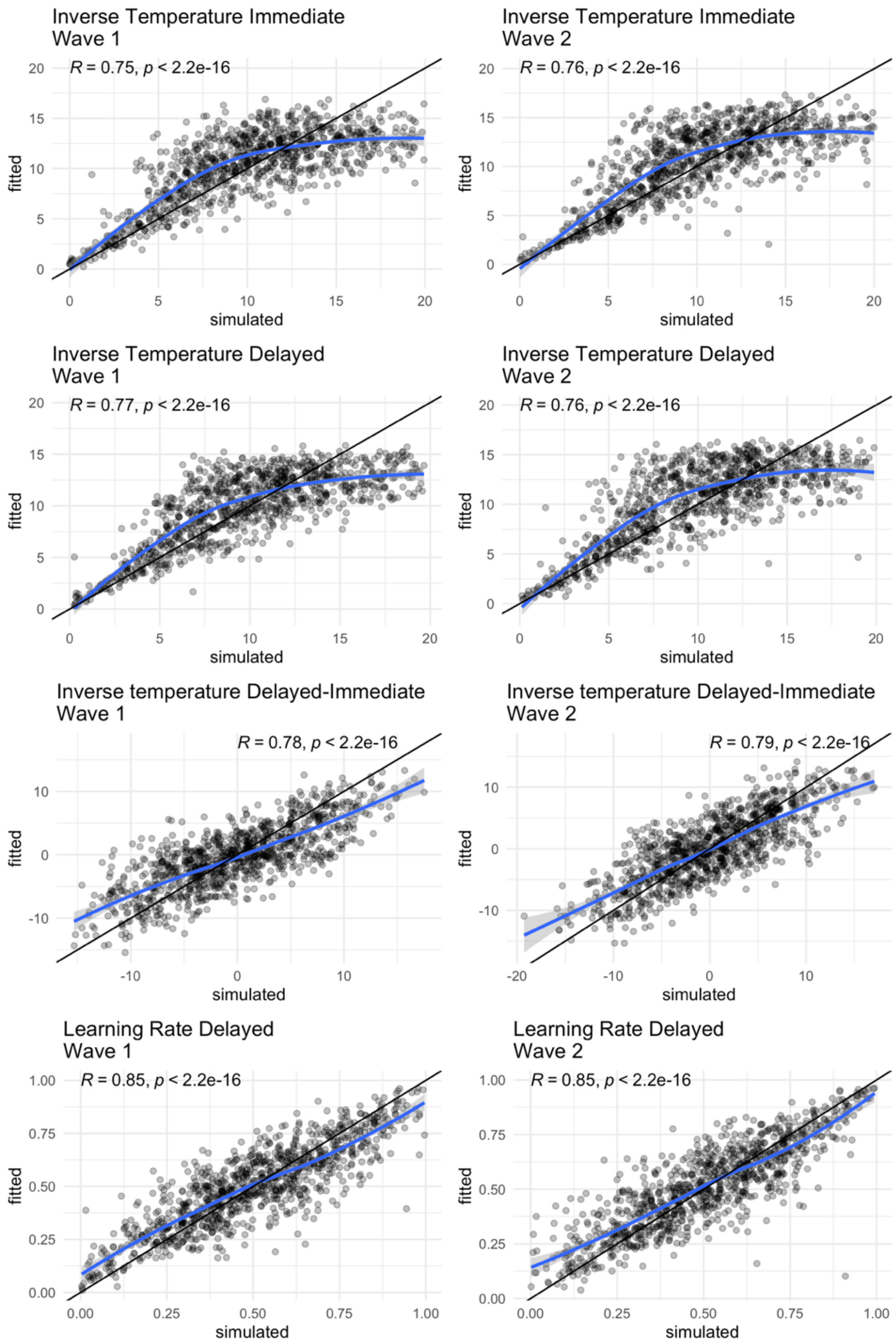
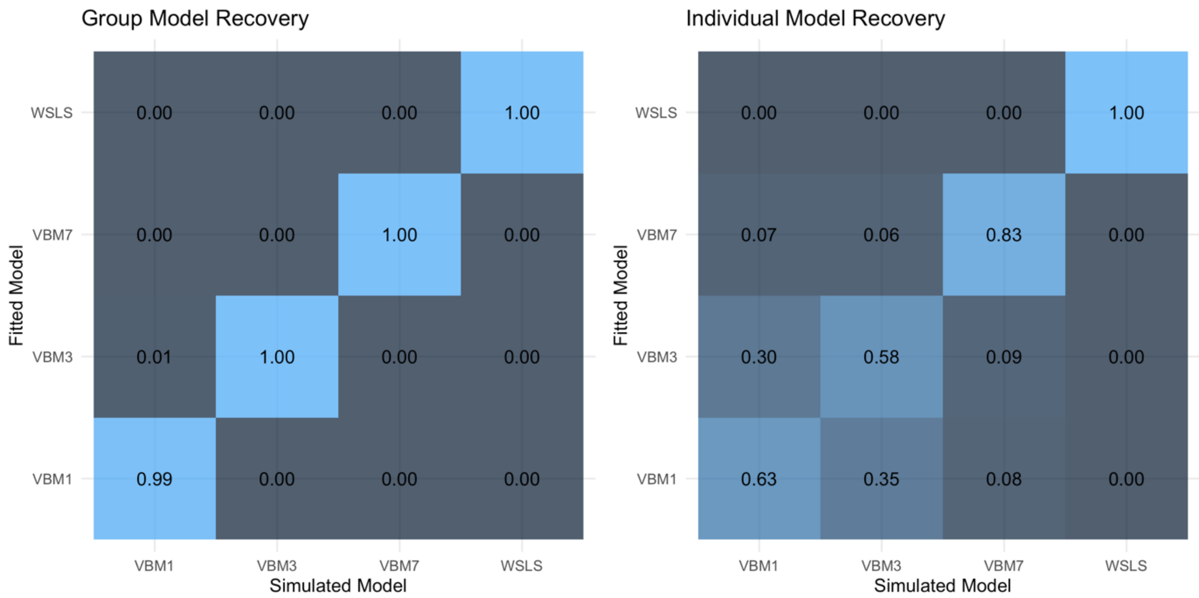
Supplementary Material 1: Parameter and Model recovery

We simulated 1000 datasets (50 groups, 20 datasets each) using a wide distribution within the boundaries for learning rate (boundaries = [0,1], *Mean* = 0.5, *SD* = 0.25) and inverse temperature (boundaries = [0,20], *Mean* = 10, *SD* = 5). We first performed a parameter recovery to see how well the winning model recovers the simulated parameters (Supplementary Figure 1). Both inverse temperature and learning rate were recovered overall well, with correlations of 0.75 – 0.77 for the inverse temperature, their condition differences correlating 0.78 – 0.79, and the learning rates correlating at 0.85. Inverse temperature values was slightly overestimated until a value of 12 and clearly underestimated above 12. The underestimation was less pronounced for the inverse temperature condition differences. Learning rate was also less biased – here, values below 0.5 sligthly overestimates and underestimated with values above 0.5. This means that more extreme values, i.e. those closer to the boundaries, were recovered closer towards the group mean. We next performed model recovery to see how well the model evidence is recovered compared to other models that were used during model comparison. Of all 10 models that were used, we performed model recovery on the two best models (winning model $vbm_3, 1\alpha, 2\tau$ and second-best model $vbm_7, 1\alpha, 2\rho$), our value-based baseline model ($vbm_1, 1\alpha, 1\tau$) and our heuristic strategy model (Supplementary Figure 2). We examined recovery on the group and individual level. On the group level, we used the model weight *Pseudo-BMA+* model for relative model evidence using Bayesian model averaging. On the individual level, we used model fit $elpd_{loo}$, which is the individual summed expected log pointwise predictive density of all trials. On the group level, model recovery was excellent, as all models were recovered with model weights of 0.99 – 1.00. On the individual level, model recovery was lower for the value-based models, with model weights of 0.58-0.83. Specifically, the models $vbm_1$and vbm_3, which only differed in whether inverse temperature was estimated separate by learning condition (immediate and delayed feedback) or across learning condition, were affected. Here, 35 % of the datasets that were simulated using separate inverse temperature fitted best on the model with one inverse temperature (and 30 % vice versa), and likely reflects the noisy property of the inverse temperature.



Supplementary Figure 1. Parameter recovery of the winning model, the black line represents the identity line, whereas the blue line is loess regression line, Correlations are calculated by Pearson's r.



Supplementary Figure 2. Model recovery on the group (left) and individual level (right). Group-level recovery values are the average model weights (across 20 groups, 50 datasets each) *Pseudo-BMA+* using Bayesian model averaging stabilized by Bayesian bootstrap using 100,000 iterations. Individual-level recovery values are the average model fits (across 1,000 datasets) $elpd_{loo}$, which is the individual summed expected log pointwise predictive density of all trials.

Supplementary Material 2: Model structure and detailed results of generalized linear mixed models (GLMM)

GLMM Random effects model structure

We ran four GLMMs with the dependent variables accuracy (1 = correct, 0 = incorrect), win-stay behavior (1 = win-stay, 0 = win-shift), lose-shift behavior (1 = lose-shift, 0 = lose-stay) and reaction time (in milliseconds) as the dependent variable (Supplementary Table 1). As fixed effects, we included within-subject factors wave (1 = wave 1, 2 = wave 2) and feedback type (1 = immediate, 2 = delayed) as well as the covariate sex (1 = girl, 2 = boy). The contrasts of the categorical variables were set using the contr.sum function to keep the mean intercept at the global mean. We first tested whether including the main effects of wave, feedback type and sex improved the model fit. We then tested whether including interaction terms between these three variables, and the model had to improve the overall model fit to be reported as the winning model. As random effects, data were clustered at the participant and learning block level, allowing fixed intercept for each of the 4 blocks (32 trials each) of each individual. As random slopes, we included within-subject factors wave and feedback type.

Supplementary Table 1. Mixed effects model structure and fixed effects results for the models using the dependent variables Accuracy (ACC), win-stay (WS), lose-shift (LS) and Reaction time (RT).

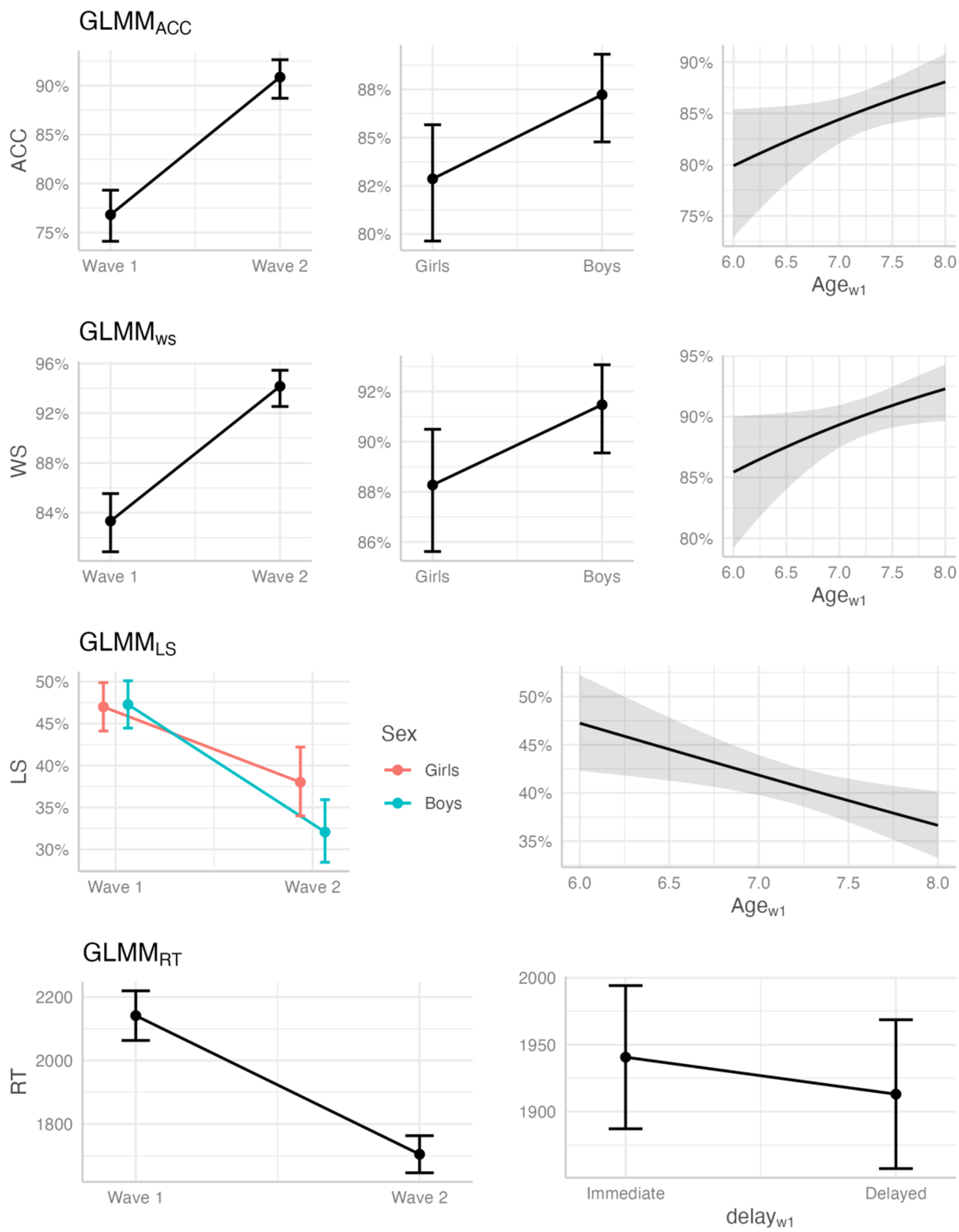
Fixed effects	$GLMM_{ACC}$	$GLMM_{WS}$	$GLMM_{LS}$	$GLMM_{RT}$
Feedback=Delayed	.013	.023	-.030	14.0*
Wave=2	.550**	.586**	-.252**	218**
Sex=Girls	-.172*	-.177*	.062	23.5
Wave 1 Age	.142*	.163*	-.100*	-24.5
Wave=1*Sex=Girls	not included	not included	.068*	not included
Random slopes				
Feedback Type	X	X	X	X
Wave	X	X	X	X
Random intercepts				
Participant ID	X	X	X	X
Block	X	X	X	X
Model fit				
ICC	0.44	0.45	0.12	0.23
Observations	33460	22013	10383	33460
Marginal R^2	0.056	0.063	0.021	0.036
Conditional R^2	0.472	0.482	0.138	0.258

Note. ** denotes significance at $\alpha < .001$, * at $\alpha < .05$. X indicates which random effects were included in the final model. ICC = intraclass correlation. Marginal R^2 = variance explained by fixed effects, Conditional R^2 = variance explained by random effects.

Detailed GLMM results

With the complete dataset, we found that increased learning accuracy was predicted at wave 2 compared to wave 1 ($\beta_{wave=2}$ = .550, *SE* = .061, z = 8.97, p < .001) and with higher age at wave 1 ($\beta_{wave\ 1\ age}$ = .142, *SE* = .070, z = 2.03, p = .043), but there were no differences in accuracy by feedback timing ($\beta_{feedback=delayed}$ = .013, *SE* = .024, z = 0.54, p = .590). Girls were overall less accurate than boys ($\beta_{sex=girls}$ = -.172, *SE* = .070, z = 2.45, p = .014). Win-stay probability was predicted to be higher at wave 2 ($\beta_{wave=2}$ = .586, *SE* = .071, z = 8.22, p < .001) and with higher age at wave 1 ($\beta_{wave\ 1\ age}$ = .177, *SE* = .078, z = 2.27, p = .024), again without differences by feedback timing ($\beta_{feedback=delayed}$ = -.023, *SE* = .032, z = -0.69, p = .489). Win-stay probability was lower for girls compared to boys ($\beta_{sex=girls}$ = -.177, *SE* = .078, z = -2.27, p = .024). The predicted Lose-shift probability was lower at wave 2 compared to wave 1 ($\beta_{wave=2}$ = -.586, *SE* = .071, z = -8.22, p < .001) and with higher age at wave 1 ($\beta_{wave\ 1\ age}$ = -.177, *SE* = .078, z = 2.27, p = .024), but did not differ by feedback type ($\beta_{feedback=delayed}$ = .036, *SE* = .020, z = 1.74, p = .081) and sex ($\beta_{sex=girls}$ = .063, *SE* = .036, z = 1.76, p = .079). Taken together, children on average improved their accuracy, while win-stay probability increased and lose-shift probability decreased between waves. Girls were on average less accurate, showed reduced win-stay behavior and a smaller decrease in lose-shift probability between waves (Supplementary Table 1 and Supplementary Figure 3).

Reaction times were predicted to be faster at wave 2 compared to wave 1 ($\beta_{wave=2}$ = -218, *SE* = 22.7, t = -9.61, p < .001), but did not differ by wave 1 age ($\beta_{age\ wave\ 1}$ = -42.5, *SE* = 25.7, t = -1.66, p = .100), and they were faster for delayed compared to immediate feedback trials ($\beta_{feedback=delayed}$ = -14.0, *SE* = 6.61, t = -2.12, p = .036). Girls were not different compared to boys ($\beta_{sex=girls}$ = 23.5, *SE* = 25.7, t = 0.91, p = .362. To summarize the reaction time results, 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.



Supplementary Figure 3. Fixed effects plots of significant predictors across behavioral variables Accuracy (ACC), win-stay (WS), lose-shift (LS) and Reaction time (RT).

Supplementary Material 3: Behavioral and model results with reduced dataset

Behavioral results

To validate our results, we examined whether the poor learning performance of some of the children in the reinforcement learning task had an effect on our findings. Therefore, we repeated the analyses with a reduced dataset that excluded children performing below 50 % accuracy in their last 20 trials. 13 out of 140 children at wave 1 (54% girls), as well as 6 out of 126 at wave 2 (67% girls) did not reach the learning criterion and were excluded in the reduced dataset. We kept the same model structure to directly compare the results. The fixed effects remained unchanged in all models.

Using the reduced dataset, the learning accuracy model did not differ in the results, accuracy was predicted by wave ($\beta_{wave=2}$ = .493, *SE* = .063, z = 7.88, p < .001) and by wave 1 age ($\beta_{age\ wave\ 1}$ = .171, *SE* = .071, z = 7.88, p < .001), there were no differences by feedback timing ($\beta_{feedback=delayed}$ = .010, *SE* = .025, z = 0.39, p = .698), and girls were less accurate ($\beta_{sex=girls}$ = -.155, *SE* = .071, z = -2.18, p = .029). The win-stay model also did not differ in the results using the reduced dataset. Win-stay probability was again predicted to be higher at wave 2 ($\beta_{wave=2}$ = .533, *SE* = .074, z = 7.26, p < .001) and by higher wave 1 age ($\beta_{age\ wave\ 1}$ = .184, *SE* = .079, z = 2.32, p = .020), and girls had a lower win-stay probability ($\beta_{sex=girls}$ = -.161, *SE* = .080, z = -2.01, p = .045). The lose-shift model did not differ using the reduced dataset, lose-shift probability was lower at wave 2 ($\beta_{wave=2}$ = -.252, *SE* = .037, z = -6.84, p < .001), did not differ by feedback type ($\beta_{feedback=delayed}$ = .030, *SE* = .022, z = 1.35, p = .178) and sex ($\beta_{gender=girls}$ = .063, *SE* = .038, z = 1.66, p = .098), but the decrease in lose-shift behavior between waves again was smaller for girls ($\beta_{sex=girls\ X\ wave=2}$ = .067, *SE* = .034, z = 1.99, p = .047). Reaction times were faster at wave 2 compared to wave 1 ($\beta_{wave=2}$ = -218, *SE* = 23.4, t = -9.32, p < .001), they were not predicted by wave 1 age ($\beta_{age\ wave\ 1}$ = -43.0, *SE* = 26.3, t = -1.63, p = .105), and they were faster at delayed compared to immediate feedback ($\beta_{feedback=delayed}$ = -17.0, *SE* = 6.75, t = -3.16, p = .013). Girls were not different compared to boys ($\beta_{sex=girls}$ = 23.6, *SE* = 26.2, t = 0.90, p = .370). To conclude, behavioral results remained the same using the reduced dataset.

Model results

We repeated model comparison with the reduced dataset by excluding the elpd (expected log pointwise predictive density) from the poor learners (Supplementary Table 2).

Supplementary Table 2. Model comparison results with the reduced dataset.

Model	Parameters	$\Delta elpd_{loo}\ [SE]$	$\Sigma elpd_{loo}\ [mean]$	*pseudo-BMA+*
step 1: heuristic strategy vs value-based learning model				
vbm_1	1α, 1τ	0 [0]	-14717.7 [-0.47]	1
ws	$1\tau_{ws}$	-1296.2 [159.1]	-16013.9 [-0.51]	0
$wsls$	$1\tau_{wsls}$	-4164.3 [281.6]	-18882.1 [-0.61]	0
step 2: value-based learning model variants				
vbm_3	1α, 2τ	0 [0]	-14609.9 [-0.47]	0.78
vbm_7	1α, 2ρ	-3.71 [0.92]	-14613.6 [-0.47]	0.19
vbm_6	2α, 1ρ	-24.34 [12.46]	-14634.8 [-0.47]	<0.01
vbm_8	2α, 2ρ	-29.20 [18.43]	-14639.1 [-0.47]	0.02
vbm_4	2α, 2τ	-43.86 [17.83]	-14653.6 [-0.47]	<0.01
vbm_2	2α, 1τ	-45.08 [17.48]	-14655.0 [-0.47]	<0.01
vbm_5	1α, 1ρ	-57.65 [16.16]	-14667.6 [-0.47]	<0.01
vbm_1	1α, 1τ	-107.8 [20.73]	-14717.7 [-0.47]	<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). $\Delta elpd_{loo}[SE]$ = differences in Bayesian leave-one-out cross-validation estimate of the expected log pointwise predictive density relative to the winning model and its standard errors. $\Sigma elpd_{loo}[mean]$ = sums of expected log pointwise predictive density of all 31,116 trials, including all participants (except poor learners) and waves as well as trial means. *Pseudo-BMA+* = model weight for relative model evidence using Bayesian model averaging stabilized by Bayesian bootstrap using 100,000 iterations.

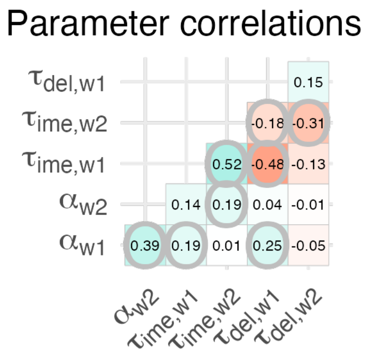
One may argue that this procedure is suboptimal, as the model parameters were fitted from the full dataset so that poor learners impacted the parameters of the remaining participants in hierarchical model estimation. However, fitting the reduced dataset only would have required a different model structure, as the amount of longitudinal datasets had been much smaller, and some participants would only have

wave 2 data. Since we used a wide prior for model estimation, the impact of poor learners on the group level is reduced. The model comparison of the reduced dataset did not differ from the results of the complete dataset. At the first step, children's learning behavior in the longitudinal data again can be better described by a value-based rather than by a heuristic strategy model. At the second step, comparison different value-based models, the winning model again suggests that feedback timing affected the inverse temperature, but not the learning rate or outcome sensitivity.
We did not find any deviations from the findings of the winning model when using the reduced dataset. The correlations between condition differences of inverse temperature and reaction times remained (r = -.288, $t(125)$ = -3.36, p = .001 at wave 1 and r = -.352, $t(118)$ = -4.09, p < .001 at wave 2).

Supplementary Material 4: Winning model parameter correlations

Parameter correlations of the winning model

Correlations between the model parameters learning rate and inverse temperature were only small (r = 0.19 – 0.25), which suggests relative independence of these parameters (Figure 3C). Negative correlations between feedback conditions (r = -0.31 – -0.48), captured by the inverse temperature, suggest individual differences feedback timing modulation. Positive correlations of the parameters across waves (r = 0.39 – 0.52) were moderate to large which suggest temporal stability and showed the appropriateness of our modeling endeavour to incorperate the within-subject data structure. Only inverse temperature for delayed feedback learning was not correlated across waves, which suggests greater temporal instability. 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, differences in the choice rule and reaction times 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.



Supplementary Figure 2. Parameter correlations of the winning model. Significant correlations are circled, *p*-values were adjusted for multiple comparisons using bonferroni correction.

Supplementary Material 5: Longitudinal brain-cognition links

Univariate latent change score (LCS) models

The model fit and model parameters of the univariate LCS models of our variables of interest (striatal volume, hippocampal volume, immediate learning score, delayed learning score) are summarized in Supplementary Table 3. Of note, learning scores were negatively covaried with sex at wave 1, suggesting reduced immediate learning scores ($\phi_{sex=girls,LS_{i,w1}}$= -0.20, z = -2.39, SE = 0.08, p = .017) and reduced delayed learning scores in girls ($\phi_{sex=girls,LS_{d,w1}}$= -0.17, z = -2.01, SE = 0.08, p = .044). When excluding poor learners, this covariation only remained significant for immediate learning scores ($\phi_{sex=girls,LS_{i,w1}}$= -0.18, z = -2.10, SE = 0.08, p = .035), but not for delayed learning scores ($\phi_{sex=girls,LS_{i,w1}}$= -0.15, z = -1.72, SE = 0.08, p = .085).

Confirmatory brain-cognition links with learning scores and episodic memory with reduced dataset

We fitted a fourvariate LCS model using the reduced dataset to check whether the reported results remained the same. The LCS again provided a good data fit (χ^2 (27) = 18.7, *CFI* = 1.00, *RMSEA* (*CI*) = 0 (0 – .030, *SRMR* = .053). Striatal volume at wave 1 again covaried with both immediate and delayed learning score ($\phi_{STR_{w1},LS_{i,w1}}$= 0.17, z = 2.19, SE = 0.08, p = .029 and $\phi_{STR_{w1},LS_{d,w1}}$= 0.16, z = 2.04, SE = 0.08, p = .041). Constraining the striatal association to immediate learning to 0 worsened model fit relative to the unrestricted model ($\Delta\chi^2$ (1) = 3.96, p = .047), but not when constraining the striatal association to delayed learning to 0 ($\Delta\chi^2$ (1) = 3.58, p = .058). Hippocampal volume did not covary with any learning scores in the reduced dataset ($\phi_{HPC_{w1},LS_{i,w1}}$ = 0.11, z = 1.52, SE = 0.08, p = .130 and $\phi_{HPC_{w1},LS_{d,w1}}$ = 0.14, z = 1.93, SE = 0.07, p = .054). We further examined whether in the reduced dataset the hippocampal contribution at delayed feedback would selectively enhance episodic memory. Episodic memory, as measured by individual corrected object recognition memory (hits – false alarms) of confident ("sure") ratings, again showed only at trend higher memory for delayed feedback ($\beta_{feedback=delayed}$ = .013, SE =.007, t = 1.87, p = .064). The results in the reduced dataset suggest that striatal associations are selective to immediate learning, while the hippocampus shows no associations to either learning conditions.

Supplementary Table 3. Model fit and parameter estimates of the univariate LCS models for immediate and delayed feedback PLS learning score as well as for striatal (STR) and hippocampal (HPC) brain volumes.

	$LS_{immediate}$	$LS_{delayed}$	*STR*	*HPC*
χ^2 (*df*)	1.75 (4)	1.25 (4)	1.61 (6)	1.77 (6)
RMSEA (*CI*)	0.08 (0 - 0.08)	0 (0 – 0.07)	0 (0 - 0)	0 (0 - 0.02)
SRMR, CFI	0.03, 1.00	0.03, 1.00	0.03, 1.00	0.03, 1.00
Mean change μ_Δ	0.74** (0.09)	0.73** (0.08)	0.06* (0.03)	0.37** (0.05)
w1 variance σ_β	0.99** (0.08)	0.99** (0.07)	0.51** (0.07)	0.46** (0.06)
Change variance σ_Δ	0.94** (0.10)	0.89** (0.10)	0.07** (0.02)	0.18* (0.08)
Intercept-change regression δ	-0.69** (0.08)	-0.73** (0.08)	-0.04 (0.04)	-0.12* (0.04)
Age onto Intercept	-0.07 (0.08)	0.11 (0.08)	0.02 (0.09)	0.15 (0.08)
Sex onto Intercept	-0.20* (0.08)	-0.17* (0.08)	-0.05 (0.09)	-0.09 (0.09)
eTIV onto Intercept	–	–	0.67** (0.09)	0.62** (0.10)

Standard errors in parentheses. ** denotes significance at α < .001, * at α < .05. sex coded as 1 = girls, -1 = boys.

Exploratory brain-cognition links with model parameters

The model parameters all showed significant mean change and variance (learning rate: $\mu_{\Delta\alpha}$ = 1.29, z = 7.41, *SE* = 0.17, $p < .001$, $\sigma_{\Delta\alpha}$= 3.73, z = 6.77, *SE* = 0.55, $p < .001$; immediate inverse temperature: $\mu_{\Delta\tau_i}$ = 0.82, z = 9.65, *SE* = 0.09, $p < .001$, $\sigma_{\Delta\tau_i}$= 0.97, z = 4.12, *SE* = 0.24, $p < .001$; delayed inverse temperature: $\mu_{\Delta\tau_d}$ = 0.84, z = 3.91, *SE* = 0.08, $p < .001$, $\sigma_{\Delta\tau_d}$= 0.84, z = 3.91, *SE* = 0.22, $p < .001$). To further understand how the found links between striatal volumes and immediate learning and between hippocampal volumes and delayed learning could be understood as effects of the model parameters, we compiled a five-variate model including brain volumes, learning rates (α) and inverse temperature (τ) for immediate and delayed learning. The LCS again provided a good data fit (χ^2 (25) = 15.8, *CFI* = 1.00, *RMSEA* (*CI*) = 0 (0 – .023, *SRMR* = .040).

For hippocampal volume, we found a positive covariance with delayed inverse temperature at wave 1($\phi_{HC_{w1},\tau_{del,w1}}$= 0.13, z = 2.30, *SE* = 0.06, p = .021), whereas striatal volume positively covaried with learning rate at ($\phi_{STR_{w1},\alpha_{w1}}$= 0.15, z = 2.05, *SE* = 0.08, p = .041). The striatal link to learning rate however was diminished when excluding children below the learning criterion. Longitudinally, striatal volume at wave 1 further predicted positive gains in learning rate ($\beta_{STR_{w1},\Delta\alpha}$= 0.44, z = 2.25, *SE* = 0.20, p = .024). Changes in learning rate covaried positively with changes in immediate inverse temperature ($\phi_{\Delta STR,\Delta\tau_i}$= 0.35, z = 2.46, *SE* = 0.14, p = .014), while changes in immediate inverse temperature covaried negatively with changes in delayed inverse temperature ($\phi_{\Delta\tau_i,\Delta\tau_d}$= -0.28, z = -3.60, *SE* = 0.08, $p < .001$). Immediate inverse temperature at wave 1 predicted negative striatal volume change ($\beta_{\tau_{i,w1},\Delta STR}$= -0.09, z = -2.38, *SE* = 0.04, p = .017), while delayed inverse temperature at wave 1 predicted negative change in hippocampal volume ($\beta_{\tau_{d,w1},\Delta HPC}$= -0.08, z = -2.06, *SE* = 0.04, p = .039) in the reduced sample, but not in the full sample. Taken together, while hippocampal volume was only linked to delayed inverse temperature at wave 1, striatal volume was linked to learning rate at wave 1 and was predictive of learning rate development. Further, there was evidence that inverse temperature was predictive of brain volume change in line with the hypothesized brain-cognition links. The inverse temperature between delayed and immediate feedback showed diverging changes, in which the change in immediate inverse temperate was similar to that of learning rate, but dissimilar to that of delayed inverse temperature. This suggests that the hippocampus might be uniquely associated with inverse temperature during delayed learning, whereas the striatum was linked to learning rates, inverse temperature and suggest a stronger contribution to the longitudinal change of learning function in general.

Confirmatory brain-cognition links with learning scores using the second best fitting model

We fitted a fourvariate LCS model using the second best fitting model to check whether separating outcome sensitivity by feedback timing would show results comparable to those of the winning model that separated inverse temperature by immediate and delayed feedback condition. Using the model-derived learning scores from the second best fitting model, our LCS model again provided a good data fit (χ^2 (27) = 10.1, *CFI* = 1.00, *RMSEA* (*CI*) = 0 (0 – 0, *SRMR* = .042). However, the brain-cognition links at baseline were not significant for both striatal volume ($\phi_{STR_{w1},LS_{i,w1}}$= 0.14, *z* = 1.66, *SE* = 0.09, *p* = .098 and $\phi_{STR_{w1},LS_{d,w1}}$= 0.14, *z* = 1.55, *SE* = 0.09, *p* = .121) and hippocampal volume ($\phi_{HPC_{w1},LS_{i,w1}}$= 0.09, *z* = 1.04, *SE* = 0.09, *p* = .297 and $\phi_{HPC_{w1},LS_{d,w1}}$= 0.11, *z* = 1.22, *SE* = 0.09, *p* = .222), suggesting no brain-cognition links at wave 1. Longitudinally, striatal volumes predicted larger gains in immediate learning scores ($\beta_{STR_{w1},\Delta ls_i}$= 0.17, *z* = 1.97, *SE* = 0.08, *p* = .049), but this effect diminished when excluding poor learners ($\beta_{STR_{w1},\Delta ls_i}$= 0.11, *z* = 1.35, *SE* = 0.08, *p* = .177). The failure to capture brain-cognition links and the relatively lower model evidence compared to the winning model during model comparison overall suggests that modulations by feedback timing could be captured better by the decision-related parameter inverse temperature rather than by the valuation-related parameter outcome sensitivity.