Supporting Information

Human collective intelligence under dual exploration-exploitation dilemmas

Toyokawa, W., Kim, H. & Kameda, T.

Supporting methods

*Defining exploration and exploitation*

We followed the traditional definitions of exploration and exploitation in the reinforcement learning literature [1]. Individual *x* facing a choice at round *t* has a unique vector , whose elements represent expected payoffs of the 30 options based on *x*’s cumulative knowledge up to round *t*. As *x*’s experience accumulates, *Q* (*x, t*) is updated by an averaging rule, in which  is equal to the mean payoff that *x* has earned from option *i* until round *t*. The averaging updating rule is a suitable assumption of the action-value function in a stationary multi-armed bandit task [1]. This model assumes that players can hold up to 30 values for the different options simultaneously in working memory and calculate a new average for each option. Although such assumptions are too strong for human cognitive capacity, we believe that this model can provide a first approximation to distinguish between exploration and exploitation in a stationary multi-armed bandit task. We categorized choice behaviour as “exploitation” when the individual *x* chose option *i* with max  in round *t* + 1 (i.e., the greedy choice). Otherwise, we categorized the choice as “exploration.” The initial *Q* values were set to zero, i.e., , because *x* had no prior knowledge about the expected payoff from each option. For the first round (*t* = 1) only, we categorized all choices as “exploration.”

*Comparing exploration frequencies between the individual and group conditions*

To compare the frequencies of exploratory choices over all 100 rounds between the individual and group conditions, we used a hierarchical Bayesian method. We modelled the probability of individual *x* in group engaging in exploration () as a mixed logit model:

 (S1)

where λ1, 0 indicates an intercept, λ1, 1 indicates a fixed effect for the condition (dummy coded: individual condition = 0; group condition = 1), and local parameters *ε*1 (*x*) and indicate random effects specific to the individual and to the group respectively ( and ).

To estimate the four parameters (λ1, 0, λ1, 1, σ1, and Σ1), we set the uninformed priors of λ1, 0 and λ1, 1 to a normal distribution with mean = 0 and variance = 104 (i.e., a very stretched distribution), and those of the hyper parameters (σ1 and Σ1) to a uniform distribution [0, 104]. We conducted Markov Chain Monte Carlo (MCMC) simulations for 5 independent sequences. The number of total iterations per chain was 240,000 (thinning rate was 250), and first 40,000 steps were discarded from analysis because of initial-value dependencies. We used 95% Bayesian credible intervals to determine the significance of each parameter.

*Comparing exploration frequencies between the frequency-only and frequency-plus-evaluation sub-conditions*

To compare the frequencies of exploratory choices over all 100 rounds between the two sub-conditions, we used a hierarchical Bayesian method. We modelled the probability of individual *x* in group engaging in exploration () as a mixed logit model:

 (S2)

where λ2, 0 indicates an intercept, λ2, 1 indicates a fixed effect for the sub-condition (dummy coded: *frequency-only* sub-condition = 0; *frequency-plus-evaluation* sub-condition = 1), and local parameters *ε*2 (*x*) and indicate random effects specific to the individual and to the group respectively ( and ).

Because the structure of the model was identical to the equation S1, the MCMC protocol was the same as described above.

*Analysis of the causality between exploration and evaluation*

To analyze factors affecting individual exploratory choices, we constructed two models. One model posits that an individual’s exploration probability at round *t* + 1 is influenced by the total number of ratings contributed in the group at round *t* (evaluation effect model). The converse model posits that the individual probability of contributing rating information at the feedback stage of round *t* is influenced by whether or not she/he has engaged in exploration at the choice stage of round *t* (exploration effect model).

For the evaluation effect model, the probability of individual *x* in group engaging in exploration at round *t* + 1 () was modelled:

 (S3)

where α1 indicates an intercept, β1 indicates a fixed effect for the total number of ratings contributed in the group at the preceding round (), and local parameters *ε*3(*x*) and indicate random effects specific to the individual and to the group respectively ( and ).

To estimate the four parameters (*α*1, *β*1, σ3, and Σ3), we set the uninformed priors of *α*1 and *β*1 to a normal distribution with mean = 0 and variance = 104 (i.e., a very stretched uniform distribution), and those of the hyper parameters (*σ*3 and Σ3) to a uniform distribution [0, 104]. We conducted MCMC simulations for 5 independent sequences. The number of total iterations per chain was 63,000 (thinning rate was 300), and first 3,000 steps were discarded from analysis because of initial-value dependencies. We used 95% credible intervals to determine the significance of each parameter.

For the exploration effect model, the probability of contributing evaluation at the feedback stage of round *t* was modelled:

 (S4)

where *ωx,g,t* indicates the probability of individual *x* contributing her/his evaluation at round *t*, α2 for an intercept, β2 for a fixed effect of exploration in the current period *t* (dummy coded: exploration = 1, exploitation = 0). Local parameters *ε*4(*x*) and *Ε*4(*g*) indicate random effects specific to the individual and to the group respectively (and ). Because the structure of the model was identical to the evaluation effect model, the MCMC protocol was the same as described above.

References

[1] Sutton, R. S. & Barto, A. G. *Reinforcement Learning: An Introduction*. (MIT press, Cambridge, 1998).