A computationally rational model of human reinforcement learning

Zeming Fang, Chris Sims

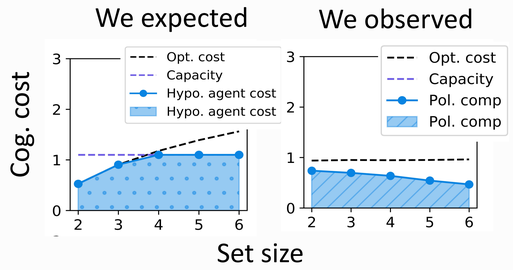
Department of Cognitive Science, Rensselaer Polytechnic Institute, Troy, NY, 12180

Background

Humans may hold a different computationally goal from many artificial reinforcement learning algorithms. The humans' goal is to not only optimize the expected utility but also consider the corresponding cognitive cost. One cognitive cost candidate is policy complexity, defined in terms of information theory as the mutual information between the sensory input and behavioral response.

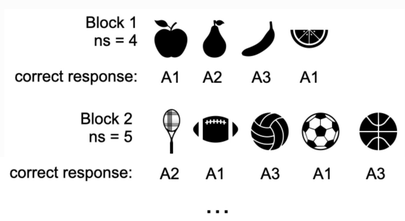
Problem

Humans show the "set set" effect in reinforcement learning tasks: humans' learning efficiency decreases when the number of the presented stimuli increases. However, using a published data set (Collins & Frank, 2012), we founded the policy complexity cannot explain the set size effect. We were prepared to interpret the suboptimality as the human brain's balancing task performance and the rising cognitive cost, but the optimal policy complexity does not necessarily increase with the set size



Experiment paradigm

In each trial of Collins & Frank experiment (2012), subjects are presented one stimuli (sampled from a stimuli pool $\mathcal{S}$) instructed to choose the correct response from **3**. A deterministic feedback is offered. The size of the stimuli varies from 2 to 6, called set size ns.



The information notions

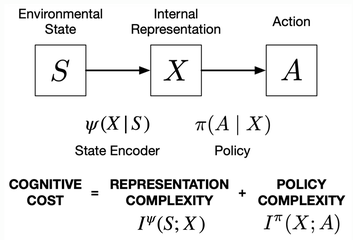
The mutual information:

$$\mathrm{I}^{\pi}(S;A) = \sum_s p(s) \sum_a \pi(a|s) \log \frac{\pi(a|s)}{p(a)}$$

The policy complexity:

$$\mathrm{I}^{\pi}(S;A) = \mathrm{H}(A) - \mathrm{H}(A|S) \leq \mathrm{H}(A) \leq \log \frac{1}{3}$$

The brain as a cascade information channel (Zenon et al., 2019):



The cognitive cost: $\mathrm{I}^{\Psi}(S;X) + \mathrm{I}^{\pi}(X;A)$

Modeling

$\mathrm{M}^{\Psi+\pi}$: An actor-critic model based on Tishby and Polani, 2011; Genewein et al., 2015.

Computational goal:

$$\max_{\pi} \mathrm{E}[r_t] \quad \text{s.t. } \mathrm{I}^{\Psi}(S;X) + \mathrm{I}^{\pi}(X;A) \leq C$$

The Lagrange multiplier:

$$\max_{\pi} \beta_{ns} \mathrm{E}[r_t] - \mathrm{I}^{\Psi}(S;X) - \mathrm{I}^{\pi}(X;A)$$

Note that β is different for each set size, the subscript denotes the set size it belongs to.

Update the critic:

$$Q^t(s_t,a_t) = Q^{t-1}(s_t,a_t) + \alpha_q[r_t - Q^{t-1}(s_t,a_t)]$$

Estimate the representation-action value function:

$$Q^t_{bel}(x,a) = \sum_x p(s|x) Q^t(s,a)$$

where $\mathrm{p}(s|x) = \psi(x|s)p(s)/p(x)$

Update the actor:

$$\pi^{\mathrm{t}}(a|x) \propto p_a^{t-1}(a) \exp[\beta_{ns} Q^t_{bel}(x,a)]$$

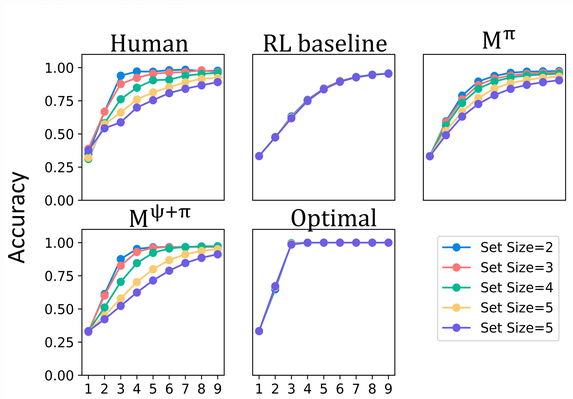
Update the marginal action policy:

$$p_a^t(a) = p_a^{t-1} + \alpha_a[p(a_t|s_t) - p_a^{t-1}(a)]$$

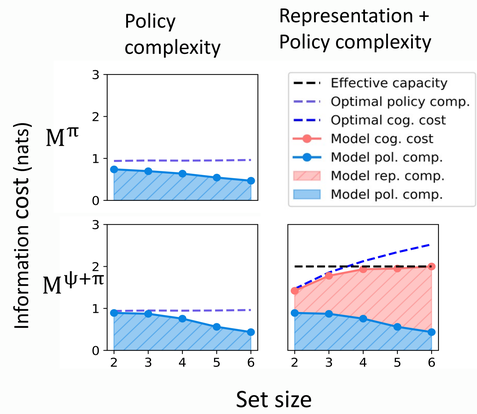
Results

M^{π} is adopted from Lai & Gershman, 2021, but fitting the tradeoff parameter β_{ns} to each set size.
$\mathrm{M}^{\Psi+\pi}$ is the model we proposed.

The set size effect:



Inspect the cognitive cost:



References:

Collins, A. G., & Frank, M. J. (2012). How much of reinforcement learning is working memory, not reinforcement learning? A behavioral, computational, and neurogenetic analysis. *European Journal of Neuroscience*, *35*(7), 1024-1035.

Lai, L., & Gershman, S. J. (2021). Policy compression: An information bottleneck in action selection. *Psychology of Learning and Motivation S*.

Zenon, A., Solopchuk, O., & Pezzulo, G. (2019). An information-theoretic perspective on the costs of cognition. *Neuropsychologia*, *123*, 5-18.