
736 Learning and Memory. 1–27 (2021).
- 737 10. Packard, M. G., Goodman, J. & Ressler, R. L. Emotional modulation of habit memory: neural
738 mechanisms and implications for psychopathology. *Curr. Opin. Behav. Sci.* **20**, 25–32 (2018).
- 739 11. Schwabe, L. & Wolf, O. T. Stress and multiple memory systems: from ‘thinking’ to ‘doing’.
740 *Trends Cogn. Sci.* **17**, 60–68 (2013).
- 741 12. Foerde, K., Race, E., Verfaellie, M. & Shohamy, D. A role for the medial temporal lobe in
742 feedback-driven learning: Evidence from amnesia. *J. Neurosci.* **33**, 5698–5704 (2013).
- 743 13. Höljtje, G. & Mecklinger, A. Feedback timing modulates interactions between feedback
744 processing and memory encoding: Evidence from event-related potentials. *Cogn. Affect. Behav.*
745 *Neurosci.* **2020** **202** **20**, 250–264 (2020).
- 746 14. Lighthall, N. R., Pearson, J. M., Huettel, S. A. & Cabeza, R. Feedback-Based Learning in
747 Aging: Contributions and Trajectories of Change in Striatal and Hippocampal Systems. *J.*
748 *Neurosci.* **38**, 8453–8462 (2018).
- 749 15. Zhang, L., Lengersdorff, L., Mikus, N., Gläscher, J. & Lamm, C. Using reinforcement learning
750 models in social neuroscience: Frameworks, pitfalls and suggestions of best practices. *Soc.*
751 *Cogn. Affect. Neurosci.* **15**, 695–707 (2020).
- 752 16. Nussenbaum, K. & Hartley, C. A. Reinforcement learning across development: What insights
753 can we draw from a decade of research? *Developmental Cognitive Neuroscience* **40**, (2019).

- 754 17. Decker, J. H., Lourenco, F. S., Doll, B. B. & Hartley, C. A. Experiential reward learning
755 outweighs instruction prior to adulthood. *Cogn. Affect. Behav. Neurosci.* **15**, 310–320 (2015).
- 756 18. Javadi, A. H., Schmidt, D. H. K. & Smolka, M. N. Differential representation of feedback and
757 decision in adolescents and adults. *Neuropsychologia* **56**, 280–288 (2014).
- 758 19. Palminteri, S., Kilford, E. J., Coricelli, G. & Blakemore, S. J. The Computational Development
759 of Reinforcement Learning during Adolescence. *PLoS Comput. Biol.* **12**, 1–25 (2016).
- 760 20. Master, S. L. *et al.* Distangling the systems contributing to changes in learning during
761 adolescence. *Dev. Cogn. Neurosci.* **41**, 100732 (2020).
- 762 21. Hauser, T. U., Iannaccone, R., Walitza, S., Brandeis, D. & Brem, S. Cognitive flexibility in
763 adolescence: Neural and behavioral mechanisms of reward prediction error processing in
764 adaptive decision making during development. *Neuroimage* **104**, 347–354 (2015).
- 765 22. Moutoussis, M. *et al.* Change, stability, and instability in the Pavlovian guidance of behaviour
766 from adolescence to young adulthood. *PLoS Comput. Biol.* **14**, (2018).
- 767 23. Van Den Bos, W., Cohen, M. X., Kahnt, T. & Crone, E. A. Striatum-medial prefrontal cortex
768 connectivity predicts developmental changes in reinforcement learning. *Cereb. Cortex* **22**,
769 1247–1255 (2012).
- 770 24. Rodriguez Buritica, J. M., Heekeren, H. R. & van den Bos, W. The computational basis of
771 following advice in adolescents. *J. Exp. Child Psychol.* **180**, 39–54 (2019).
- 772 25. Galván, A. The Teenage Brain: Sensitivity to Rewards. *Curr. Dir. Psychol. Sci.* **22**, 88–93
773 (2013).
- 774 26. van Duijvenvoorde, A. C. K. *et al.* A cross-sectional and longitudinal analysis of reward-
775 related brain activation: Effects of age, pubertal stage, and reward sensitivity. *Brain Cogn.* **89**,
776 3–14 (2014).
- 777 27. Eckstein, M. K., Wilbrecht, L. & Collins, A. G. E. What do RL Models Measure ? Interpreting
778 Model Parameters in Cognition and Neuroscience. *Curr. Opin. Behav. Sci.* **41**, 128–137 (2021).
- 779 28. Cohen, A. O., Nussenbaum, K., Dorfman, H. M., Gershman, S. J. & Hartley, C. A. The rational
780 use of causal inference to guide reinforcement learning strengthens with age. *npj Sci. Learn.* **5**,
781 1–9 (2020).
- 782 29. Raznahan, A. *et al.* Longitudinal four-dimensional mapping of subcortical anatomy in human
783 development. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1592 (2014).
- 784 30. Wierenga, L. *et al.* Typical development of basal ganglia, hippocampus, amygdala and
785 cerebellum from age 7 to 24. *Neuroimage* **96**, 67–72 (2014).
- 786 31. Giedd, J. N. Structural Magnetic Resonance Imaging of the Adolescent Brain. *Ann. N. Y. Acad.*
787 *Sci.* **1021**, 77–85 (2004).
- 788 32. Uematsu, A. *et al.* Developmental Trajectories of Amygdala and Hippocampus from Infancy to
789 Early Adulthood in Healthy Individuals. *PLoS One* **7**, e46970 (2012).
- 790 33. Giedd, J. N. *et al.* Child Psychiatry Branch of the National Institute of Mental Health

- 791 Longitudinal Structural Magnetic Resonance Imaging Study of Human Brain Development.
792 *Neuropsychopharmacology* **40**, 43 (2015).
- 793 34. Goodman, J., Marsh, R., Peterson, B. S. & Packard, M. G. Annual research review: The
794 neurobehavioral development of multiple memory systems--implications for childhood and
795 adolescent psychiatric disorders. *J. Child Psychol. Psychiatry.* **55**, 582–610 (2014).
- 796 35. Goddings, A. L. *et al.* The influence of puberty on subcortical brain development. *Neuroimage*
797 **88**, 242–251 (2014).
- 798 36. Dima, D. *et al.* Subcortical volumes across the lifespan: Data from 18,605 healthy individuals
799 aged 3–90 years. *Hum. Brain Mapp.* 1–18 (2021). doi:10.1002/hbm.25320
- 800 37. Lavenex, P. & Banta Lavenex, P. Building hippocampal circuits to learn and remember:
801 Insights into the development of human memory. *Behavioural Brain Research* **254**, 8–21
802 (2013).
- 803 38. Mandolesi, L., Petrosini, L., Menghini, D., Addona, F. & Vicari, S. Children’s radial arm
804 maze performance as a function of age and sex. *Int. J. Dev. Neurosci.* **27**, 789–797 (2009).
- 805 39. Raffington, L. *et al.* Stable longitudinal associations of family income with children’s
806 hippocampal volume and memory persist after controlling for polygenic scores of educational
807 attainment. *Dev. Cogn. Neurosci.* **40**, 100720 (2019).
- 808 40. Raffington, L. *et al.* Effects of stress on 6- and 7-year-old children’s emotional memory differs
809 by gender. *J. Exp. Child Psychol.* **199**, 104924 (2020).
- 810 41. Fischl, B. FreeSurfer. *Neuroimage* **62**, 774–781 (2012).
- 811 42. Phan, T. V., Smeets, D., Talcott, J. B. & Vandermosten, M. Processing of structural
812 neuroimaging data in young children: Bridging the gap between current practice and state-of-
813 the-art methods. *Dev. Cogn. Neurosci.* **33**, 206–223 (2018).
- 814 43. Schoemaker, D. *et al.* Hippocampus and amygdala volumes from magnetic resonance images
815 in children: Assessing accuracy of FreeSurfer and FSL against manual segmentation.
816 *Neuroimage* **129**, 1–14 (2016).
- 817 44. Bates, D., Mächler, M., Bolker, B. & Walker, S. Fitting Linear Mixed-Effects Models Using
818 lme4. *J. Stat. Softw.* **67**, 1–48 (2015).
- 819 45. Brown, V. M. *et al.* Reinforcement Learning Disruptions in Individuals with Depression and
820 Sensitivity to Symptom Change following Cognitive Behavioral Therapy. *JAMA Psychiatry*
821 (2021). doi:10.1001/jamapsychiatry.2021.1844
- 822 46. Stan Development Team. RStan: the R interface to Stan. R package version 2.21.2. [http://mc-](http://mc-stan.org)
823 [stan.org](http://mc-stan.org) (2021).
- 824 47. R Core Team. R: A Language and Environment for Statistical Computing. (2021).
- 825 48. Ahn, W.-Y., Haines, N. & Zhang, L. Revealing Neurocomputational Mechanisms of
826 Reinforcement Learning and Decision-Making With the hBayesDM Package. *Comput.*
827 *Psychiatry* **1**, 24 (2017).

- 828 49. Brown, V. M., Chen, J., Gillan, C. M. & Price, R. B. Improving the Reliability of
829 Computational Analyses: Model-Based Planning and Its Relationship With Compulsivity. *Biol.*
830 *Psychiatry Cogn. Neurosci. Neuroimaging* **5**, 601–609 (2020).
- 831 50. Vehtari, A., Gelman, A. & Gabry, J. Practical Bayesian model evaluation using leave-one-out
832 cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
- 833 51. Yao, Y., Vehtari, A., Simpson, D. & Gelman, A. Using Stacking to Average Bayesian
834 Predictive Distributions (with Discussion). *Bayesian Anal.* **13**, 917–1007 (2018).
- 835 52. Crawley, D. *et al.* Modeling flexible behavior in childhood to adulthood shows age-dependent
836 learning mechanisms and less optimal learning in autism in each age group. *PLoS Biol.* **18**, 1–
837 25 (2020).
- 838 53. Wilson, R. C. & Collins, A. G. E. Ten simple rules for the computational modeling of
839 behavioral data. *Elife* **8**, 1–33 (2019).
- 840 54. Kievit, R. A. *et al.* Developmental cognitive neuroscience using latent change score models: A
841 tutorial and applications. *Dev. Cogn. Neurosci.* **33**, 99–117 (2018).
- 842 55. Ferrer, E. & McArdle, J. J. Longitudinal modeling of developmental changes in psychological
843 research. *Curr. Dir. Psychol. Sci.* **19**, 149–154 (2010).
- 844 56. Sluis, S. van der, Verhage, M., Posthuma, D. & Dolan, C. V. Phenotypic Complexity,
845 Measurement Bias, and Poor Phenotypic Resolution Contribute to the Missing Heritability
846 Problem in Genetic Association Studies. *PLoS One* **5**, e13929 (2010).
- 847 57. Little, T. *Longitudinal structural equation modeling*. (Guilford Press, 2013).
848 doi:10.2/JQUERY.MIN.JS
- 849 58. Chierchia, G. *et al.* Confirmatory reinforcement learning changes with age during adolescence.
850 *Dev. Sci.* e13330 (2021). doi:10.1111/desc.13330
- 851 59. Habicht, J., Bowler, A., Moses-Payne, M. E. & Hauser, T. U. Children are full of optimism, but
852 those rose-tinted glasses are fading – reduced learning from negative outcomes drives
853 hyperoptimism in children. (2021).
- 854 60. Overman, W. H. Sex differences in early childhood, adolescence, and adulthood on cognitive
855 tasks that rely on orbital prefrontal cortex. *Brain Cogn.* **55**, 134–147 (2004).
- 856 61. Evans, K. L. & Hampson, E. Sex-dependent effects on tasks assessing reinforcement learning
857 and interference inhibition. *Front. Psychol.* **6**, 1–10 (2015).
- 858 62. Shing, Y. L. *et al.* Episodic memory across the lifespan: The contributions of associative and
859 strategic components. *Neurosci. Biobehav. Rev.* **34**, 1080–1091 (2010).
- 860 63. Keresztes, A. *et al.* Hippocampal maturity promotes memory distinctiveness in childhood and
861 adolescence. *Proc. Natl. Acad. Sci. U. S. A.* **114**, 9212–9217 (2017).
- 862 64. Ghetti, S. & Bunge, S. A. Neural changes underlying the development of episodic memory
863 during middle childhood. *Dev. Cogn. Neurosci.* **2**, 381–395 (2012).
- 864 65. Dennis, N. A. & Cabeza, R. Age-related dedifferentiation of learning systems: An fMRI study

- 865 of implicit and explicit learning. *Neurobiol. Aging* **32**, 2318.e17-2318.e30 (2011).
- 866 66. Ferbinteanu, J. Contributions of Hippocampus and Striatum to Memory-Guided Behavior
867 Depend on Past Experience. *J. Neurosci.* **36**, 6459–6470 (2016).
- 868 67. White, N. M. & McDonald, R. J. Multiple Parallel Memory Systems in the Brain of the Rat.
869 *Neurobiol. Learn. Mem.* **77**, 125–184 (2002).
- 870 68. Shadlen, M. N. N. & Shohamy, D. Decision Making and Sequential Sampling from Memory.
871 *Neuron* **90**, 927–939 (2016).
- 872 69. Browning, M., Paulus, M. & Huys, Q. J. M. What is computational psychiatry good for? *Biol.*
873 *Psychiatry* **0**, (2022).
- 874 70. Hackman, D. A., Farah, M. J. & Meaney, M. J. Socioeconomic status and the brain:
875 mechanistic insights from human and animal research. *Nat. Rev. Neurosci.* *2010 119* **11**, 651–
876 659 (2010).
- 877 71. Frodl, T. *et al.* Childhood stress, serotonin transporter Gene and Brain structures in major
878 depression. *Neuropsychopharmacology* **35**, 1383–1390 (2010).
- 879 72. Lucassen, P. J., Korosi, A., Krugers, H. J. & Oomen, C. A. *Early Life Stress- and Sex-*
880 *Dependent Effects on Hippocampal Neurogenesis. Stress: Neuroendocrinology and*
881 *Neurobiology* **2**, (Elsevier Inc., 2017).
- 882 73. Rahman, M. M., Callaghan, C. K., Kerskens, C. M., Chattarji, S. & O’Mara, S. M. Early
883 hippocampal volume loss as a marker of eventual memory deficits caused by repeated stress.
884 *Sci. Rep.* **6**, 1–15 (2016).
- 885 74. Levin, M. E., Haeger, J., Ong, C. W. & Twohig, M. P. An Examination of the Transdiagnostic
886 Role of Delay Discounting in Psychological Inflexibility and Mental Health Problems. *Psychol.*
887 *Rec.* **68**, 201–210 (2018).
- 888 75. Von Siebenthal, Z. *et al.* Decision-making impairments following insular and medial temporal
889 lobe resection for drug-resistant epilepsy. *Soc. Cogn. Affect. Neurosci.* **12**, 128–137 (2017).
- 890 76. Maren, S., Phan, K. L. & Liberzon, I. The contextual brain: Implications for fear conditioning,
891 extinction and psychopathology. *Nat. Rev. Neurosci.* **14**, 417–428 (2013).

892