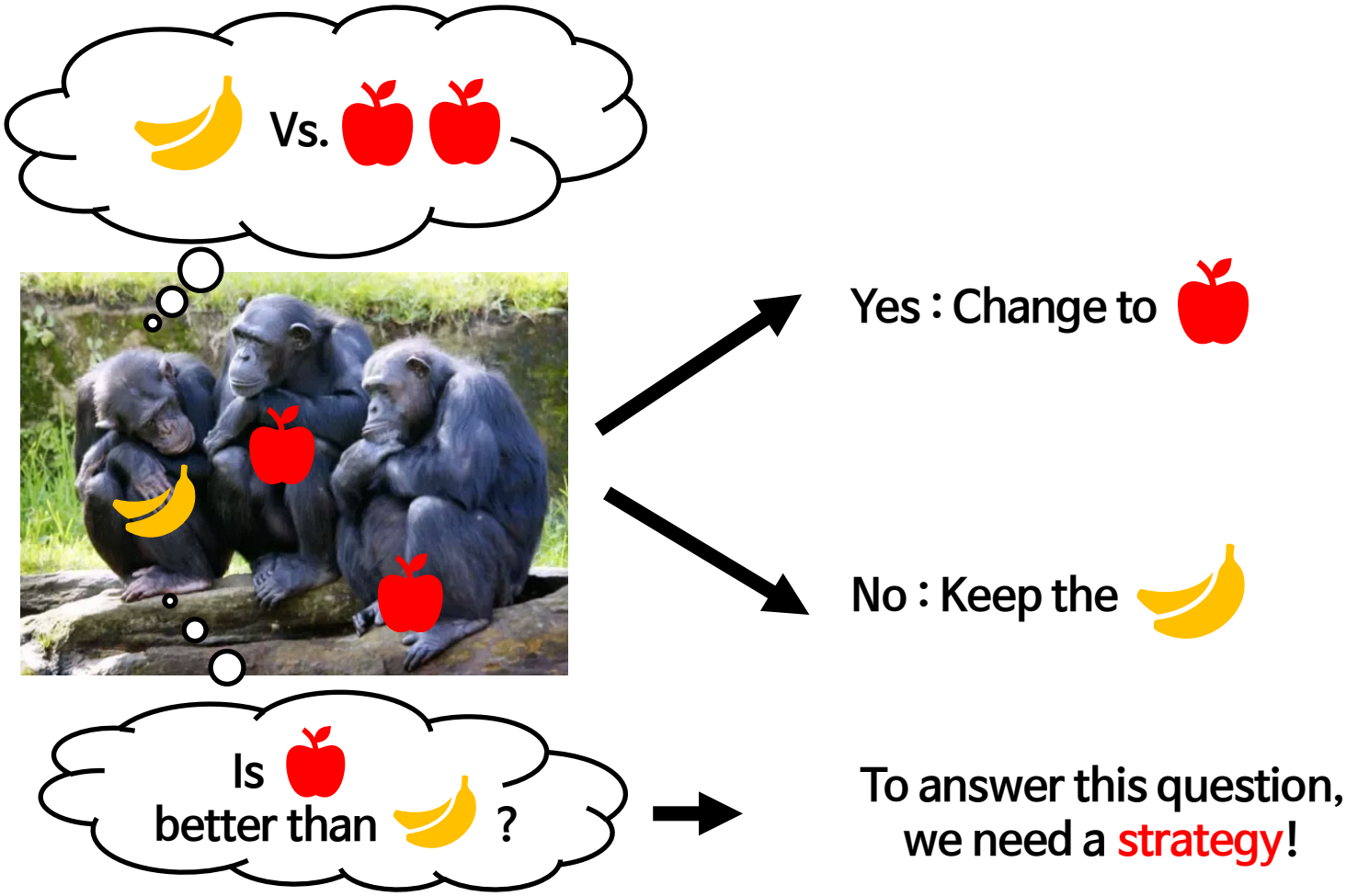
2023-07-26
ICML 2023

Social learning spontaneously emerges by searching optimal heuristics with deep reinforcement learning

Seungwoong Ha

KAIST SANTA FE INSTITUTE

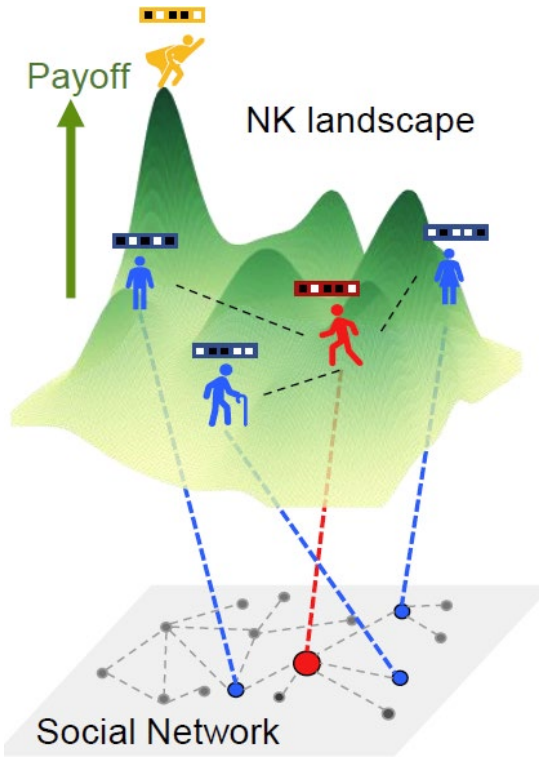
Social Learning Strategies (SLSs)



Strategy for social learning is critical for high utility.

Model for social environment

NK landscape (Kauffman, 1987)



N=5, K=2

Self, Neighbors

Lookup table

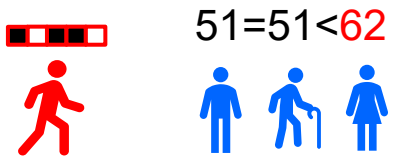
Comp.	Fitness
000	0.25
001	0.13
010	0.01
011	0.98
100	0.54
101	0.33
110	0.71
111	0.40

$$S = \sum_i s_i$$

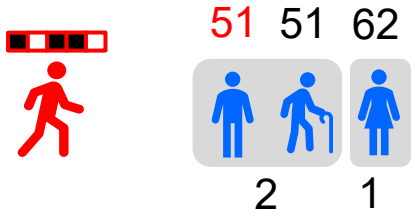
$$S_{\text{norm}} = \left(\frac{S}{S_{\max}} \right)^8$$

Conventional Strategies

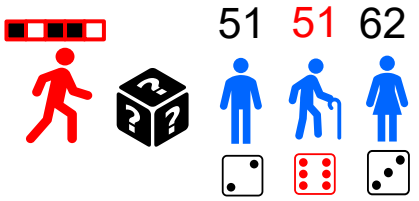
- **Best Imitator (BI)**
  - Follows the best.



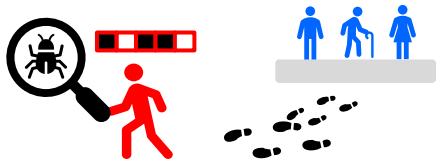
- **Conformist (CF)**
  - Follows the majority.



- **Random Imitator (RI)**
  - Follows randomly chosen one.

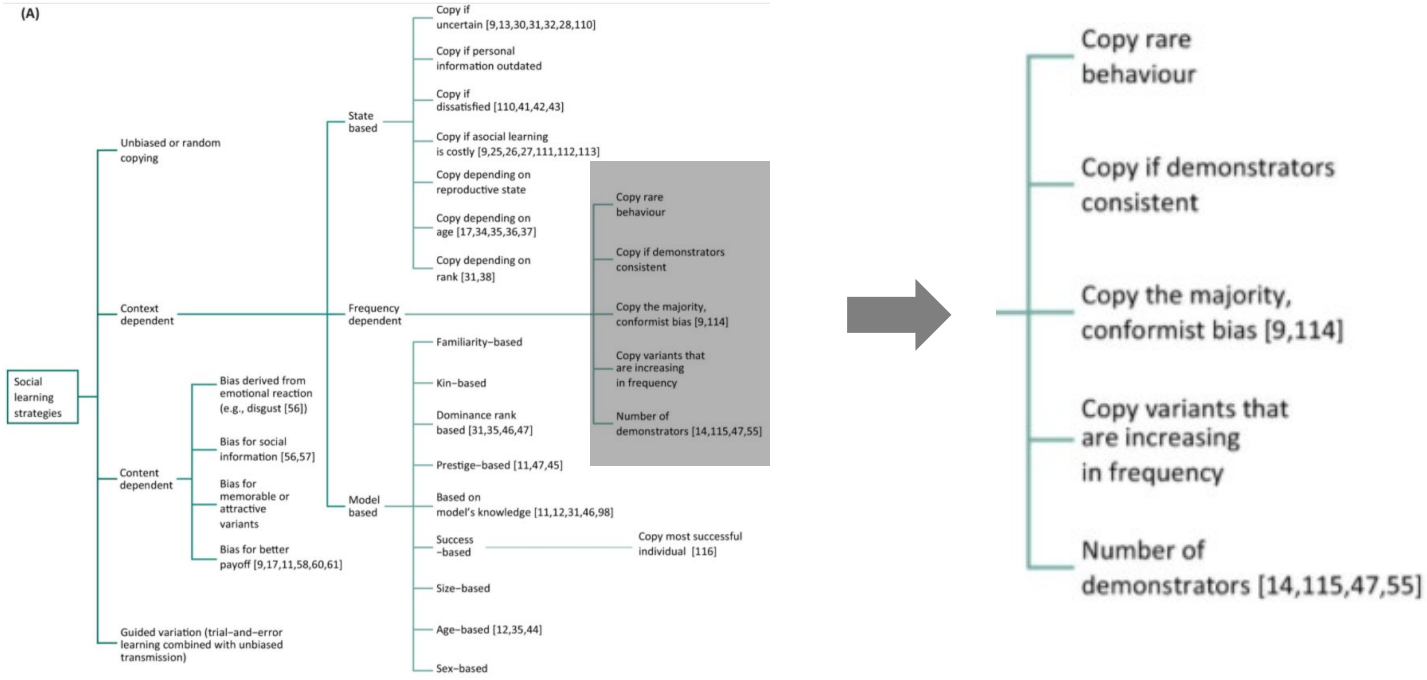


- **Pure-Individualist (PI)**
  - Do not follow, explore on its own.



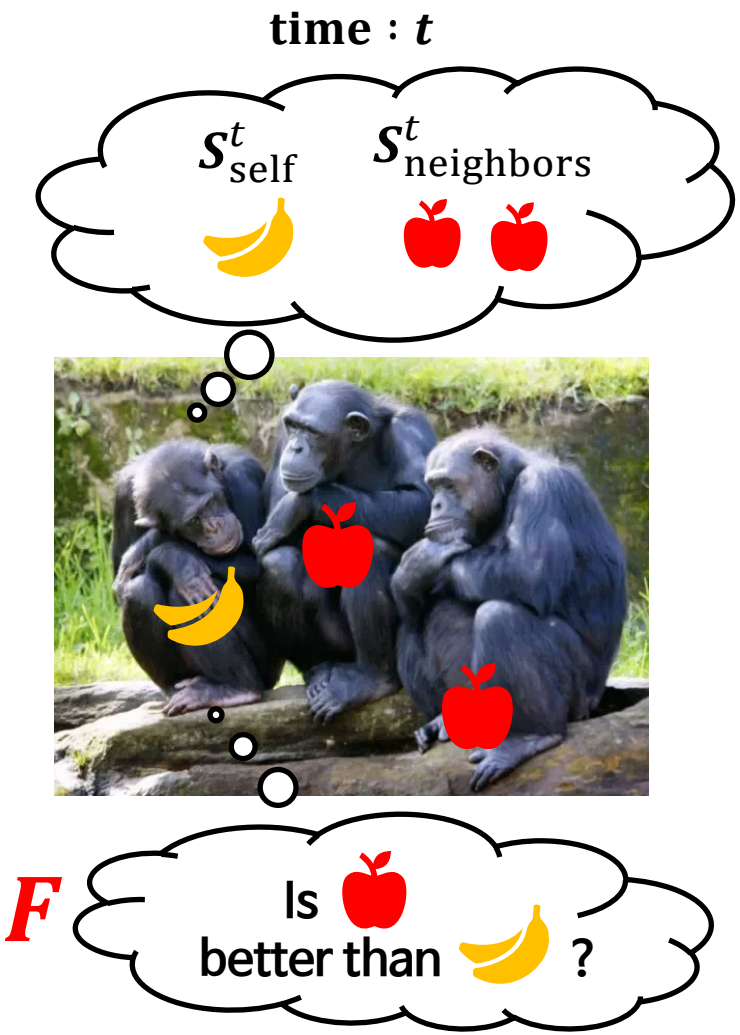
But what is the most suitable SLS?

- **Finding optimal SLS for given settings is not an easy task.**
  - There are so many options for ways of Individual/Social learning.
  - This is a problem of finding an '**optimal optimization method**' for given settings, which is a meta-problem itself → General meta-heuristics!

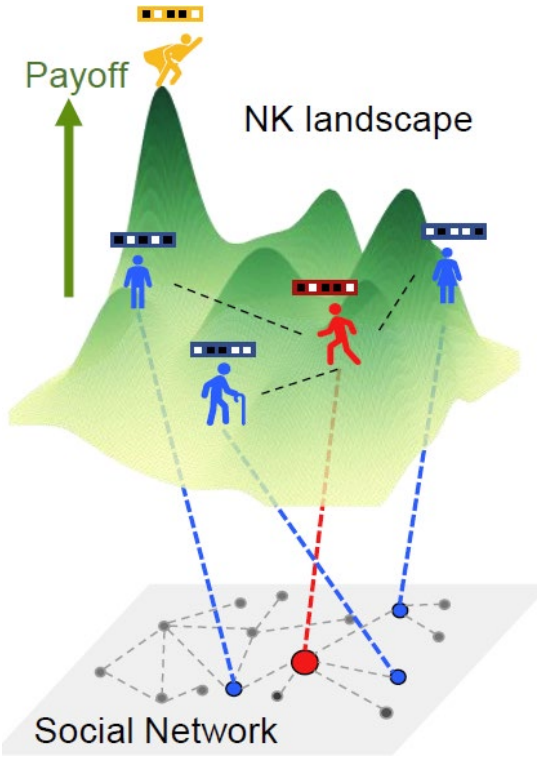


Kendal, R. L., Boogert, N. J., Rendell, L., Laland, K. N., Webster, M., & Jones, P. L. (2018). Social learning strategies: Bridge-building between fields. *Trends in cognitive sciences*, *22*(7), 651-665.

Learning heuristics of social learning



Social animals perform 'Social learning'

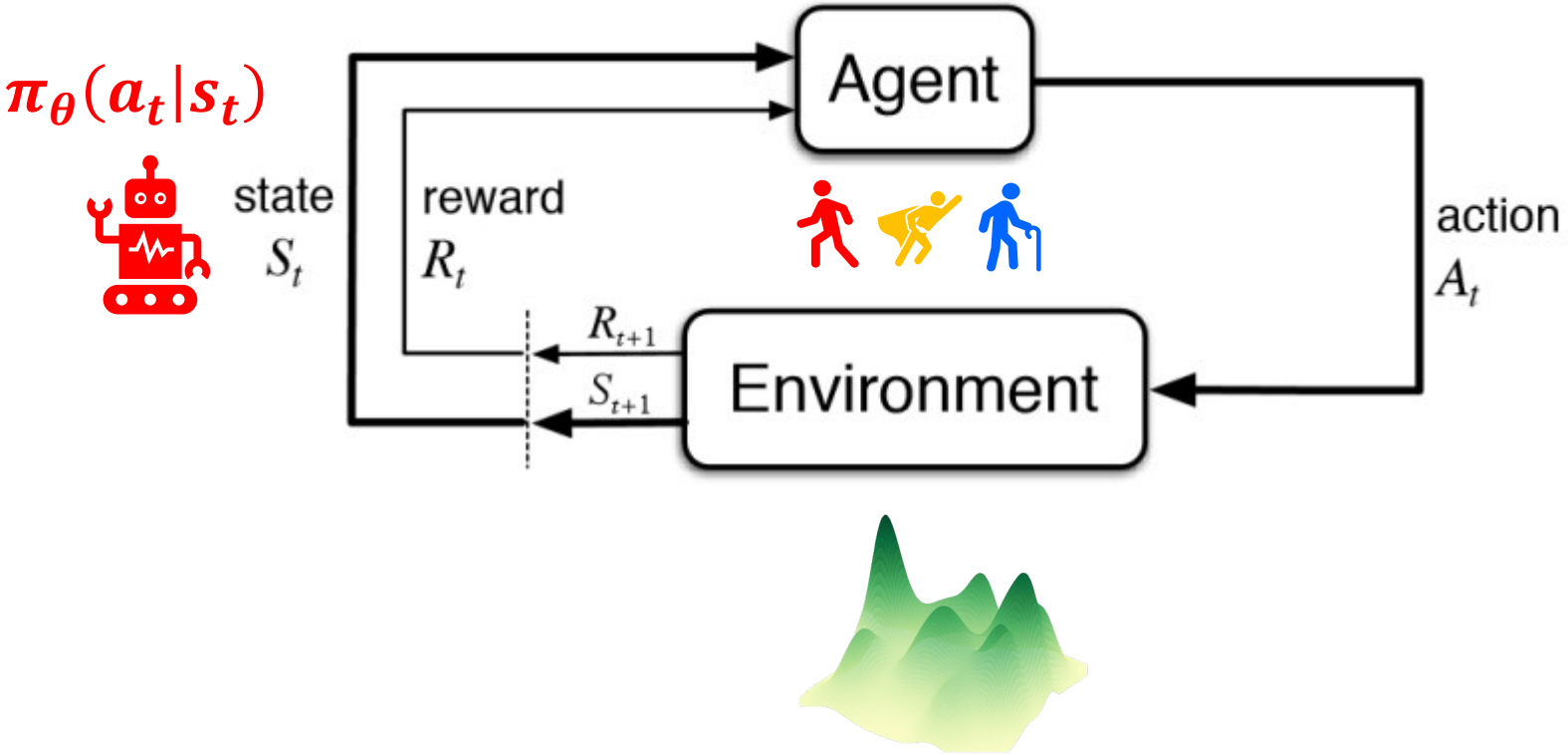


$$F\left(s^t_{\text{self}}, s^t_{\text{neighbors}}\right) = s_i^{t+1}$$

What would be the optimal strategy?
(after t = 200?)

Finding an optimal SLS via deep learning

- SLS can be formulated as a policy of reinforcement learning (RL).



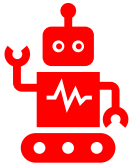
Proximal policy optimization (PPO)

- **To obtain the heuristics, we directly optimize the policy.**
  - PPO = Actor-Critic model + Importance sampling + Clipping + (entropy loss)
  - Simple description for Actor-Critic model

$$A_t = A_t\big(V(s_t)\big)$$

$$L_{\text{actor}} = \mathbb{E}_t[\pi_\theta(a_t|s_t)A_t]$$

Generalized Advantage Estimator (GAE)

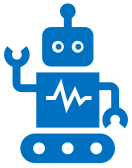


Actor

Goal : Maximize the expected reward

$$L_{\text{critic}} = \big(R_t - V(s_t)\big)^2$$

Value function

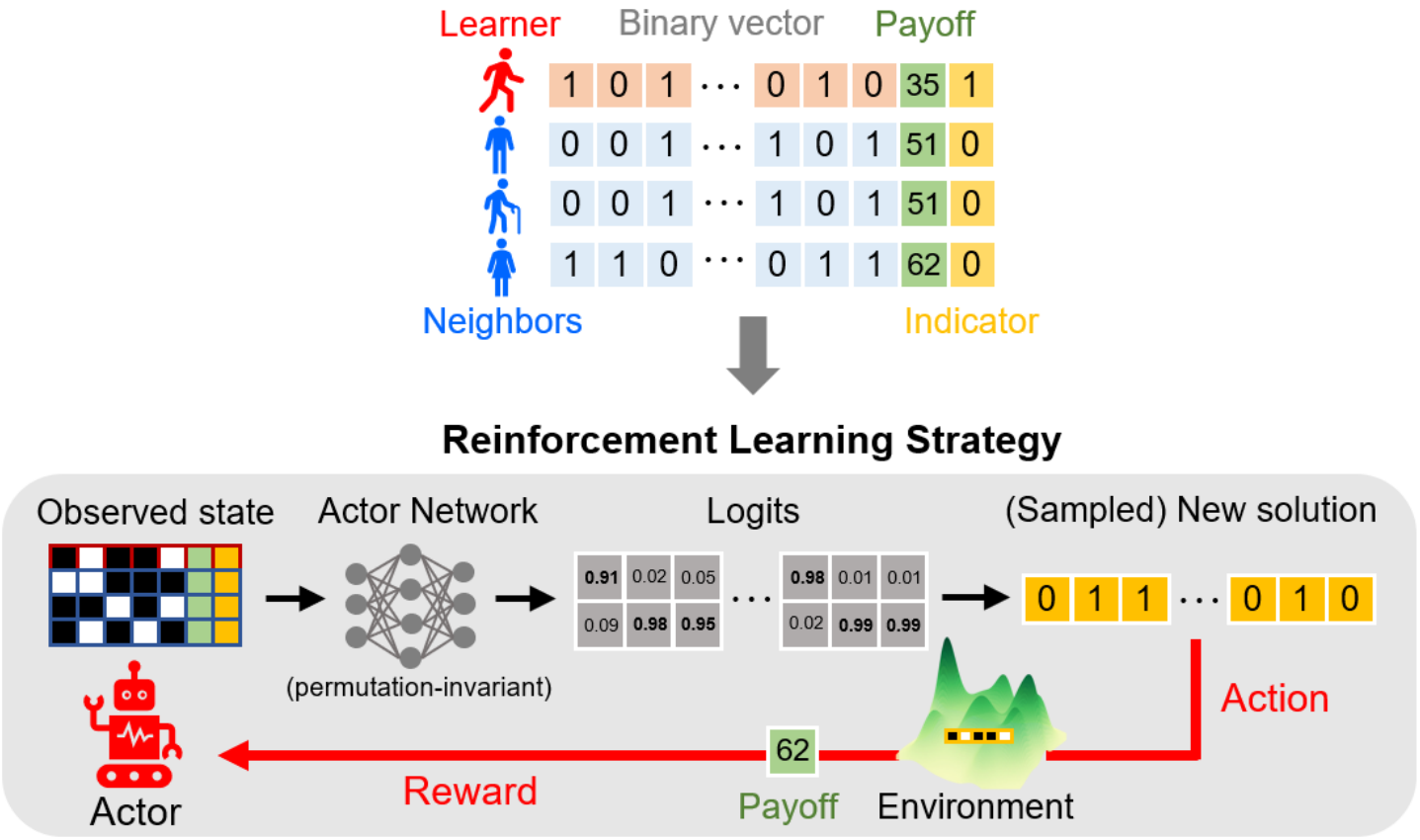


Critic

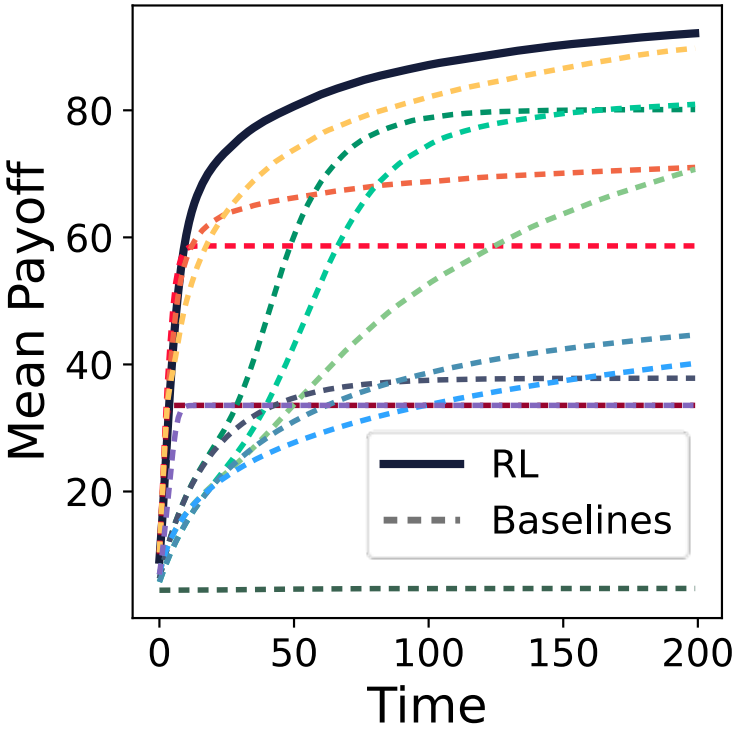
Goal : Predict the expected reward

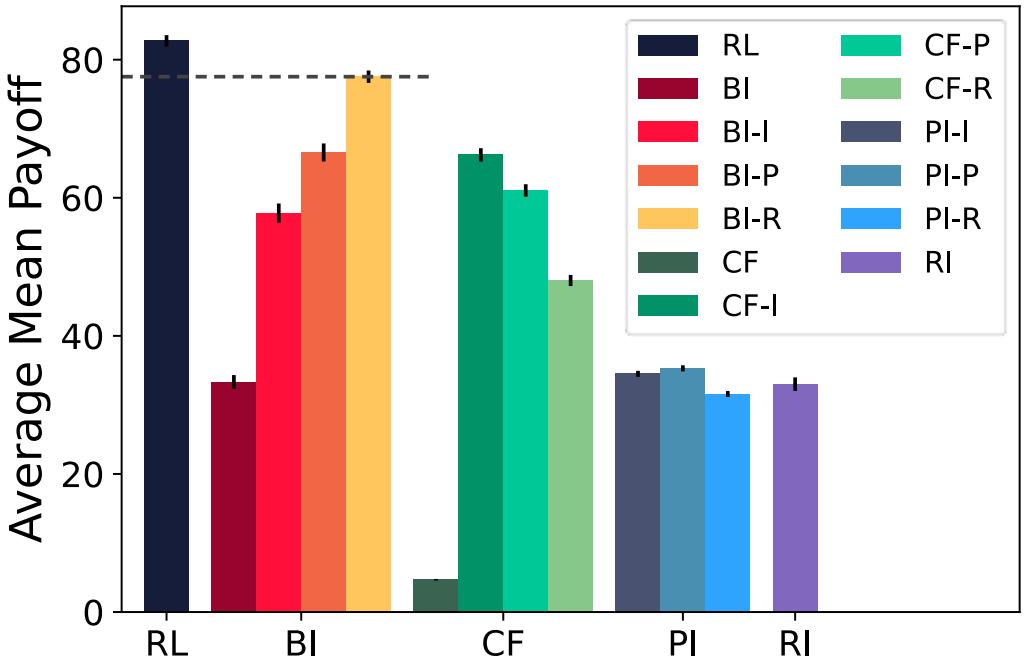
Reinforcement learning

- **SLS can be formulated as a policy of reinforcement learning (RL).**
  - Actor network is consisted of Set Transformer architecture.
  - In practice, we use 'frequency' and 'rank' as indicators.



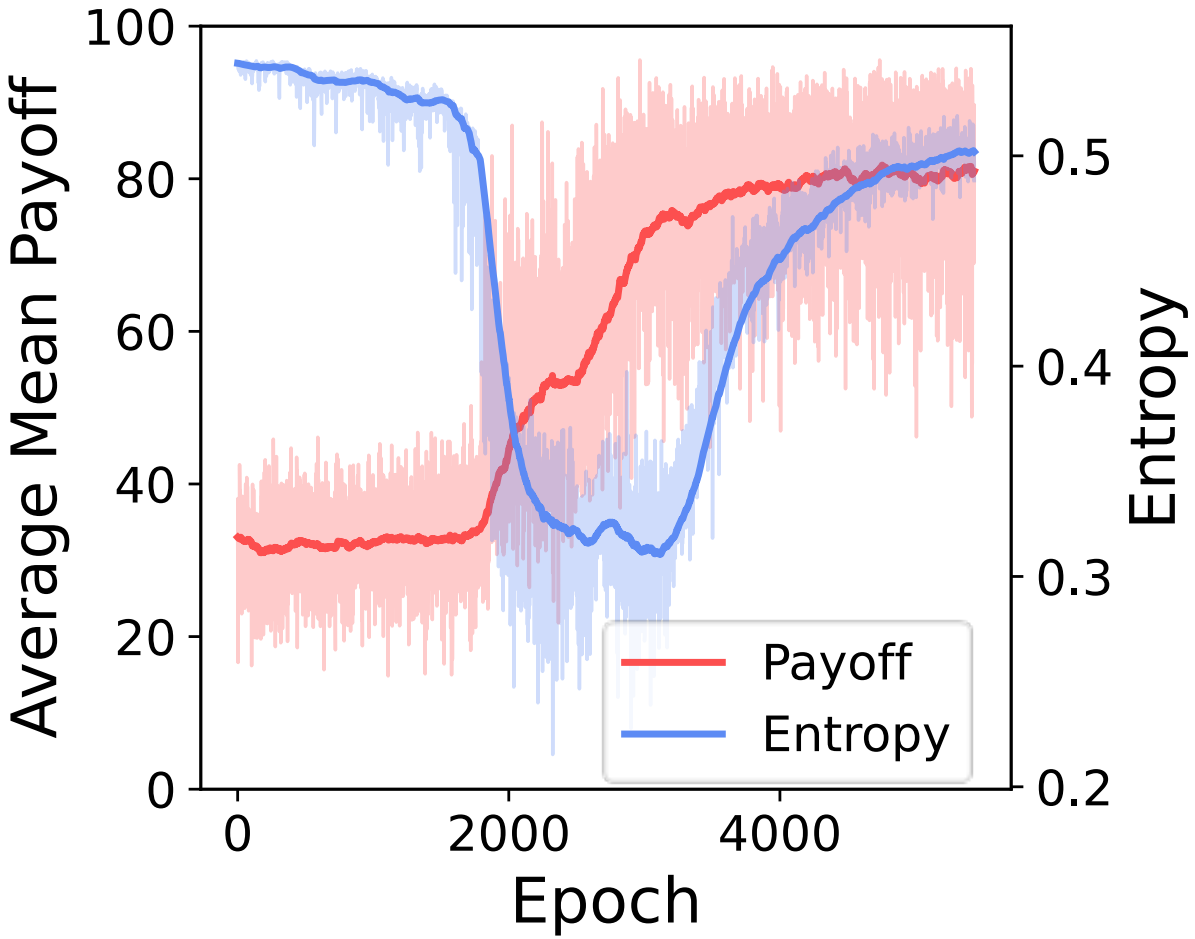
Default setting (N15K7, L200)





RL agent outperforms all baseline SLSs.

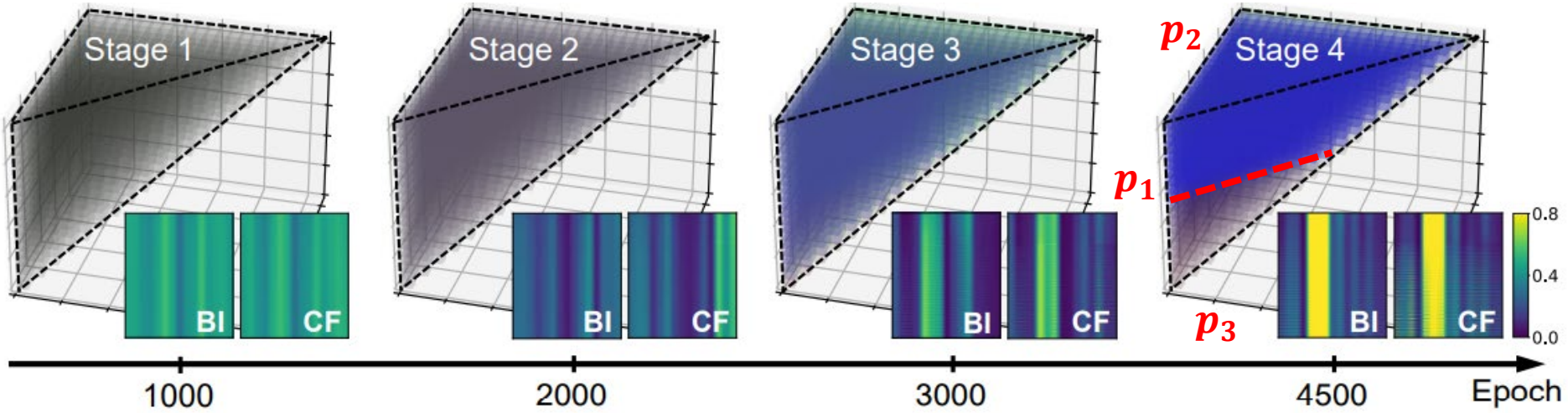
Default setting (N15K7, L200)

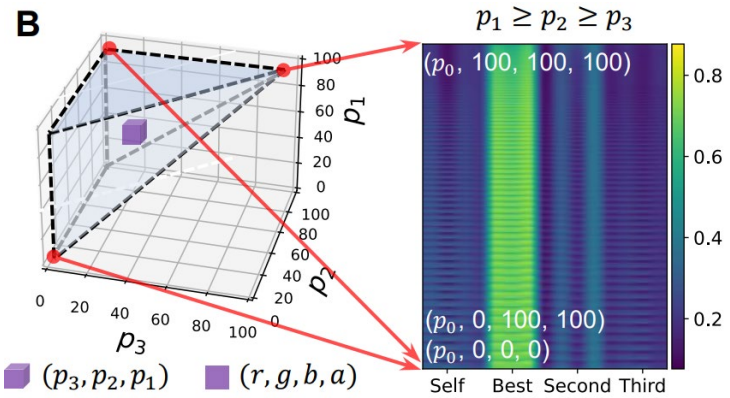


Interestingly, entropy shows non-monotonic behavior.

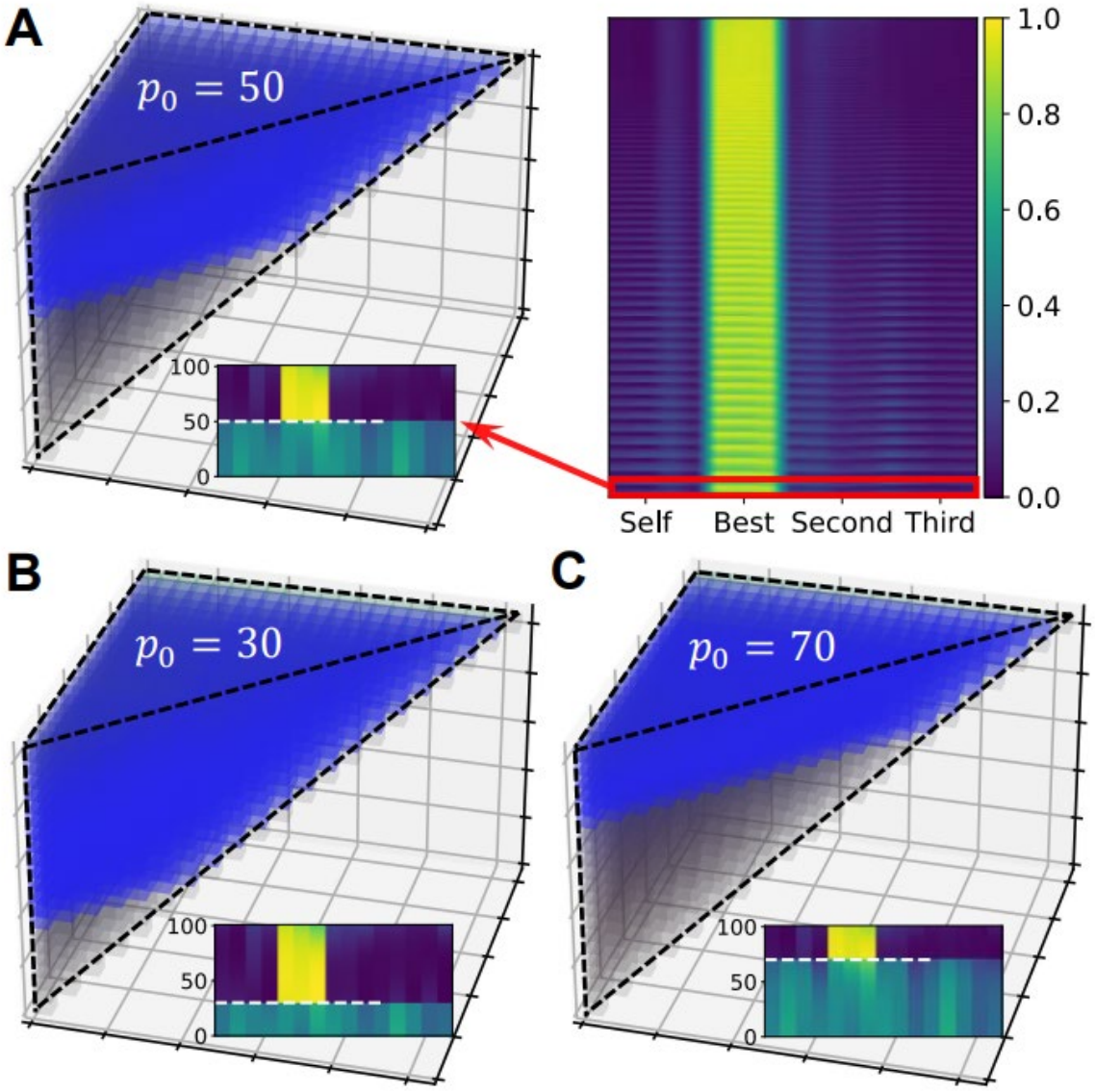
Default setting (N15K7, L200)

Self score $p_0 = 50$

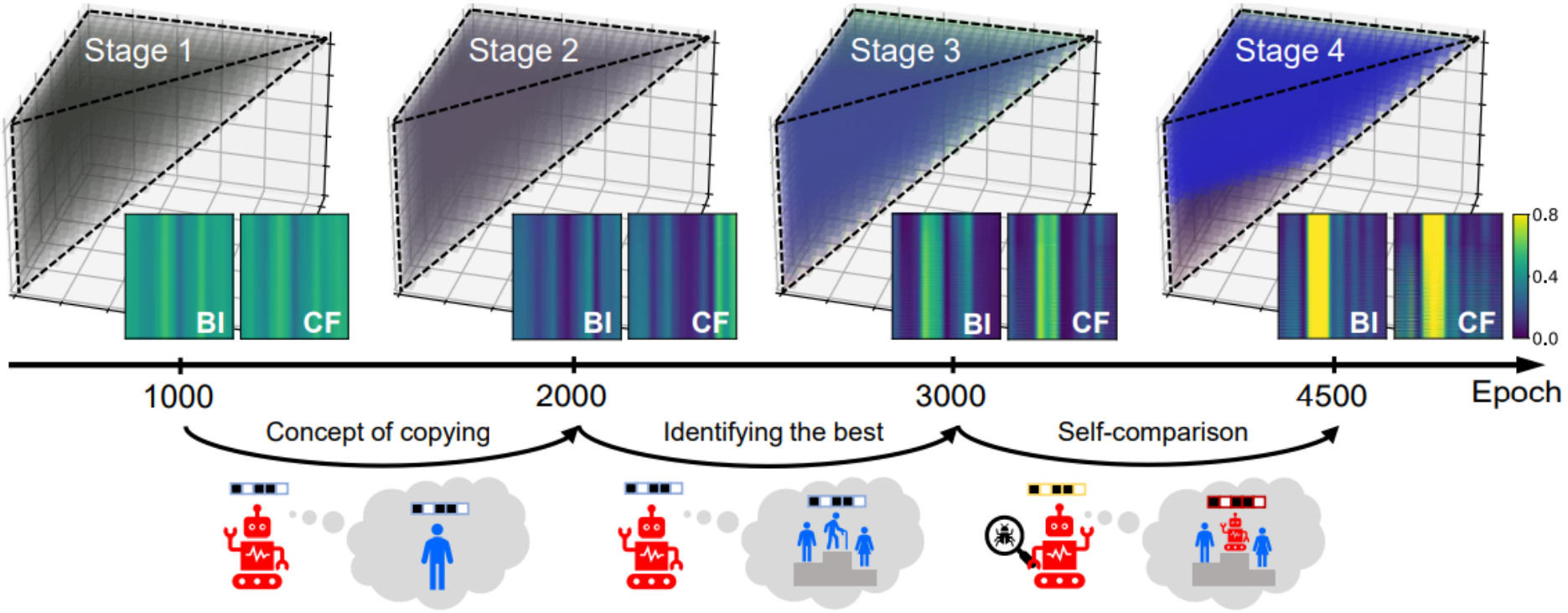




Default setting (N15K7, L200)

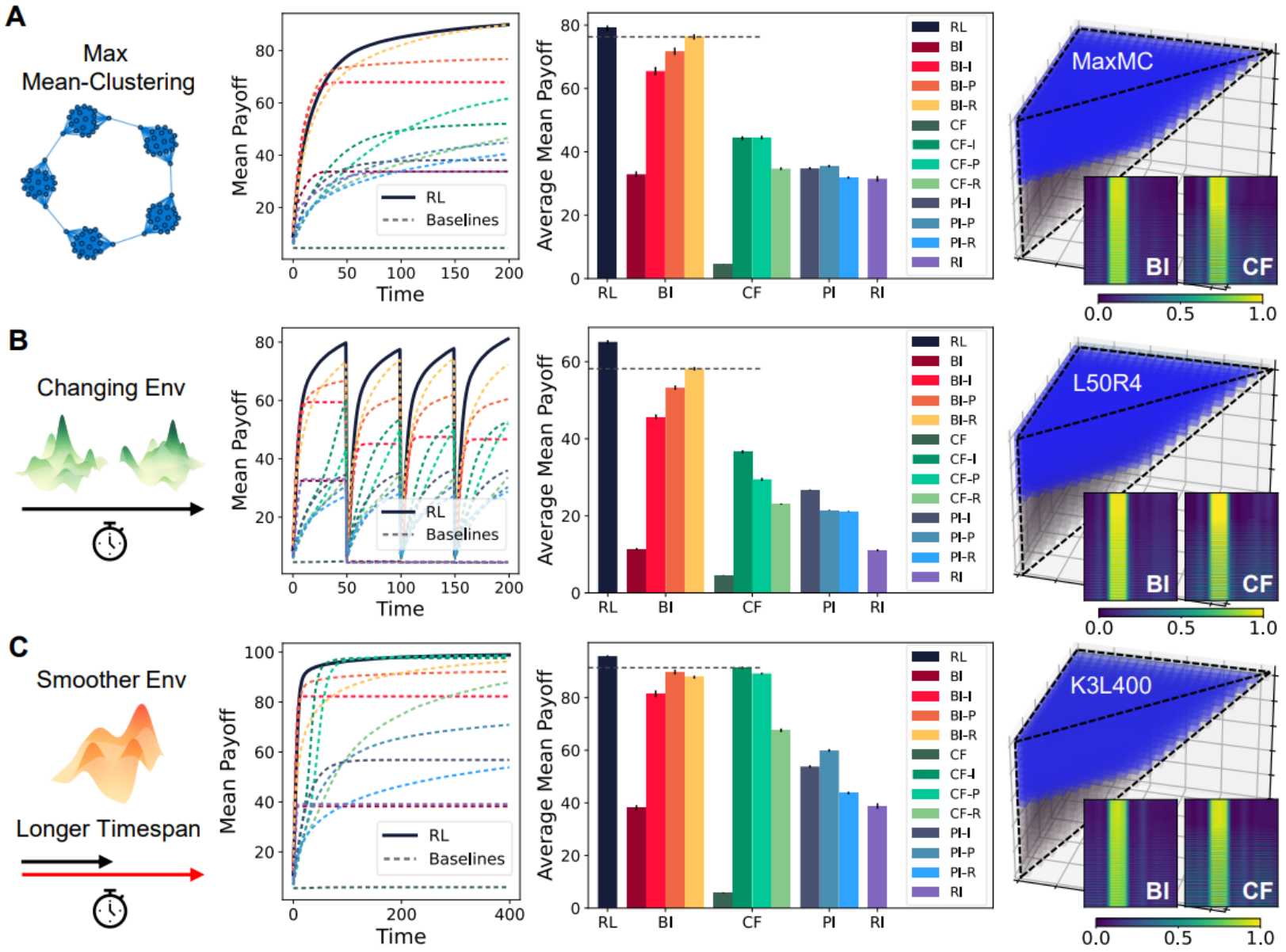


Default setting (N15K7, L200)

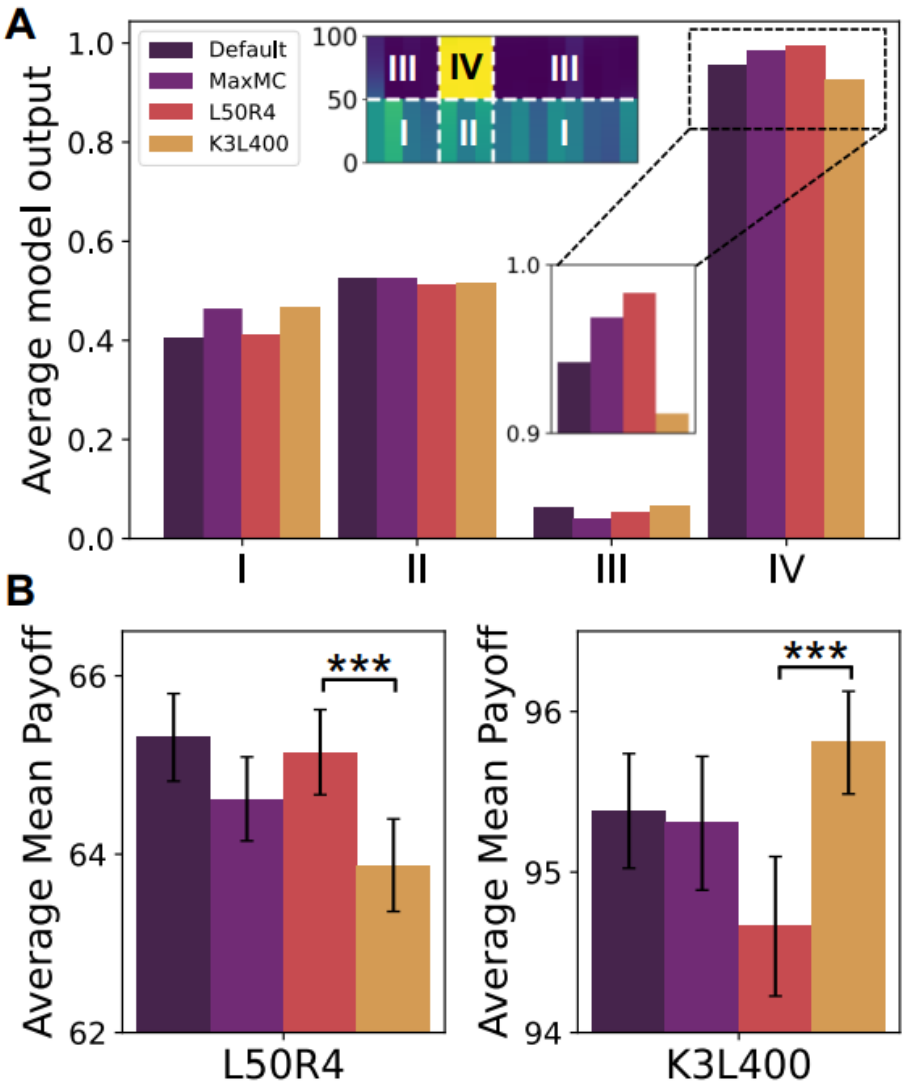


RL agent learns key concepts in SL by stages.

Various Environmental settings



Various Environmental settings



Thank you for listening!

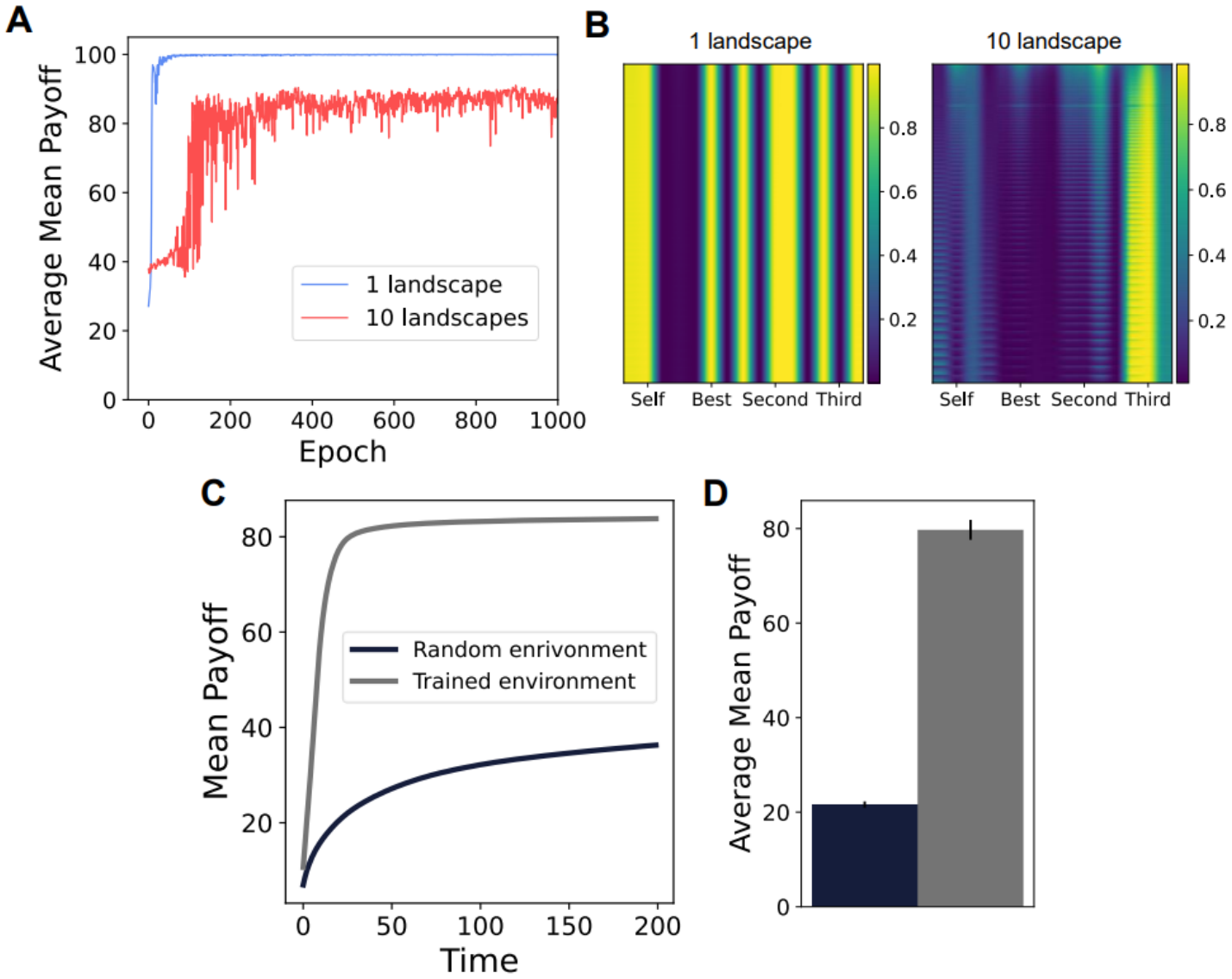
2023-07-26
ICML 2023

Social learning spontaneously emerges by searching optimal heuristics with deep reinforcement learning

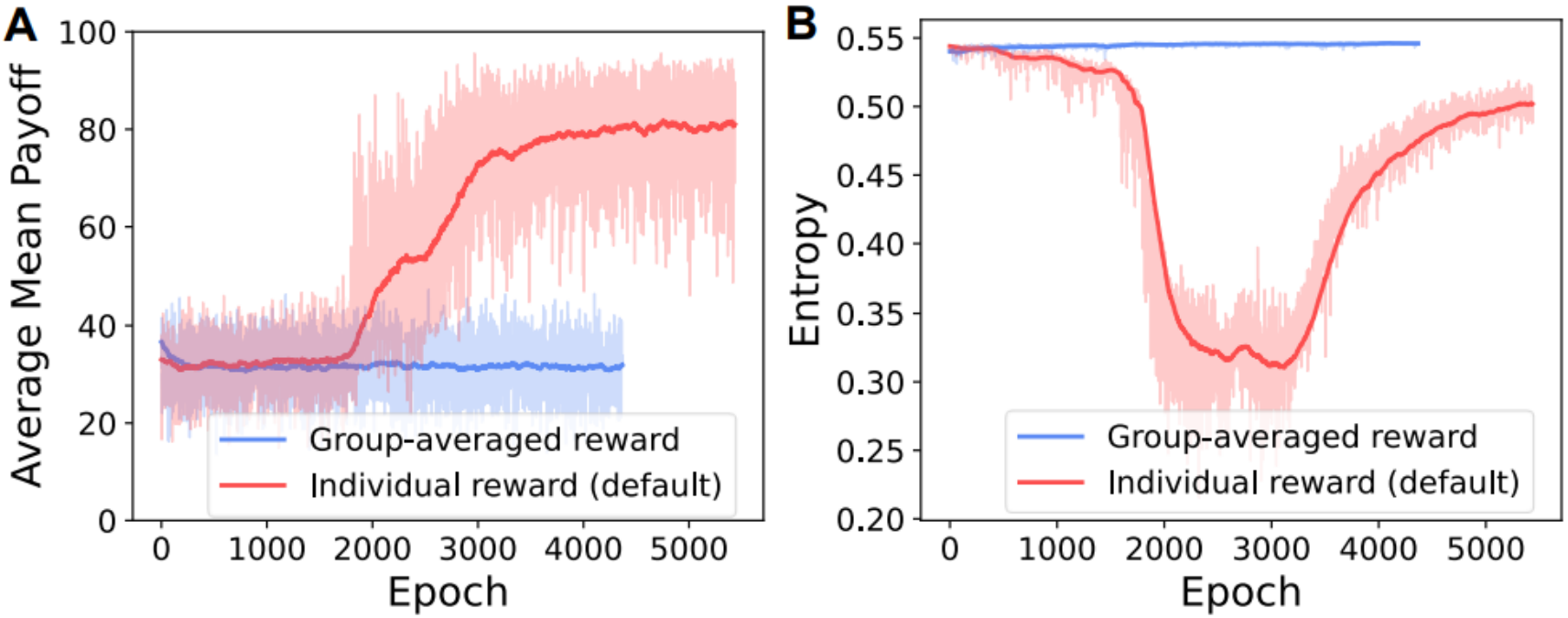
Seungwoong Ha

KAIST SANTA FE INSTITUTE

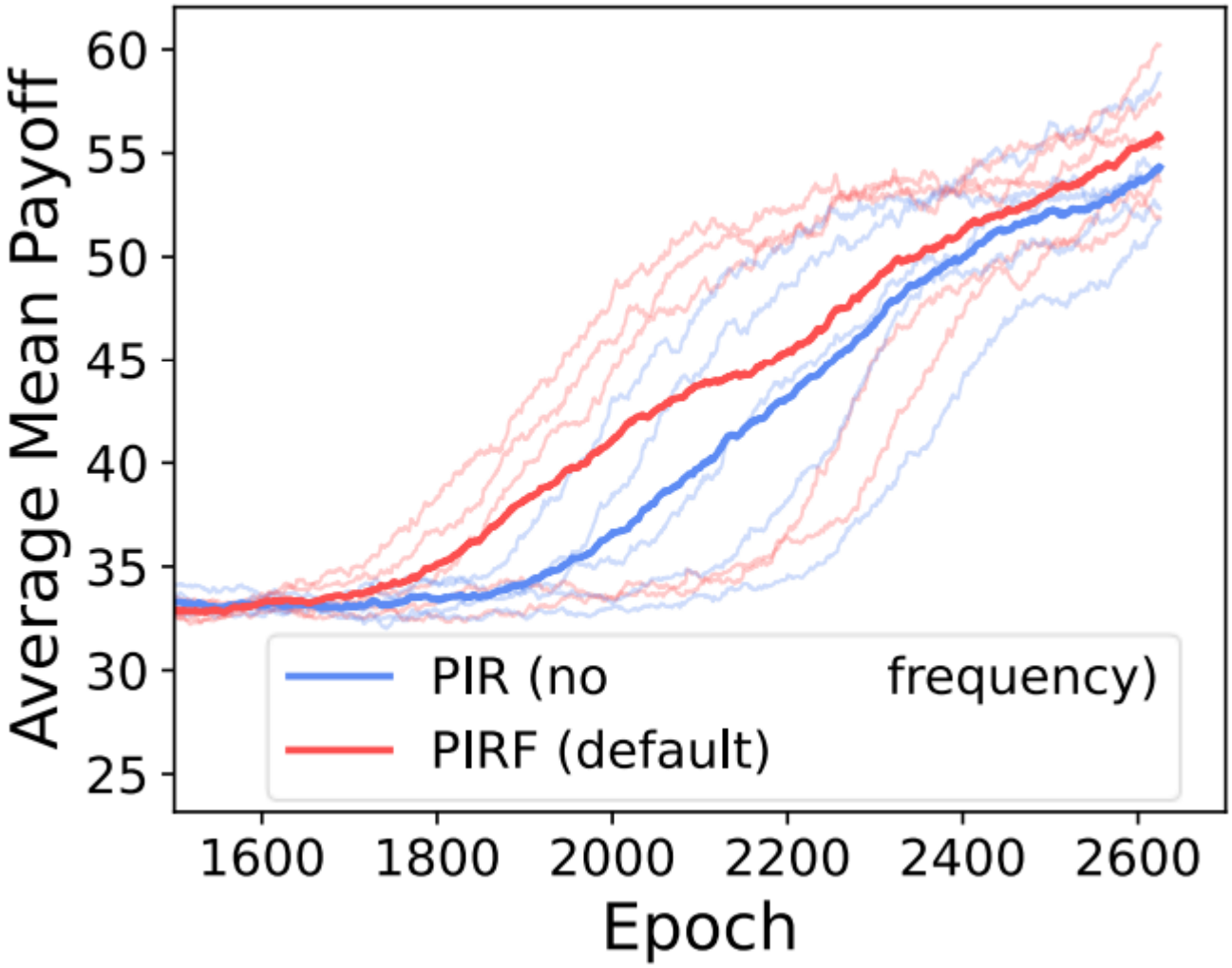
Training with fixed environment



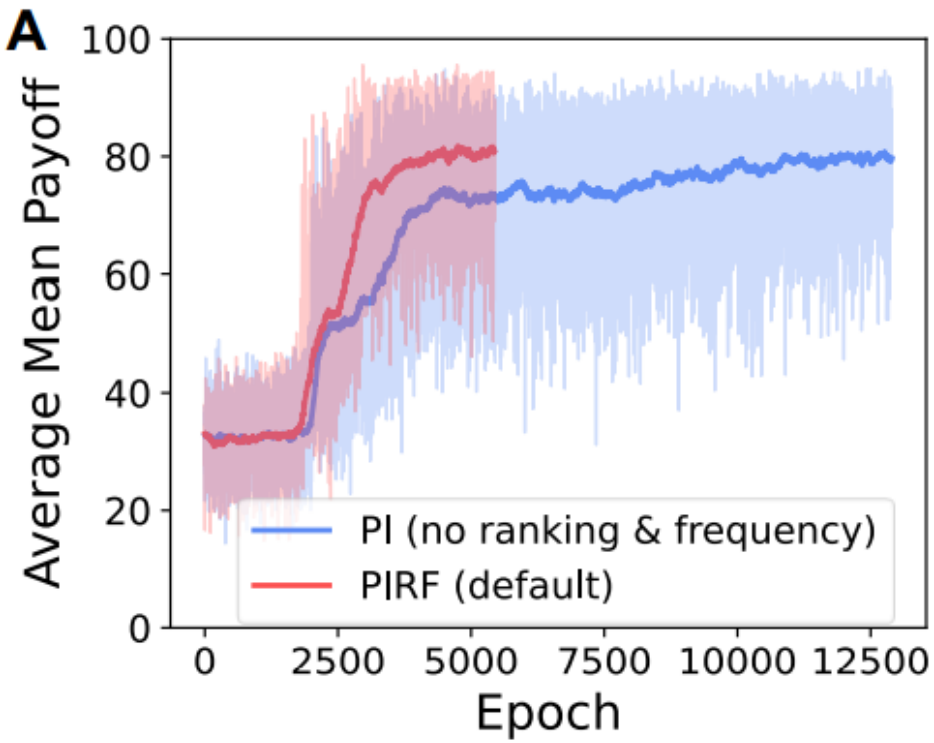
Group-averaged reward

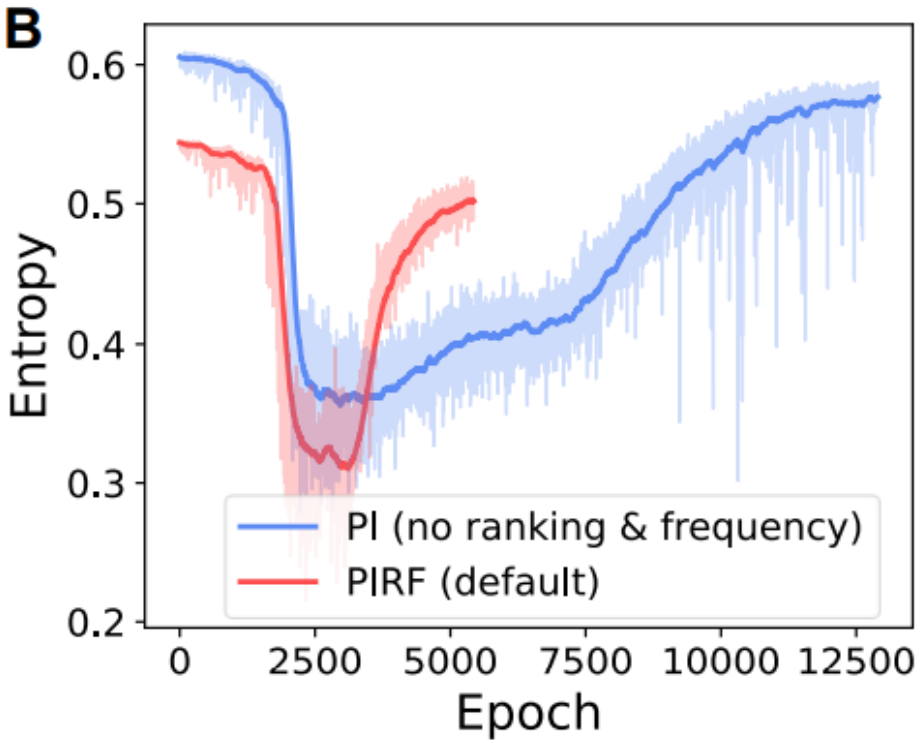


Different input features

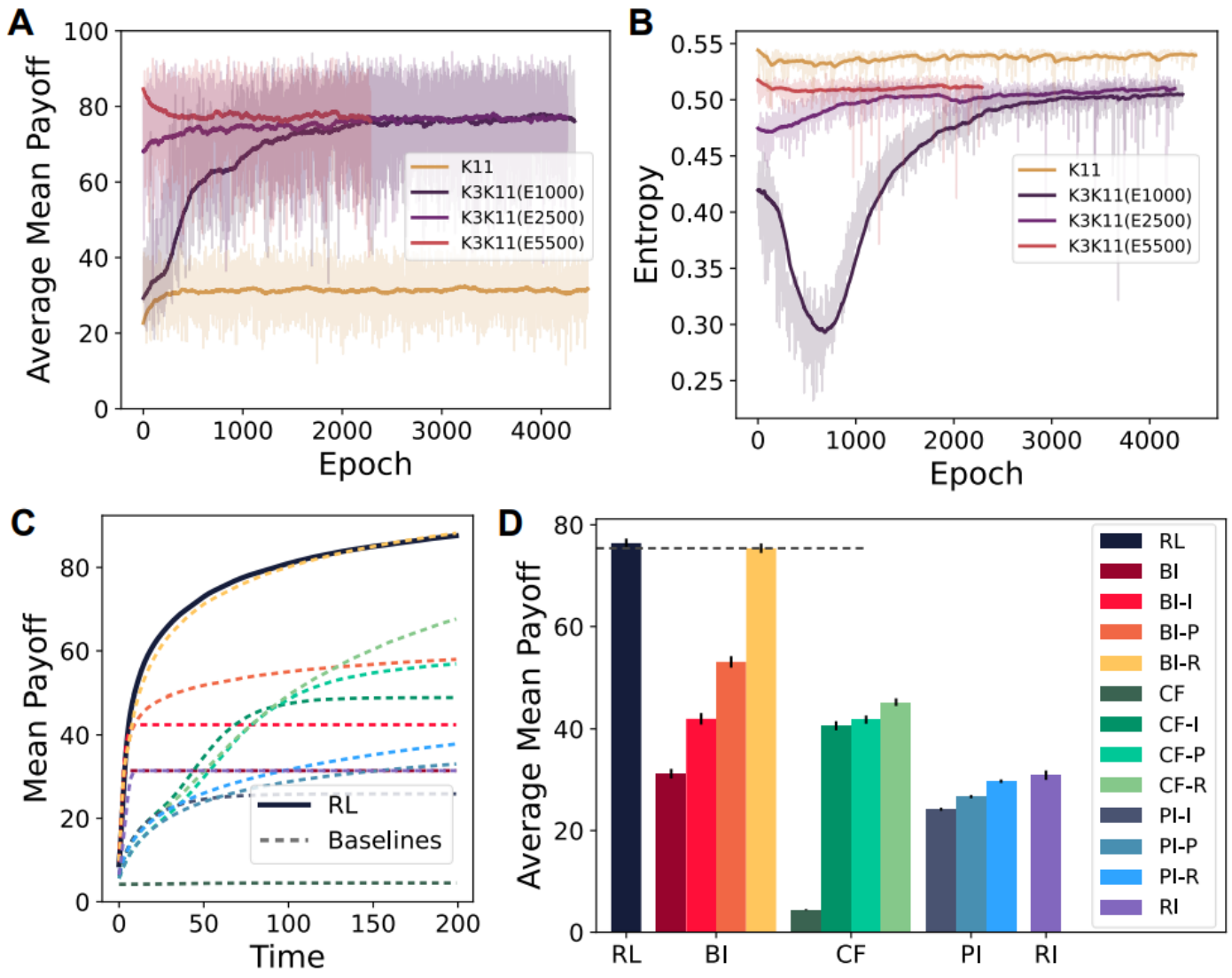


Different input features

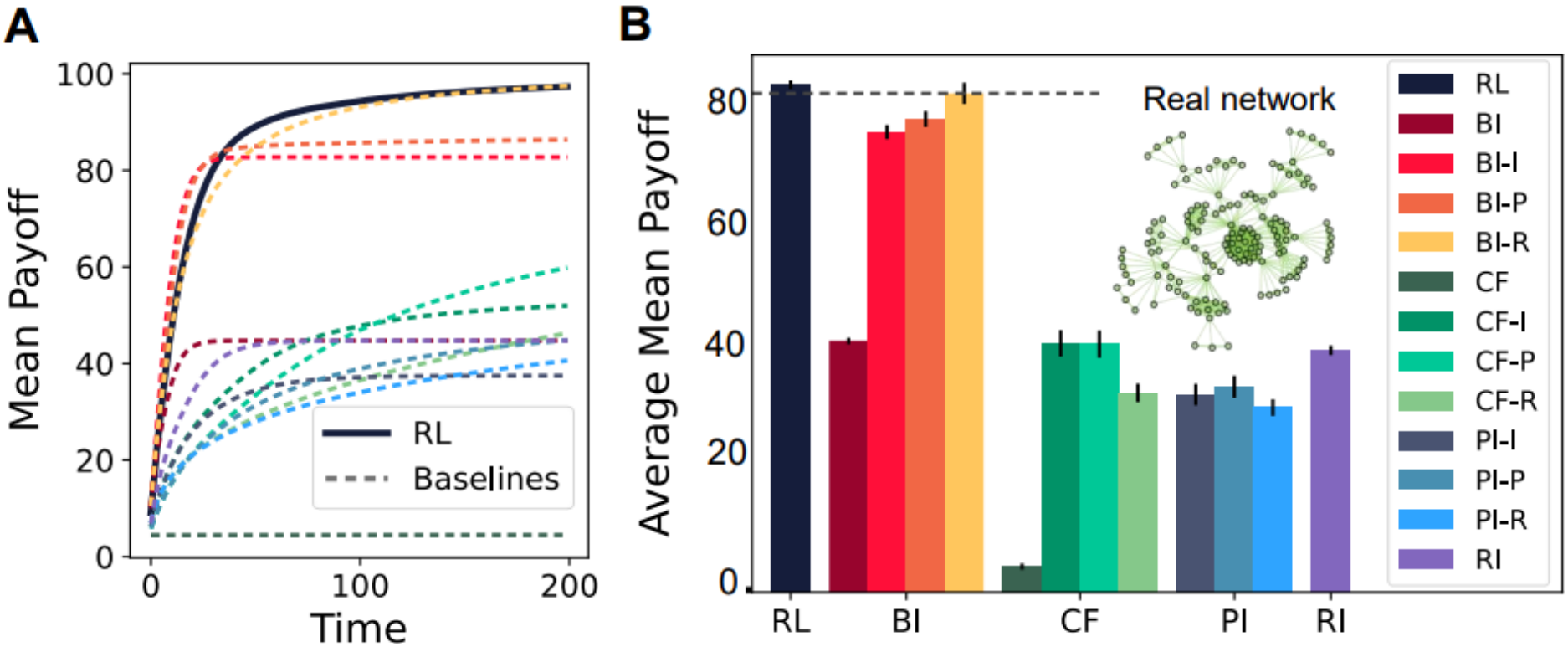




Highly rugged landscape with learning schedule



Real world network (Averaged over 88)



Thank you for listening!

2022-10-20
CCS 2022

Social learning spontaneously emerges by searching optimal heuristics with deep reinforcement learning

Seungwoong Ha
Advisor : Hawoong Jeong

KAIST