P-402

Does the prefrontal cortex guide optimal foraging?

Min Jun Kang[1], Jae Hoon Shin[1], and Sang Wan Lee[1,2,3,4]*

[1]Department of Bio and Brain Engineering, KAIST, [2]Department of Brain and Cognitive Sciences, KAIST, [3]Program of Brain and Cognitive Engineering, KAIST, [4]KAIST Center for Neuroscience-inspired AI, KAIST

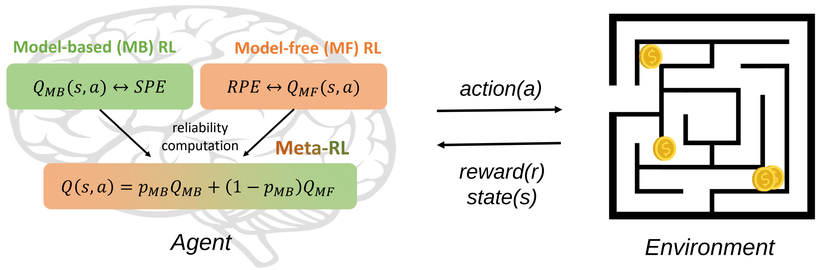
Center for Neuroscience-inspired Artificial Intelligence

Laboratory for Brain and Machine Intelligence

1. Introduction

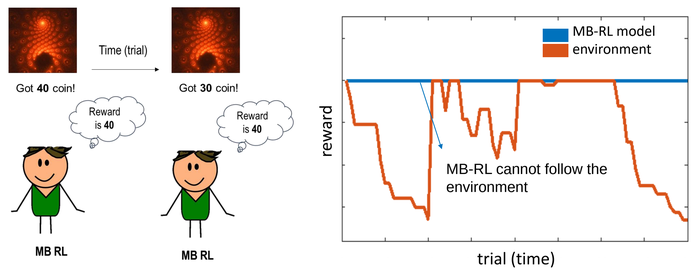
☐ Meta-RL can explain human reinforcement learning (RL)

(Lee et al., 2014)



☐ However, Meta-RL often fails to accommodate reward changes

: Even though MB-RL possesses flexibility due to its model,
it cannot follow the implicit change (e.g. rewards) in the environment

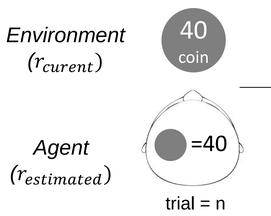


Q. Then, what strategy would a model-based system use to adapt to a dynamic environment?

2. Hypothesis Model Suggestion

☐ Model-based system would use the temporal difference (TD) rule to update the internal reward estimation

How to update the estimation?



Hypothesis reward estimation

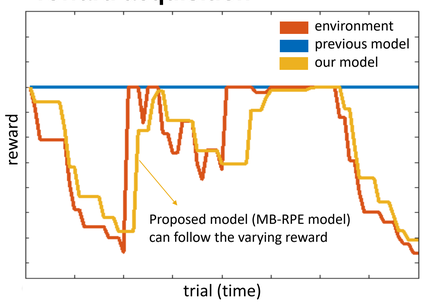
Step1. Calculate prediction error
$\delta_{MB-RPE} = r_{estimated}(s) - r_{current}(s)$

Step2. Update the estimated reward
$r_{estimated}(s) = r_{estimated}(s) + \alpha\delta_{MB-RPE}$

δ_{MB-RPE}: model-based reward prediction error (**MB-RPE**)
α: learning rate / s: state

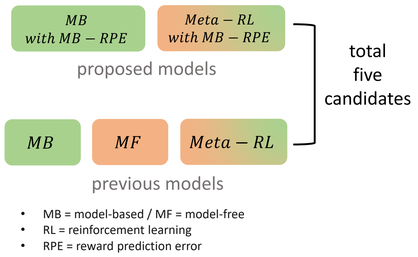
-> Using the TD rule, a model-based system would generate MB-RPE and update the estimated rewards

☐ Our model explains subject's reward acquisition



☐ Model comparison

Which model best explains human behavior patterns?



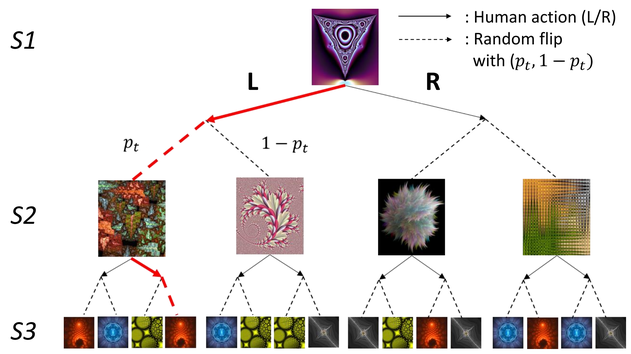
- MB = model-based / MF = model-free
- RL = reinforcement learning
- RPE = reward prediction error

☐ Research purposes

- **Test whether *Meta-RL with MB-RPE* model best explains human behavior**
- **Find neural evidence that brains generate MB-RPE signals**

3. Task Design

☐ Two-step Markov decision task



- Task with context θ_t: $\{p_t, r_t(\blacksquare), r_t(\blacksquare), r_t(\blacksquare), r_t(\blacksquare)\}$
- Subjects make two consecutive actions (L/R) to get the reward
- Next states are also determined by an internal variable (p_t)
- State transition probability (p_t) and the final rewards (r_t)
  -> map parameters that can change over trials
- Visited state reward decays (default setting)

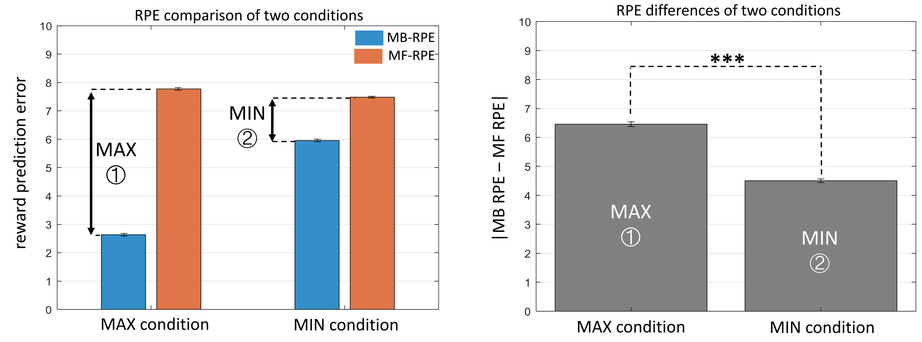
☐ Foraging rules for task design: two task conditions (MAX, MIN)

To simulate MB, MF RL either compete or cooperate, we designed two task conditions

Goals for each condition	Possible task controls per each trial	Results of task controls (20 trials)
MAX condition	1: do nothing	**MAX condition**
Maximize |MB RPE – MF PRE|	2: shift p_t from 0.9 to 0.5 (or vice versa)	1 3 1 4 3 1 3 3 3 3 4 1 1 4 4 1 5 3 3 1
MIN condition	3: reward recovery of visited state	**MIN condition**
Minimize |MB RPE – MF RPE|	4: reward recovery of unvisited state	1 1 1 1 1 1 1 1 1 1 2 1 2 1 2 2 2 1 2

☐ RPEs comparing in MAX and MIN conditions

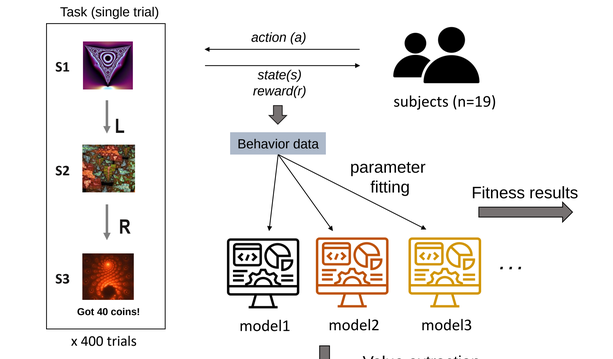
paired t-test (*: p<0.05, **:p<1e-2, ***: p<1e-3)



4. Results

☐ Task condition validation after model-fitting

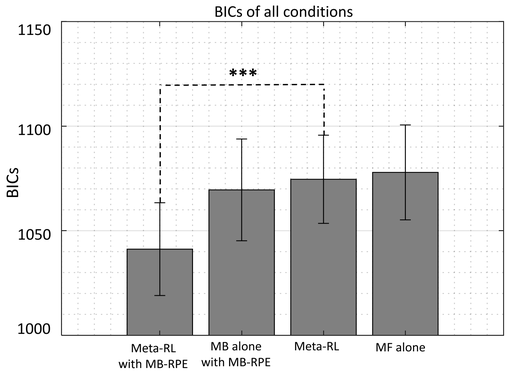
paired t-test (*: p<0.05, **:p<1e-2, ***: p<1e-3)



Model fitting shows that two task conditions were controlled as intended initially

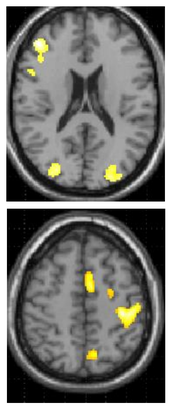
☐ Model fitness analysis (n = 19)

BIC = Bayesian Information Criterion
(lower BIC = more preferred model)



☐ Neural data GLM analysis (n = 25)

- different subjects from behavior analysis
- 25 subjects, ten females, 23.7±3.8 years



1. negative MB-RPE β map
- z = 20mm
- p < 0.05, FWE
- dlPFC activation (Z-score: 5.99)
- (x,y,z)=(46,40,20)

2. positive MB-RPE β map
- z = 54mm
- p < 0.001
- Left SMC activation (Z-score: 12.24)
- (x,y,z)=(-2, 4, 58)

- **Found neural evidence that the prefrontal cortex guides foraging**
- ***Meta-RL with MB-RPE* explains subjects' behavior patterns best**
- ***Meta-RL with MB-RPE* shows consistent accountability regardless of task conditions**

5. Conclusions

- We proposed a strategy for the **model-based system to adapt to a dynamic environment** with varying rewards
- Through a simulation study, we designed **foraging tasks** with the Markov decision process with two different conditions
- From the behavior data analysis, our proposed model **best explained the human behavior data regardless of the environmental conditions**
- From fMRI data analysis, we found evidence that the **prefrontal cortex guides the foraging**

Acknowledgments
This work was supported by (i) Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2023-00233251, System3 reinforcement learning with high-level brain functions), (ii) the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (NRF-2019M3E5D2A01066267) and (iii) the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. 2021M3E5D2A01022493).