S3 DATA

Best-fit Parameters of the Learning model

The reinforcement-based state-space model has three free parameters: \( \alpha, \sigma_m, \) and \( \sigma_e. \) To find the best-fit parameters and their corresponding confidence intervals, we performed a bootstrap optimization fitting procedure. Fitting was done using a two-step process. For all optimizations, we used a Nelder-Mead optimization routine contained in the SciPy module (optimize, minimization) in Python.

Step 1: Finding Initial Guesses with Behavioural Estimates and ‘Warm-start’ Optimization

We used a ‘warm-start’ bootstrap optimization procedure to find an initial guess of \( \alpha, \) where \( \sigma_m \) and \( \sigma_e \) were held as constants (0.85 and 0.99, respectively). Estimates of \( \sigma_m \) and \( \sigma_e \) were found using a trial-by-trial analysis (see S2 DATA).

For each bootstrap \( (N = 1000), \) we resampled with replacement separately for participants experiencing a steep landscape and participants experiencing a shallow landscape. For each group, we averaged across participants to find the mean reach angle of each trial. Then we fit an exponential function (Eq. 6) to these two experimental groups. Consequently, we obtained a time-constant that defines the rate of learning for participants experiencing the steep reinforcement landscape \( (\lambda_{\text{behaviour} \text{steep}}) \) and a time-constant for participants experiencing the shallow reinforcement landscape \( (\lambda_{\text{behaviour} \text{shallow}}). \) For each exponential function, we also found the best-fit term that defines the asymptotic reach angle \( (a). \)

Simultaneously, with our learning model (Eq. 1) we simulated 2000 ‘individuals’ experiencing the steep reinforcement landscape and another 2000 ‘individuals’ experiencing the shallow reinforcement landscape. Like the behavioural data, we averaged across simulated individuals to obtain the mean reach angle for each trial. We then fit an exponential function to the simulated individuals that experienced the steep reinforcement landscape and the simulated individuals that experienced a shallow reinforcement landscape.
landscape. This yielded a time-constant that defines the rate of learning for simulated individuals experiencing the steep reinforcement landscape ($\lambda_{steep}^{model}$) and a time-constant for simulated individuals experiencing the shallow reinforcement landscape ($\lambda_{shallow}^{model}$). We also found the best-fit term that defines the asymptotic reach angle ($a$) for each exponential function.

For each bootstrap ($i$), we then found the best-fit $\alpha^i$ by minimizing error between the behavioural time-constants and model time-constants according to following equation:

$$\alpha^i = \arg \min_{\alpha} \left\{ \left| \left( \lambda_{steep}^{behaviour} - \lambda_{steep}^{model} \right) \right|^2 + \left| \left( \lambda_{shallow}^{behaviour} - \lambda_{shallow}^{model} \right) \right|^2 \right\} \quad (C1).$$

This process produces a single best-fit $\alpha^i$ per bootstrap.

Repeating this process 1000 times yields a distribution of $\alpha_i$. We then took the mean of this distribution to find a single $\alpha$, which was then used as the initial guess for the best-fit optimization procedure described below.

**Step 2: Finding the Best-fit parameters and their 95th percentile confidence intervals**

The procedure for finding the best-fit parameters was the same as described above, except that all three parameters ($\alpha, \sigma_m, \sigma_e$) of the model were free to vary:

$$\alpha_{opt}, \sigma_{m, opt}, \sigma_{e, opt} = \arg \min_{\alpha, \sigma_m, \sigma_e} \left\{ \left| \left( \lambda_{steep}^{behaviour} - \lambda_{steep}^{model} \right) \right|^2 + \left| \left( \lambda_{shallow}^{behaviour} - \lambda_{shallow}^{model} \right) \right|^2 \right\} \quad (C2).$$

From each of the resulting parameter distributions, we used the mean as the best-fit value for each parameter. Further, an advantage of this bootstrap optimization procedure is that confidence intervals can be found from these parameter distributions. The best-fit values (and corresponding 95th percentile confidence intervals) of $\alpha$, $\sigma_m$, and $\sigma_e$ were $0.40(0.25, 0.63)$, $0.9(0.813, 1.02)$, and $0.81(0.63, 0.97)$, respectively.