Value learning model

Matteo Lisi

2025-03-06

The model assumes that on every trial t the child makes a choice action $a \in \{1, 2\}$ and obtains a "reward" $r \in \{0, 1\}$. It is assumed that the participant maintains and updates their estimate of the value (that is the expected, long-run, reward) of each choice option — the so-called Q-values. These Q-values are updated after each choice according to

$$Q_{t+1}(a) = Q_t(a) + \eta\, \delta_t$$

where η is the learning rate and δ_t is the reward prediction error at trial t, calculated as

$$\delta_t = r_t - Q_t(a)$$

A logistic sigmoid (softmax) function is used to transform the Q-values of each symbol into the probability that the participant choose it in a given trial t

$$P_t(a) = \frac{e^{\beta\, Q_t(a)}}{\sum_i^2 e^{\beta\, Q_t(i)}}$$

where β is an "inverse temperature" parameter that controls the randomness of the choices.