Supplementary Information

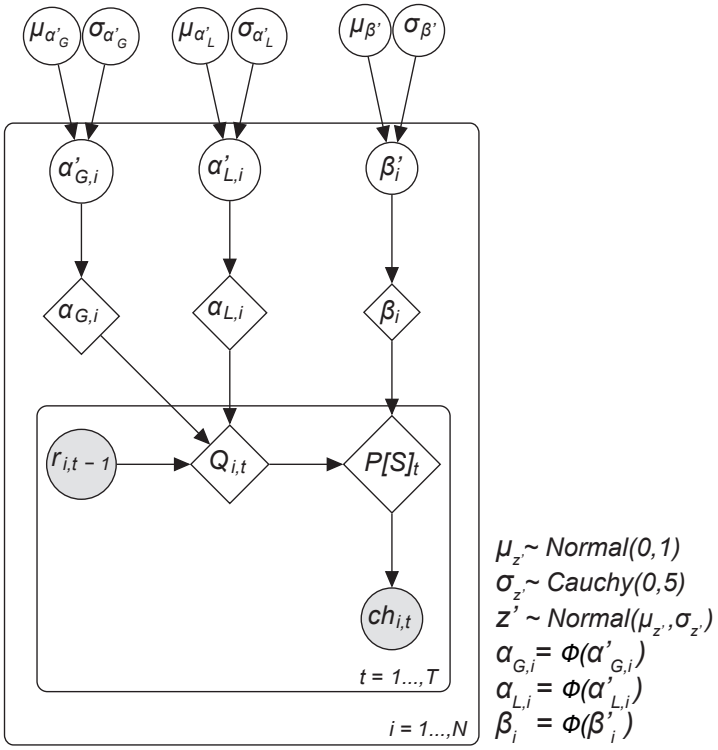
Spontaneous eye blink rate predicts individual differences in exploration and exploitation during reinforcement learning

Joanne C. Van Slooten[1]* *Sara Jahfari*[2,3] *Jan Theeuwes*[1]

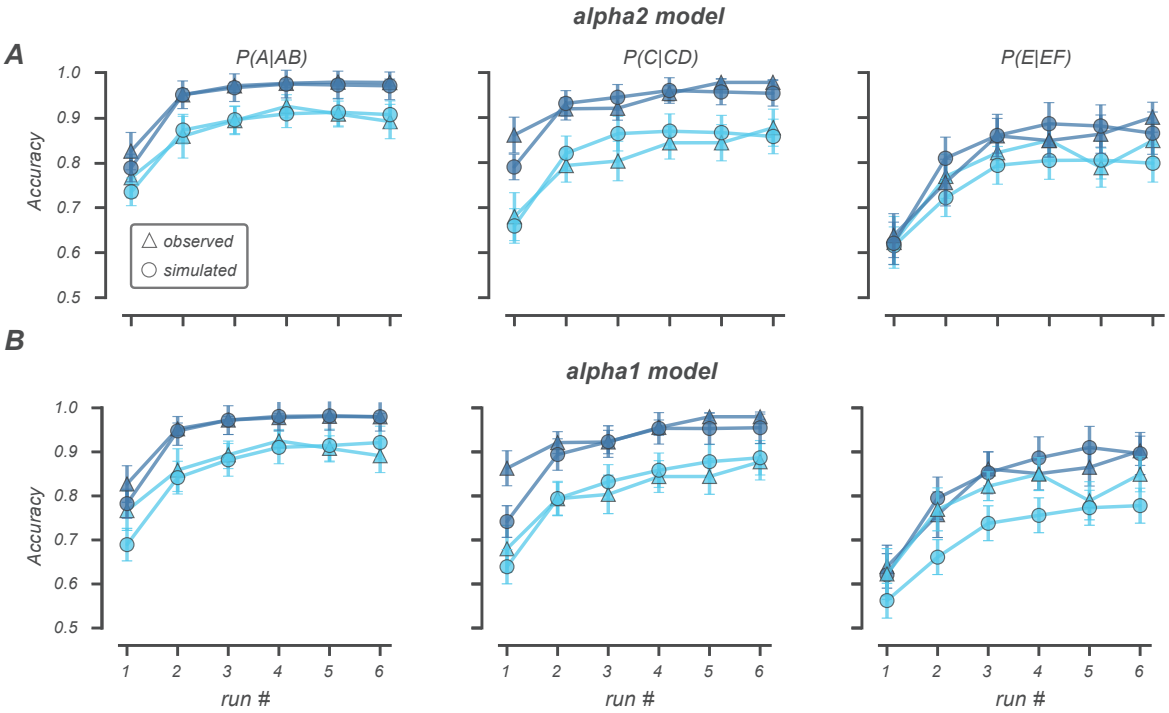
[1] *Department of Experimental and Applied Psychology, Vrije Universtiteit Amsterdam, Amsterdam, The Netherlands.*

[2] *Spinoza Centre for Neuroimaging, Royal Academy of Sciences, Amsterdam, The Netherlands.*

[3] *Department of Psychology, University of Amsterdam, Amsterdam, The Netherlands*



Supplementary Figure 1. **Graphical representation of the hierarchical Bayesian Q-learning model.** The inner plane represents within-subject trial-by-trial RL behaviour. Variables r_i *(t-1)* (outcome for participant *i* on trial *t-1*) and $ch_i(t)$ (choice of participant *i* on trial *t*) were obtained from the behavioural data. The outer plane represents per-participant parameter estimates α_{Gi} (α_{Gain} participant *i*), α_{Li} (α_{Loss} participant *i*) and β_i (β participant *i*) that were fit separately for participants in the low and high sEBR group. Per-participant parameter estimates were modelled using a probit transform z'_i ($\alpha'_{Gi}, \alpha'_{Li}, \beta'_i$). z'_i were drawn from group-level normal distributions with mean $\mu_{z'}$ and standard deviation $\sigma_{z'}$. The outermost layer represents group-level mean and standard deviations of the Q-learning model parameters. A normal prior was assigned to all group-level means, $\mu_{z'} \sim N(0,1)$, and a half-Cauchy prior to all group-level standard deviations, $\sigma_{z'} \sim Cauchy(0,5)$. A weakly informative prior such as this is recommended in small sample sizes to reduce the influence of the priors on posterior distributions[77]. Shaded variables are obtained from the behavioural data and used to fit the model. Diamond shaped nodes are deterministic, as they are derived from the model fit. Circular unshaded nodes indicate continuous variables. Arrows indicate dependencies between variables. $\Phi()$ represents the probit transform.



Supplementary Figure 2. Posterior predictive checks of mean choice accuracy for the alpha2 and alpha1 Q-learning model. Participants' mean choice accuracy data (observed; *triangle markers*) were compared to the mean accuracy of the posterior predictive distribution (simulated; *circle markers*), separately for the different option pairs (AB, CD and EF) and for six bins of trials within the learning phase. The alpha1 model is the simplest Q-learning model with one learning rate (α) that is agnostic to the sign of reward prediction errors and one explore-exploit parameter (β). The alpha2 model has a separate α for learning from positive and negative prediction errors. Model comparison using PSIS-LOO indicated that the alpha2 model best fit the data (elpd difference = 289.23, SD=51.98). This becomes evident when comparing both models' simulated choices for the EF pair, where the alpha1 model consistently underestimates EF choice accuracy of high sEBR individuals. Dark blue = low sEBR; light blue = high sEBR. Error bars are SEM.

Model Comparison

Models	P(M)	P(M | data)	BF_M	BF_{10}	R^2
Null model	0.125	0.007	0.052	1.000	0.000
β	0.125	0.353	3.814	47.859	0.297
$\beta + \alpha_{Gain}$	0.125	0.314	3.205	42.618	0.349
$\beta + \alpha_{Loss}$	0.125	0.196	1.710	26.634	0.327
$\beta + \alpha_{Gain} + \alpha_{Loss}$	0.125	0.116	0.917	15.722	0.352
α_{Loss}	0.125	0.006	0.045	0.868	0.071
α_{Gain}	0.125	0.004	0.031	0.598	0.045
$\alpha_{Gain} + \alpha_{Loss}$	0.125	0.003	0.020	0.389	0.077

Supplementary Table 1. Bayesian linear regression analysis of Q-learning model parameter modes on sEBR. Compared to the null model, the data provide strong evidence in favour of the model in which the β-parameter explains individual variability in sEBR.

Model Comparison

Models	P(M)	P(M | data)	BF_M	BF_{10}	R^2
Null model	0.125	0.048	0.350	1.000	0.000
$\beta + \alpha_{Loss}$	0.125	0.338	3.574	7.108	0.260
α_{Loss}	0.125	0.168	1.411	3.527	0.159
$\beta + \alpha_{Gain} + \alpha_{Loss}$	0.125	0.134	1.079	2.809	0.260
β	0.125	0.133	1.069	2.786	0.145
$\beta + \alpha_{Gain}$	0.125	0.092	0.713	1.944	0.185
$\alpha_{Gain} + \alpha_{Loss}$	0.125	0.064	0.475	1.336	0.161
α_{Gain}	0.125	0.025	0.177	0.518	0.035

Supplementary Table 2. Bayesian linear regression analysis of Q-learning model parameter modes on avoidance accuracy in the transfer phase. Compared to the null model, the data provide moderate evidence in favour of the model in which both the β-parameter and α_{Loss}-parameter explain individual variability in avoidance behavior in the transfer phase.

Weights matrix

Variable	Network					
	α_{Gain}	α_{Loss}	approach	avoid	β	sEBR
α_{Gain}	0.000	0.522*	-0.016	-4.340e -4	-0.220	-0.192
α_{Loss}	0.522*	0.000	0.026	0.302	0.055	-0.053
approach	-0.016	0.026	0.000	-0.043	0.139	0.037
avoid	-4.340e -4	0.302	-0.043	0.000	0.278	-0.034
β	-0.220	0.055	0.139	0.278	0.000	-0.515*
sEBR	-0.192	-0.053	0.037	-0.034	-0.515*	0.000

Supplementary Table 3. Partial correlation weights matrix of all network variables. Asterisks indicate significant partial correlations between variables in the network.

Supplementary references

77. Ahn, W-Y., Haines, N. & Zhang, L. Revealing neurocomputational mechanisms of reinforcement learning and decision-making with the hBayesDM Package. *Computational Psychiatry, **1**, 24-57 (2017).*