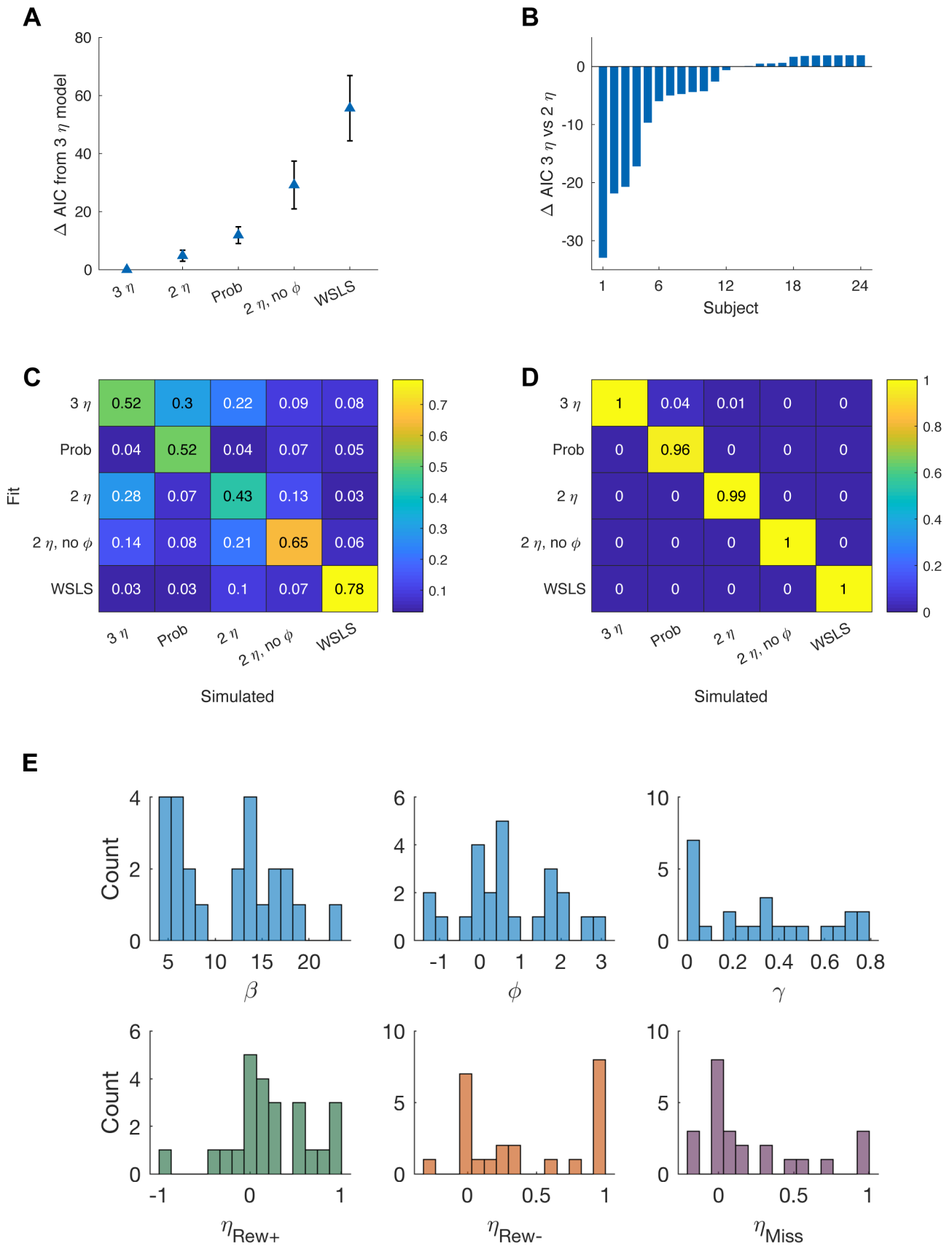

Figure S1. Model Comparisons and parameter values; related to Figures 2, 4, and STAR Methods. (**A**) Increase in Akaike information criterion (AIC) score in tested models, relative to the winning model (3

η). The 3 η model outperformed all other models (one sample *t*-tests, all *p*'s < 0.02). Lower values reflect better model fits. (**B**) Individual subject fits, comparing the best fit model (3 η) versus the second best model (2 η). Negative values indicate better fit with the 3 η model. (**C**-**D**) Confusion matrices from the simulate-and-fit analysis, with the ground-truth simulated model on the x-axis and the model used to fit the simulated data on the y-axis. Color indicates the fraction of simulations best fit by each model. Color in (**D**) uses Akaike weights (an approximation of the conditional probability of one model over the others) over summed AIC values across the sample to compute the fitted model's probability of reflecting the underlying simulated model. (We note that summing AIC values across subject samples tends to inflate model fit differences). Error bars = 1 s.e.m. (**E**) Binned distributions of fitted model parameters from the winning model (3 η). Clockwise from top left: Inverse temperature (β), persistence (Φ), decay (γ), learning rate for Miss trials (η_{Miss}), learning rate for Rew- trials (η_{Rew-}), and learning rate for Rew+ trials (η_{Rew+}). A Gamma (2,3) prior was imposed for the inverse temperature parameter during fitting to discourage extreme values. The persistence parameter, representing choice "stickiness" after Miss trials, was significantly greater than 0 (p = 0.006). Note that high values in learning rates (i.e., ~1.0) tended to represent true behavioral strategies in our task; for instance, the mean switching rate after Rew- trials was 88% for the subset of participants with fitted η_{Rew-} values of 1. Negative η values were allowed during fitting due to a lack of *a priori* predictions for the behavior of the η_{Miss} parameter.

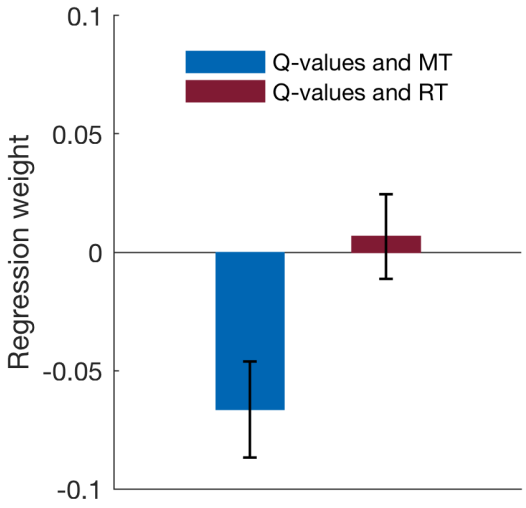

Figure S2. Choice value, movement time, and reaction time; related to Figure 2. Trial-by-trial Q-values derived from the fitted reinforcement learning model (3 η), movement times (MT), and reaction times (RT) were extracted, de-trended, and z-scored for each subject. Regression weights were computed for the effect of Q-values on MT (blue) and RT (red). Error bars = 1 s.e.m.

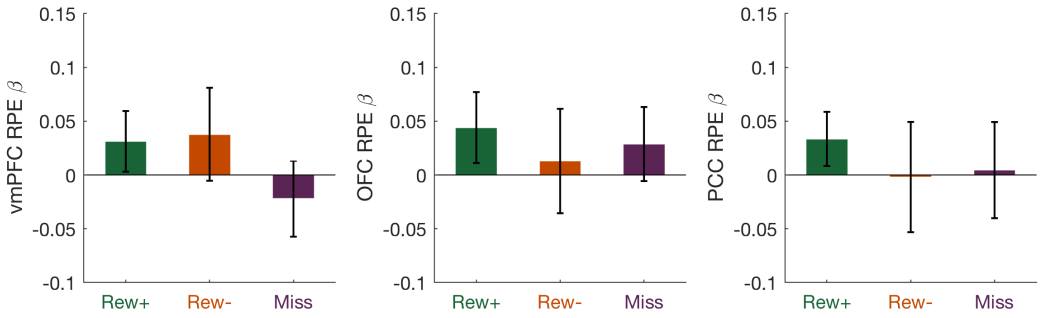

Figure S3: **RPE effects in non-striatal ROIs; related to Figures 3 and 4.** Beta values for RPE encoding in the non-striatal ROIs revealed by the Rew+ contrast in the first GLM (see Figure 3 in main text). vmPFC = ventromedial prefrontal cortex; OFC = orbitofrontal cortex; PCC = posterior cingulate cortex. Error bars = 1 s.e.m.

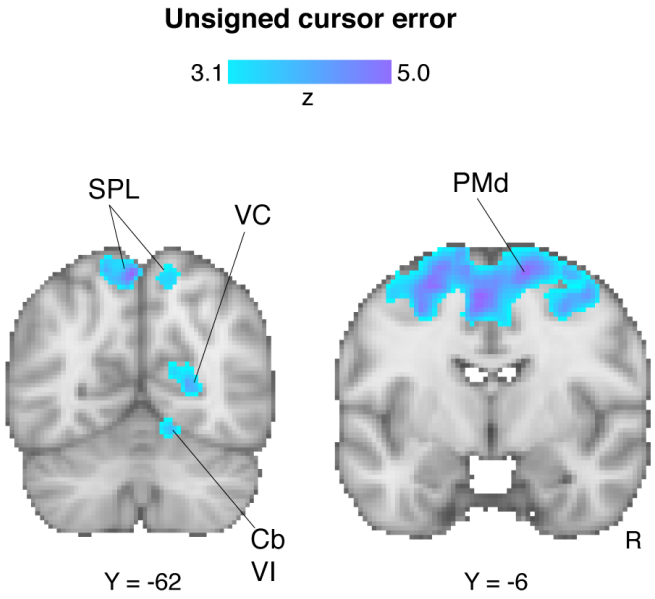

Figure S4: **Whole-brain main effects for unsigned cursor error regressor; related to Figure 3.** Significant activations produced by the parametric regressor representing the unsigned cursor error on Miss trials. SPL = superior parietal lobe; VC = visual cortex; Cb VI = cerebellum, lobule six; PMd = dorsal premotor cortex.

Analysis/Region	x (mm)	y (mm)	z (mm)	# voxels
Rew+ > (Rew- & Miss)				
Striatum	5	11	-5	704
vmPFC	0	46	2	3017
L OFC	-36	34	-12	633
PCC	1	-50	31	1199
Miss > Rew-				
SMA/PMC/ACC	10	-2	61	1787
R IPL	60	-24	33	1411
L IPL	-55	-25	27	793
Rew- > Miss				
R LOC/FG	34	-74	-3	1161
L LOC/FG	-32	-82	-7	1221
Unsigned Error				
M1/PMC/SMA	-1	-22	56	10309
V1/R Cb	13	-80	10	2823

Table S1: **Significant Clusters; related to Figure 3**. All clusters survived cluster correction at the $p < 0.05$ level (FSL FLAME 1) with a cluster-forming threshold of $p < 0.001$. Coordinates are in MNI space and correspond to the cluster's center of gravity. vmPFC = ventromedial prefrontal cortex; OFC = orbitofrontal cortex; PCC = posterior cingulate cortex; SMA = supplementary motor area; ACC = anterior cingulate cortex; IPL = inferior parietal lobule; IFG = inferior frontal gyrus; M1 = primary motor cortex; PMC = premotor cortex; V1 = primary visual cortex; LOC = lateral occipital cortex; FG = fusiform gyrus; Cb = cerebellum.

Model	Parameter								
	η Rew+	***η Rew-***	***η Miss***	***η Rew-/ Miss***	***η Prob***	***η Payoff***	***β***	***Φ***	**γ**
3 *η* (“Gating”)	✓	✓	✓				✓	✓	✓
Probability					✓	✓	✓	✓	✓
2 *η*	✓			✓			✓	✓	✓
2 *η* (no Φ)	✓			✓			✓		✓
WSLS (“win-stay-lose-switch”)							✓	✓	✓

Table S2: **Models Tested and Parameters in each; related to Figures 2, S1, and S2**. *η* = learning rate (with context-specific assignment below); *β* = inverse temperature; *Φ* = persistence; *γ* = decay.