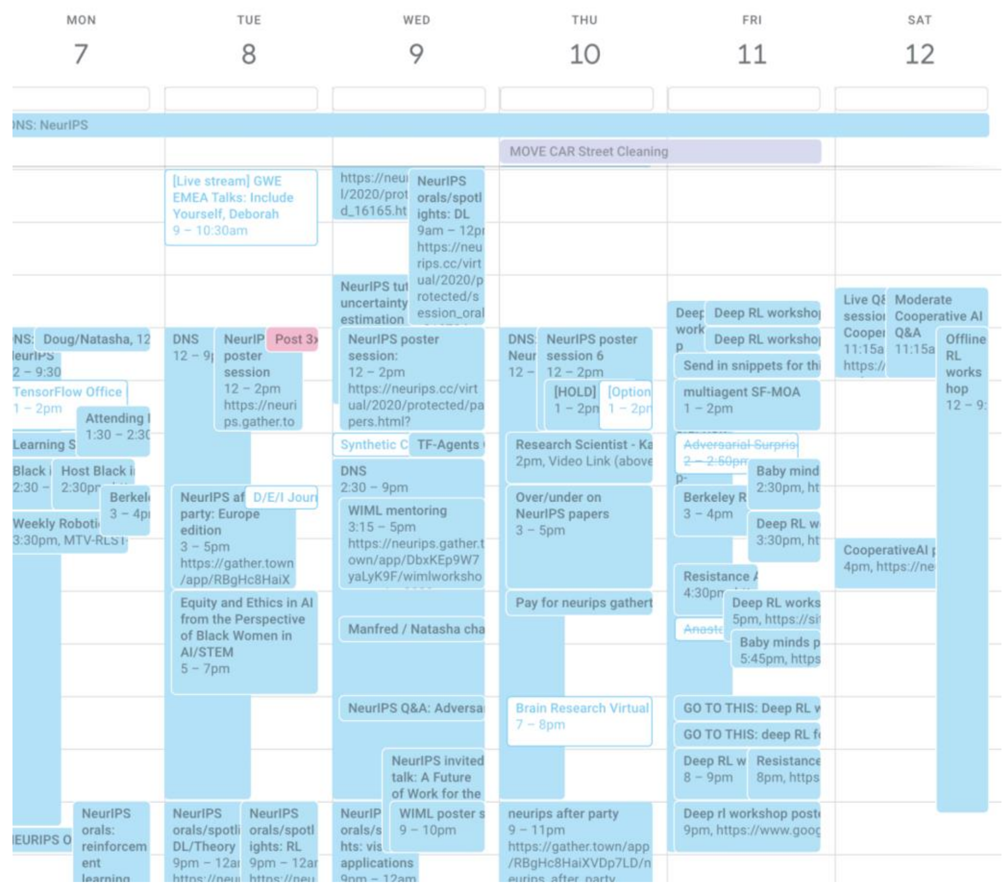
Social Reinforcement Learning

Natasha Jaques

BAIR
BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH
Google
mit media lab
affective computing

What abilities does an AI assistant need?

https://clipart.me/free-vector/kitchen-robot



Coordinate in shared spaces

Learn from **human interaction**

Learn **complex** tasks

Generalize to new environments

Hypothesis: social learning can help address all of these desiderata

Coordinate in shared spaces

Learn from **human interaction**

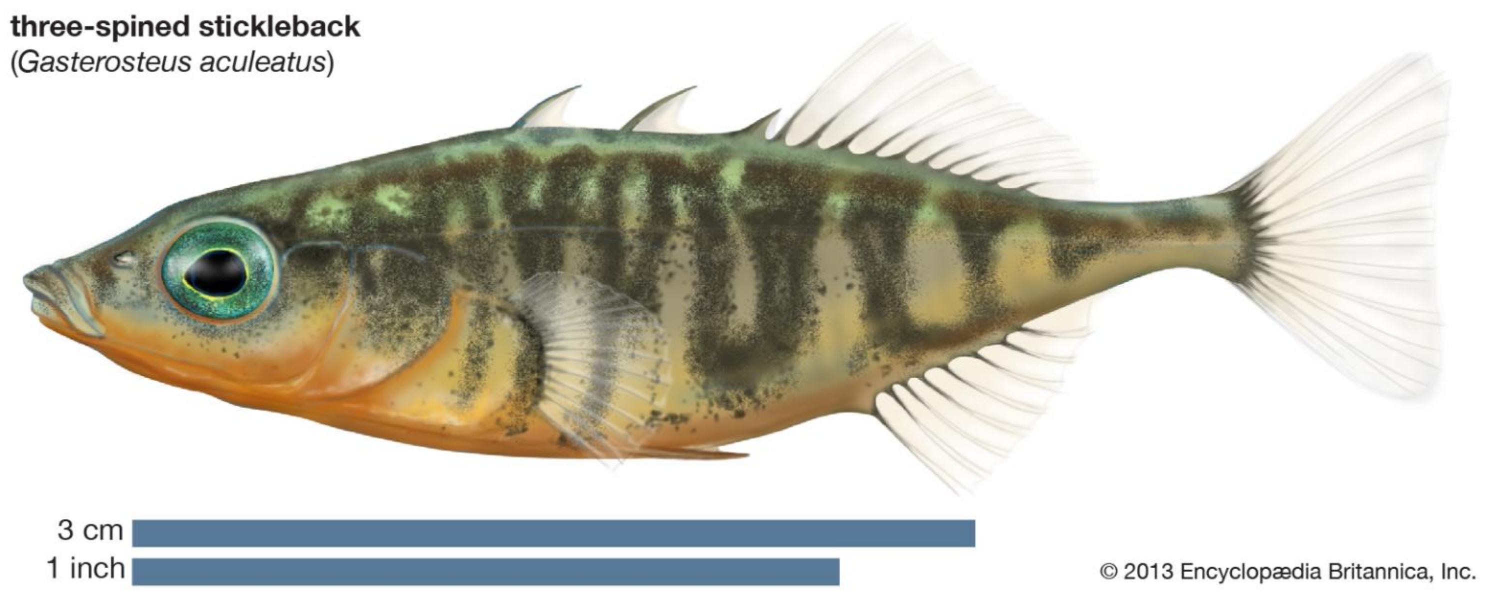
Learn **complex** tasks

Generalize to new environments

Social learning

= Learning from other intelligent agents in your environment

- Helps humans and animals...
  (Laland, 2017; Henrich, 2015)

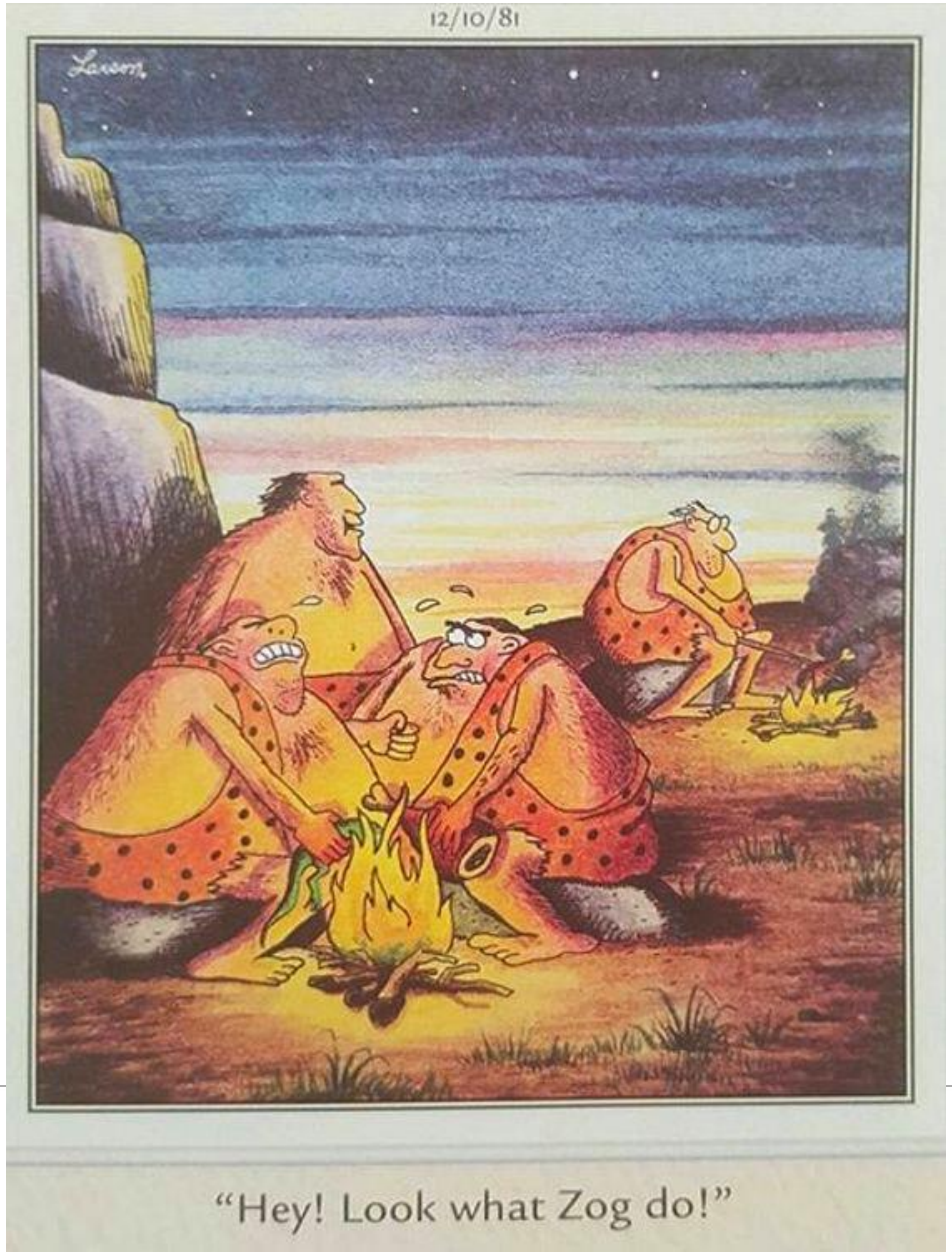


Generalize to new environments

Learn complex behavior

- Is key to **human** cognitive development, cultural and technological evolution
  (Henrich, 2015; Humphrey, 1976, Tomasello, 2009)

Social learning can accelerate learning

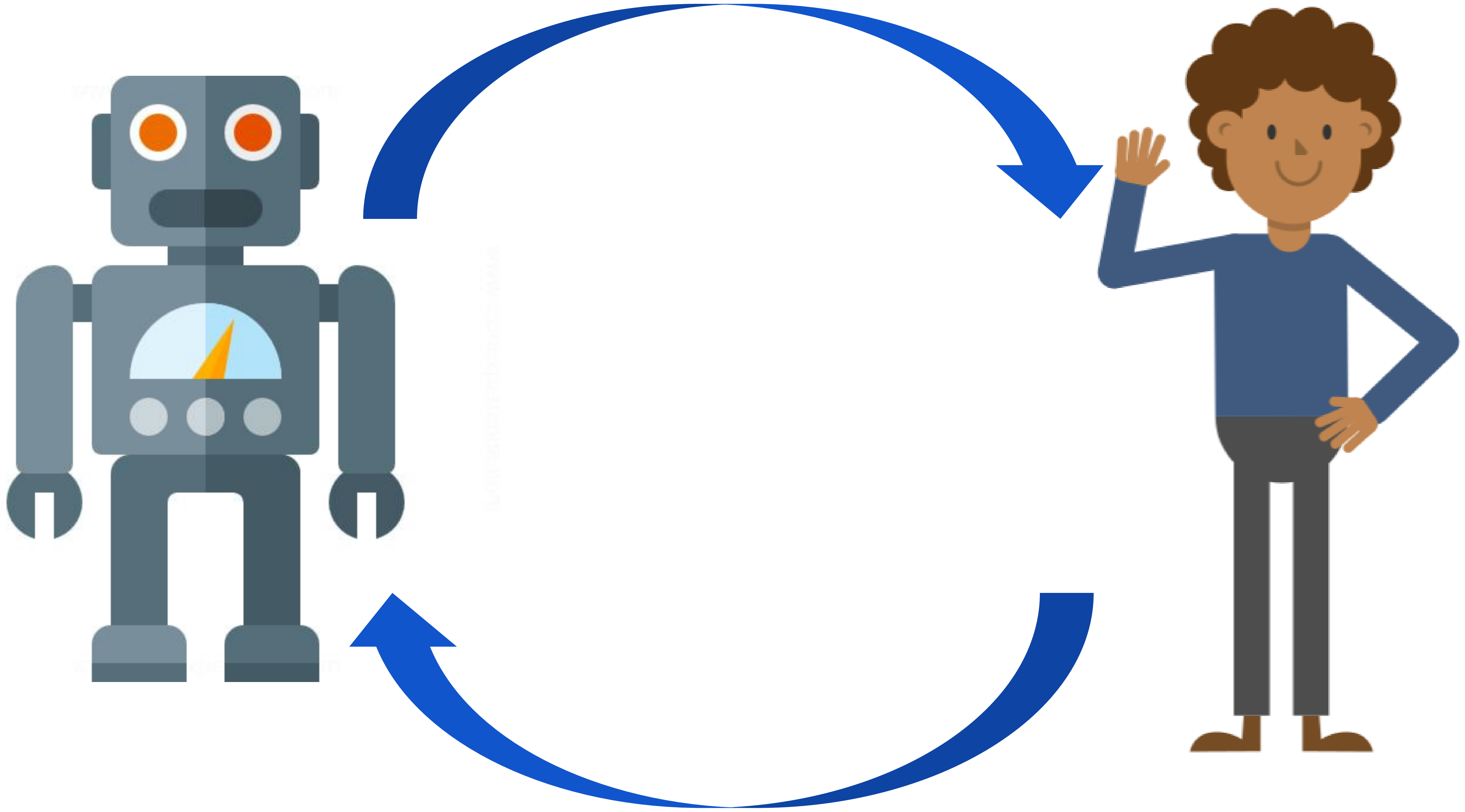


Individual learning can be unsafe, error prone, time consuming

Social learning can enable you to "stand on the shoulders of giants"

Social Reinforcement Learning

Multi-agent interaction

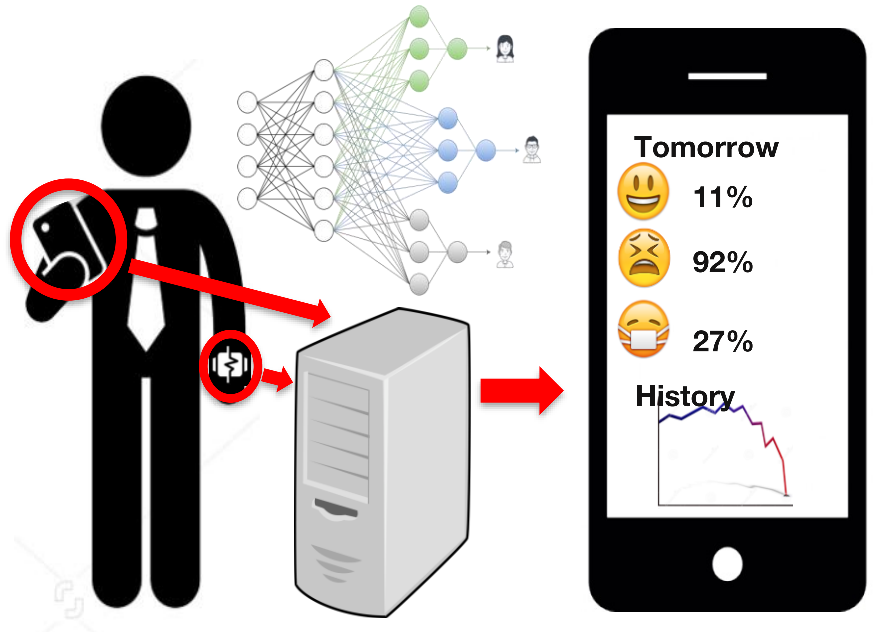


Human-AI interaction

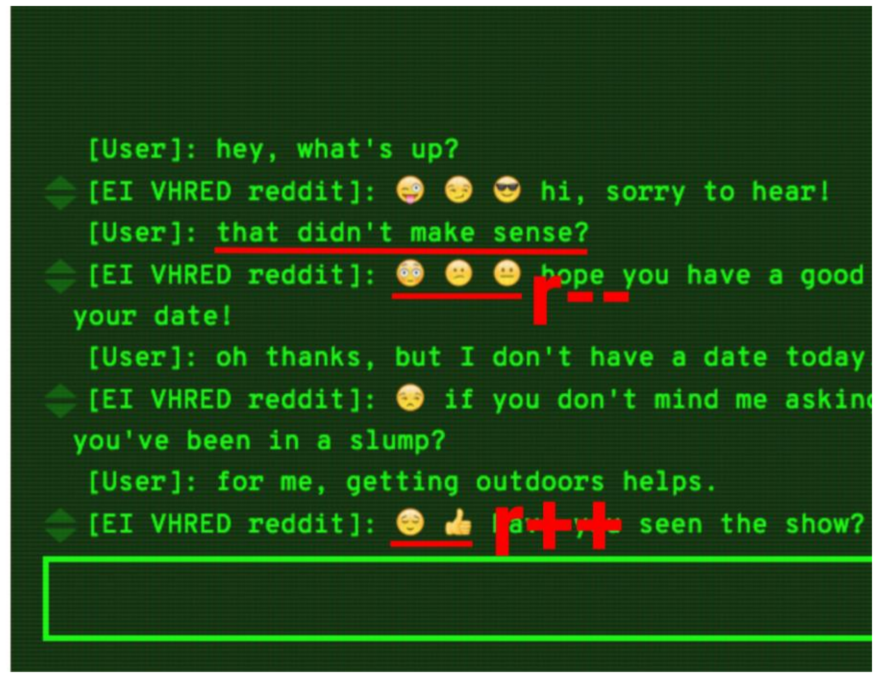
Broad overview of Social RL

Human-AI ⟷ **Multi-agent**

Detecting social & affective cues

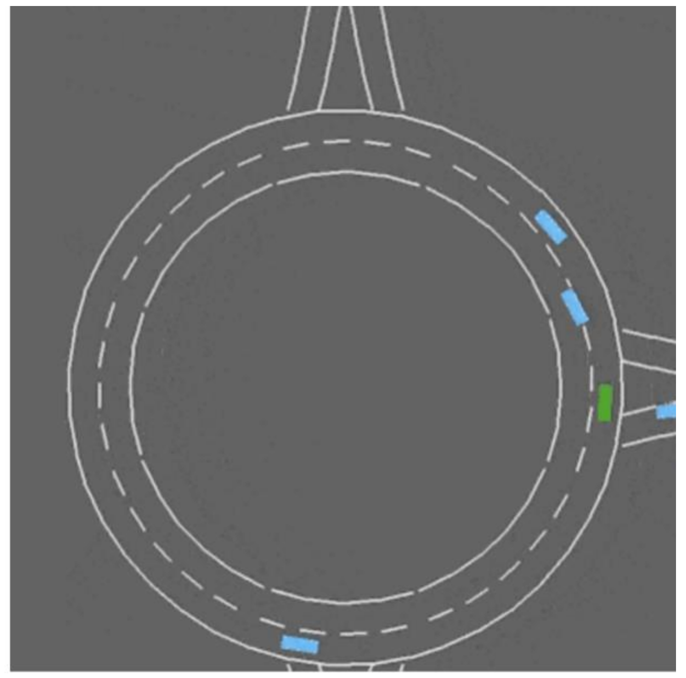


Learning from human social cues

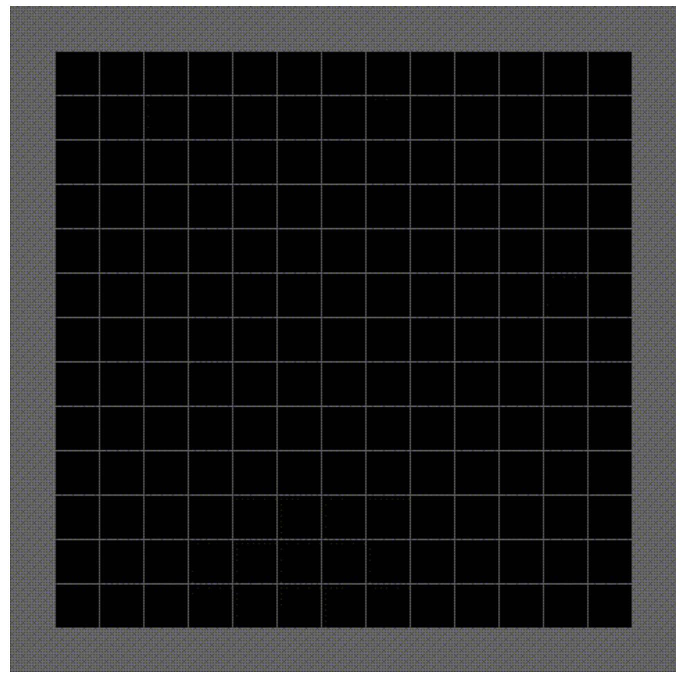


Coordination

Multi-agent social learning



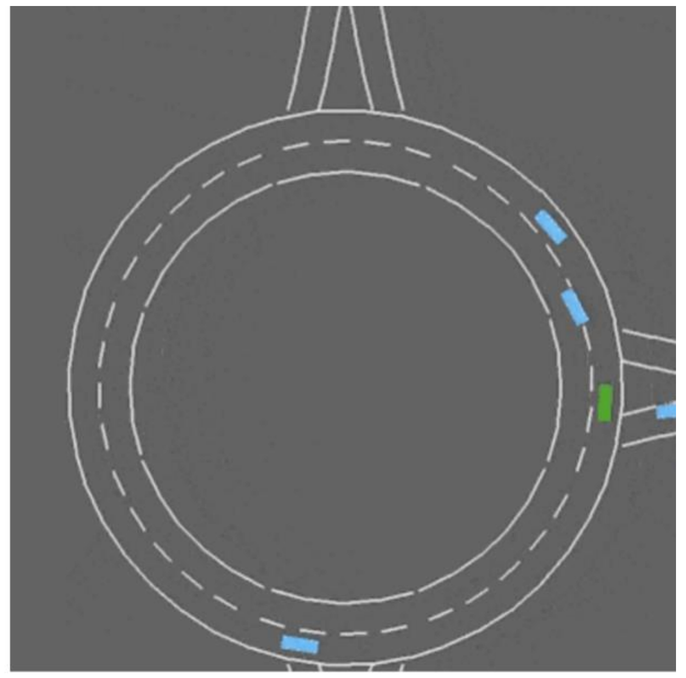
Emergent Complexity



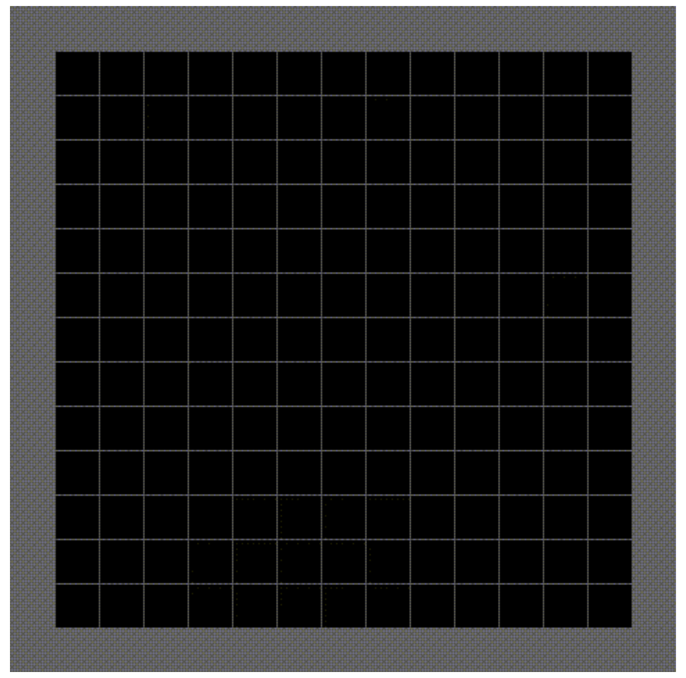
Broad overview of Social RL

Human-AI ⟷ **Multi-agent**

Multi-agent social learning

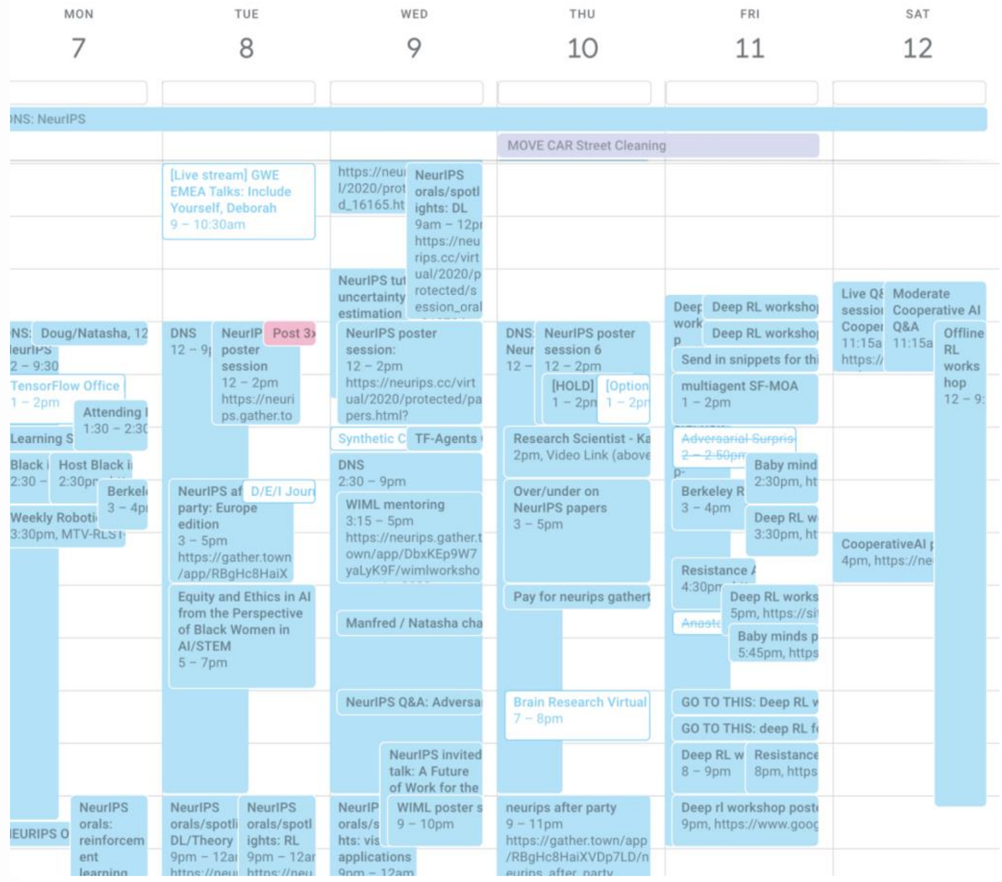


Emergent Complexity



What abilities does an AI assistant need?

https://clipart.me/free-vector/kitchen-robot



Coordinate in shared spaces

Learn from **social cues**

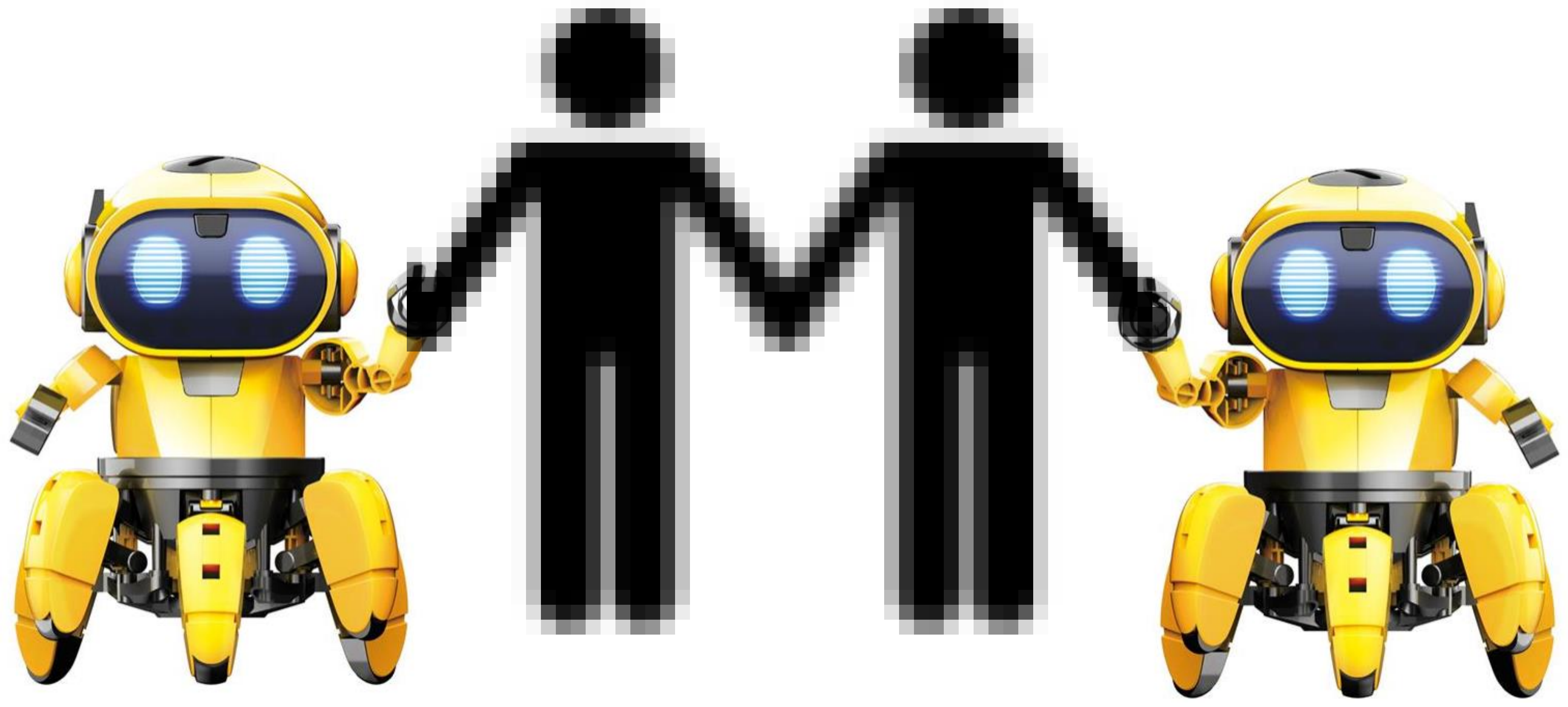
Learning complex tasks

Generalize to new environments

Outline

Multi-agent…

1. Emergent complexity
2. Social Learning

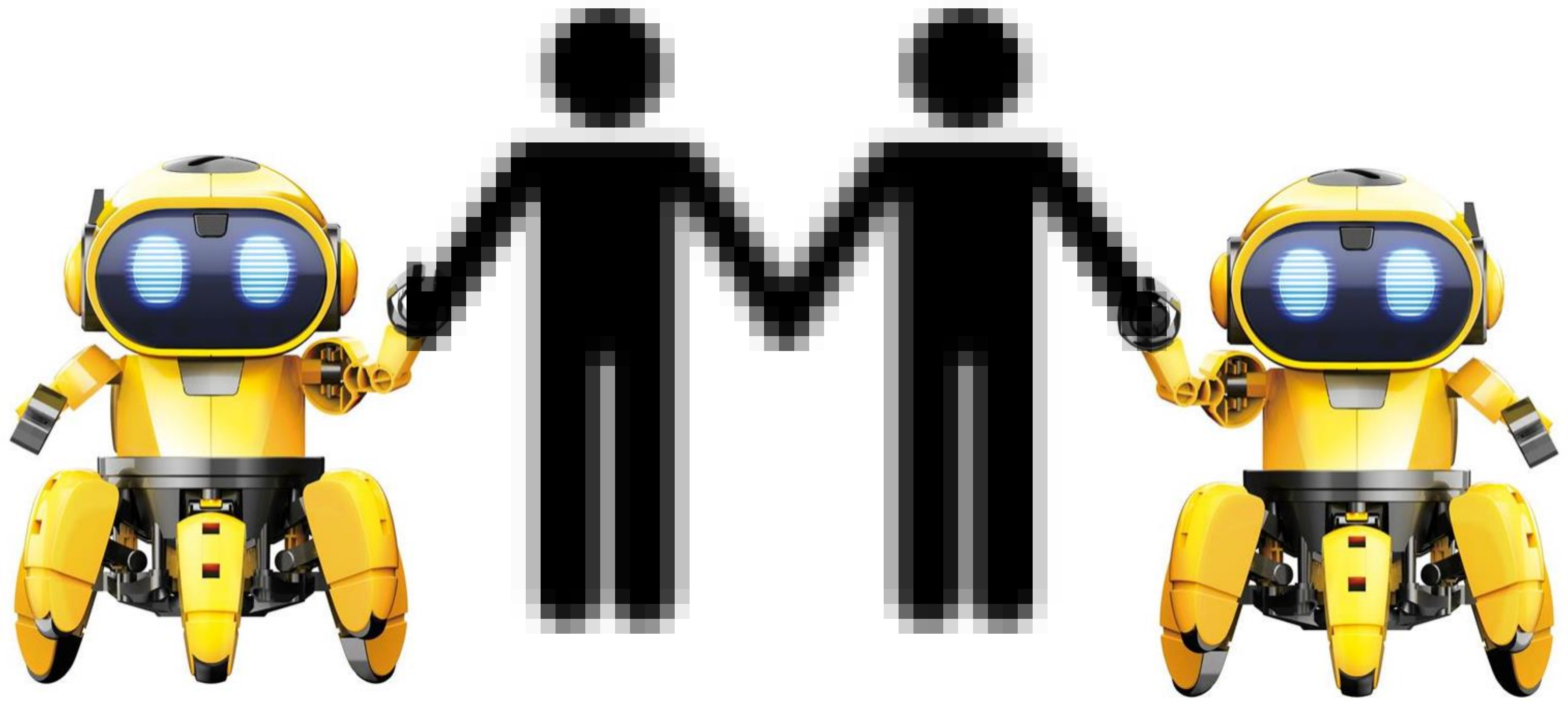


Outline

Multi-agent…

1. **Emergent complexity**

2. Social Learning



Generalization and transfer in RL

Domain Randomization:
[1] Fereshteh Sadeghi and Sergey Levine. Cad2rl: Real single-image flight without a single real image. arXiv preprint arXiv:1611.04201, 2016.
[2] Josh Tobin, Rachel Fong, Alex Ray, Jonas Schneider, Wojciech Zaremba, and Pieter Abbeel. Domain randomization for transferring deep neural networks from simulation to the real world. In 2017 IEEE/RSJ international conference on intelligent robots and systems (IROS), page 23–30. IEEE, 2017.

How can we get:

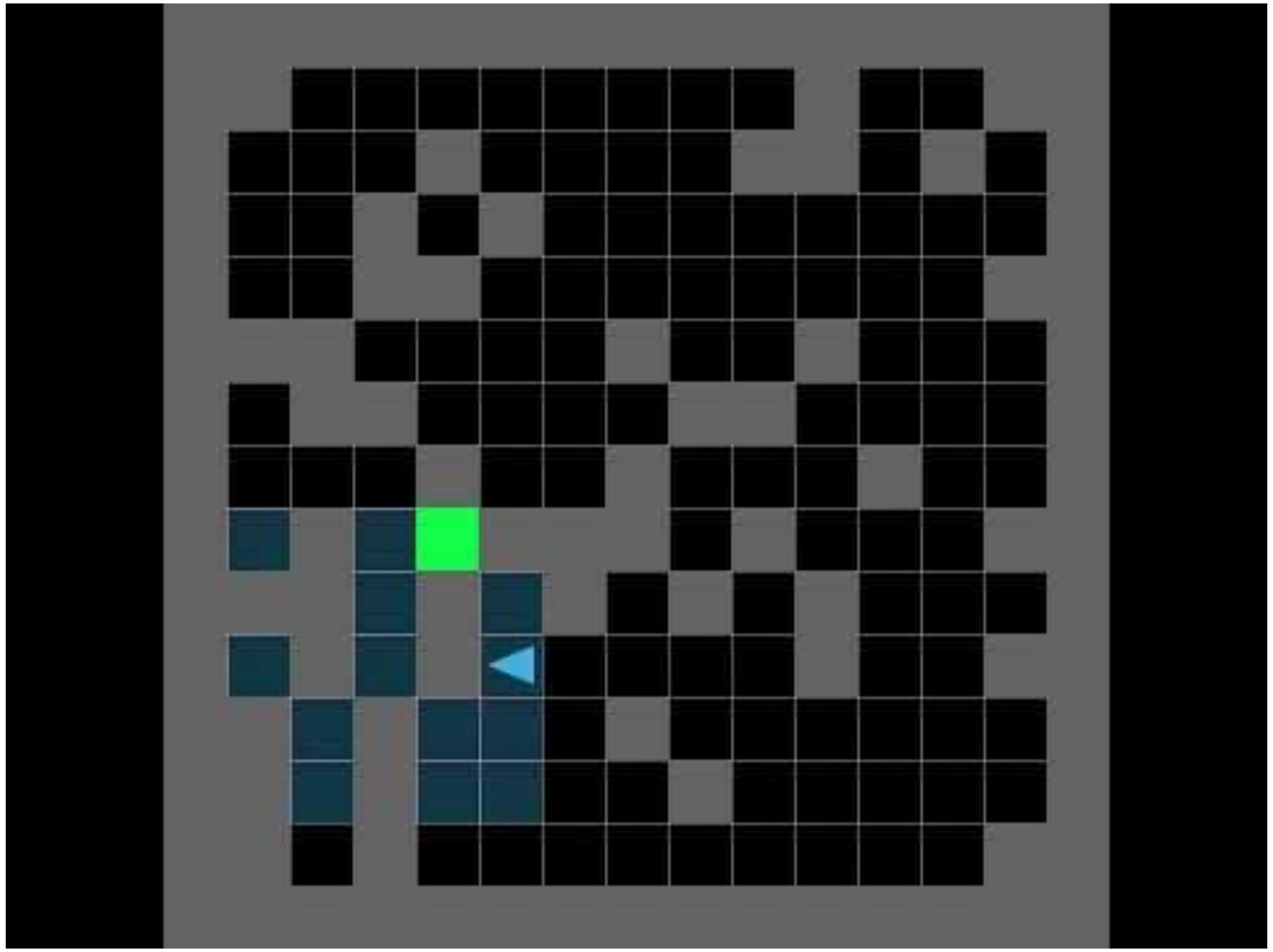
- **complex** training environments
- that cover **unknown, real world** challenges
- without having to program them ourselves **by hand**?

Adversarial environment generation

Environment-generating adversary policy

$$\min_{\pi^E} \max_{\pi^P} R(\tau)$$

Player agent policy

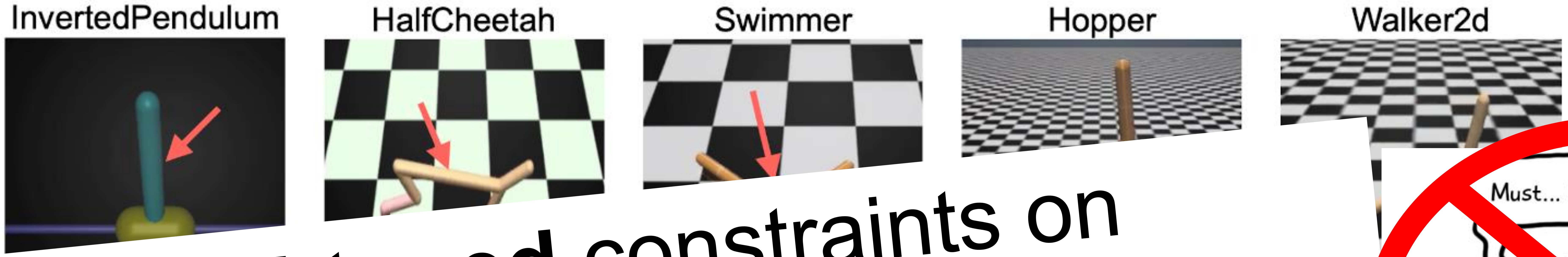


Minimax adversary

Pinto, L., Davidson, J., Sukthankar, R., & Gupta, A. (2017). Robust adversarial reinforcement learning. arXiv preprint arXiv:1703.02702.
Wang, R., Lehman, J., Clune, J., & Stanley, K. O. (2019). Paired open-ended trailblazer (poet): Endlessly generating increasingly complex and diverse learning environments and their solutions. *arXiv preprint arXiv:1901.01753*.

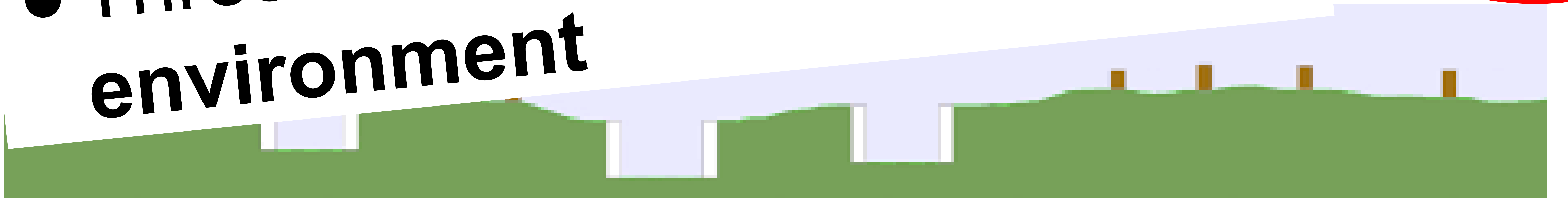
Prior work with deep RL adversaries

RARL



- **Hand-tuned** constraints on adversary
- Thresholds **specific to environment**

POET



Pinto, L., Davidson, J., Sukthankar, R., & Gupta, A. (2017). Robust adversarial reinforcement learning. arXiv preprint arXiv:1703.02702.
Wang, R., Lehman, J., Clune, J., & Stanley, K. O. (2019). Paired open-ended trailblazer (poet): Endlessly generating increasingly complex and diverse learning environments and their solutions. *arXiv preprint arXiv:1901.01753*.

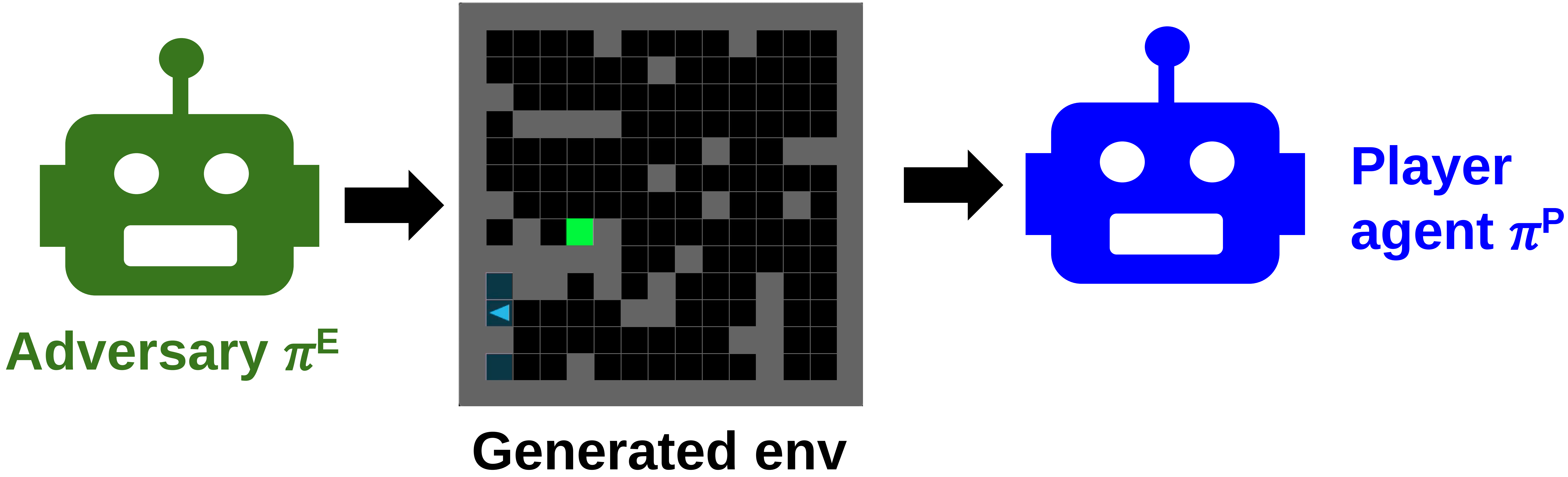
Is there a more elegant way to ensure the adversary does not create **impossible** environments?

Or even environments **tailored to current skill** of the agent? (**automatic curriculum)**

→ **Let’s make this more multi-agent**

PAIRED

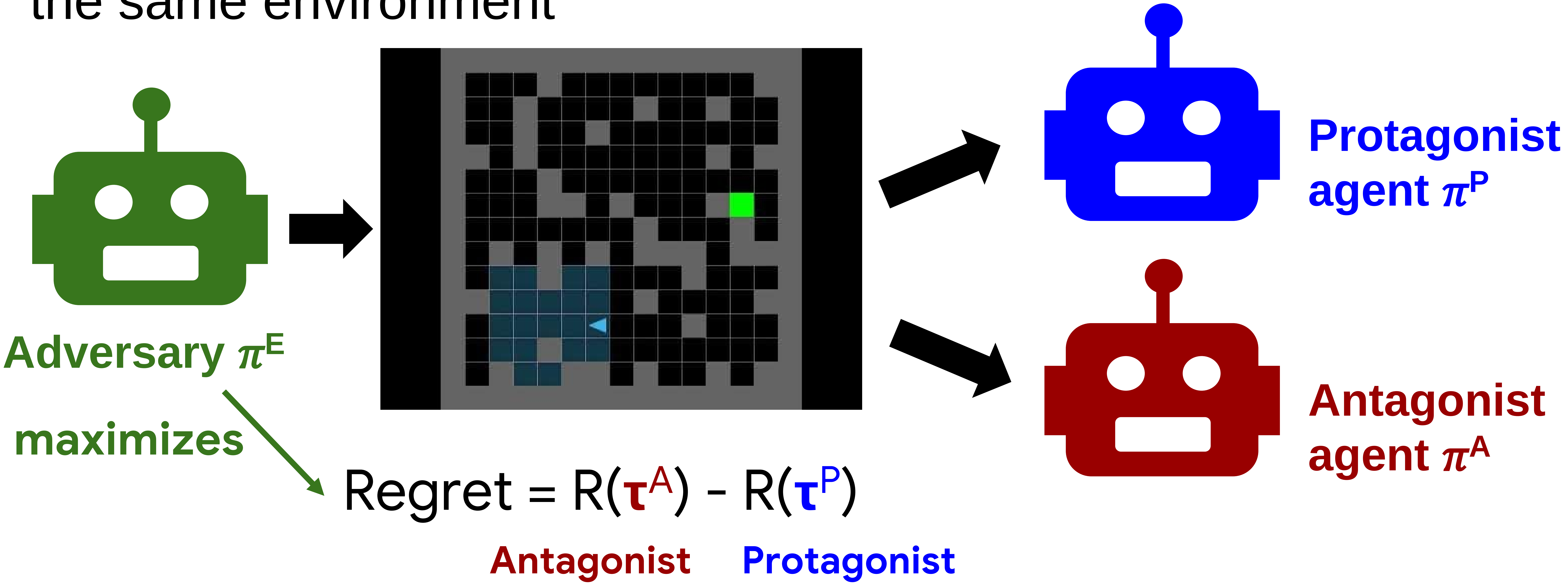
- Constrain the adversary using the performance of a **second agent** in the same environment



Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design. Dennis*, Jaques*, Vinitsky, Bayen, Russell, Critch, Levine (2020). *Equal contribution. Neural Information Processing Systems (NeurIPS) Oral (top 1%).

PAIRED *(Protagonist Antagonist Induced Regret Environment Design)*

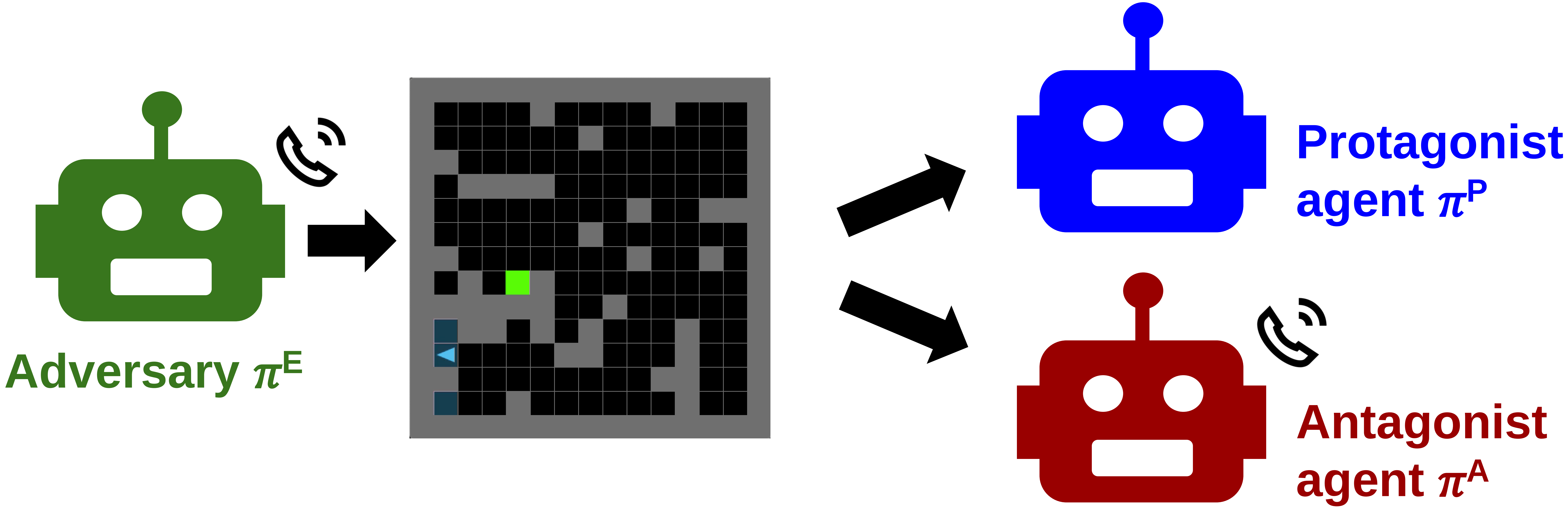
- Constrain the adversary using the performance of a **second agent** in the same environment



Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design. Dennis*, Jaques*, Vinitsky, Bayen, Russell, Critch, Levine (2020). *Equal contribution. Neural Information Processing Systems (NeurIPS) Oral (top 1%).

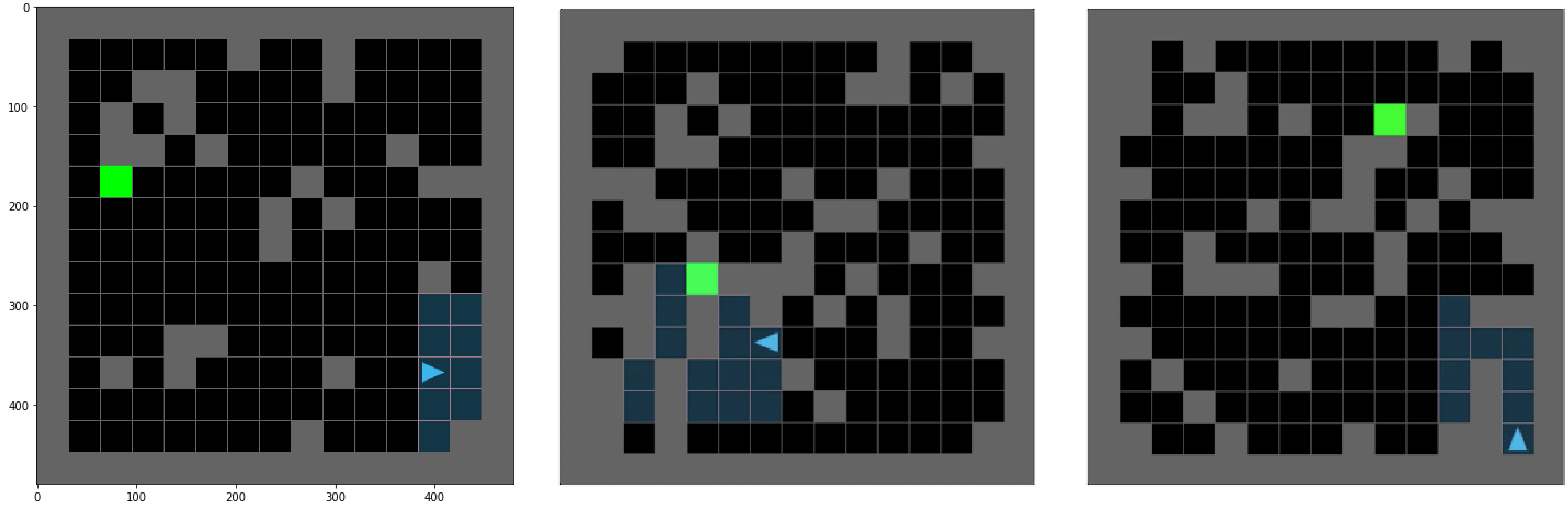
PAIRED as Minimax Regret

Theorem 2. *Let* $(\pi^P, \pi^A, \vec{\theta})$ *be in Nash equilibrium and the pair* $(\pi^A, \vec{\theta})$ *be jointly a best response to* π^P. *Then* $\pi^P \in \underset{\pi^P \in \Pi^P}{\operatorname{argmin}}\{\underset{\pi^A, \vec{\theta} \in \Pi^A \times \Theta^T}{\operatorname{argmax}} \{\textsc{Regret}^{\vec{\theta}}(\pi^P, \pi^A)\}\}$.



Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design. Dennis*, Jaques*, Vinitsky, Bayen, Russell, Critch, Levine (2020). *Equal contribution. Neural Information Processing Systems (NeurIPS) Oral (top 1%).

Generated environments



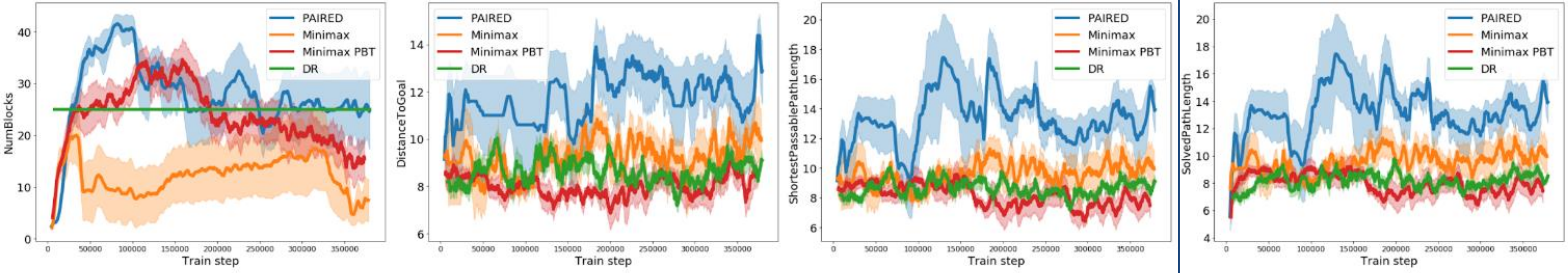
(a) Domain Randomization (b) Minimax Adversarial (c) PAIRED (ours)

Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design. Dennis*, Jaques*, Vinitsky, Bayen, Russell, Critch, Levine (2020). *Equal contribution. Neural Information Processing Systems (NeurIPS) Oral (top 1%).

Emergent complexity and curriculum

Generated environments

Agent learning



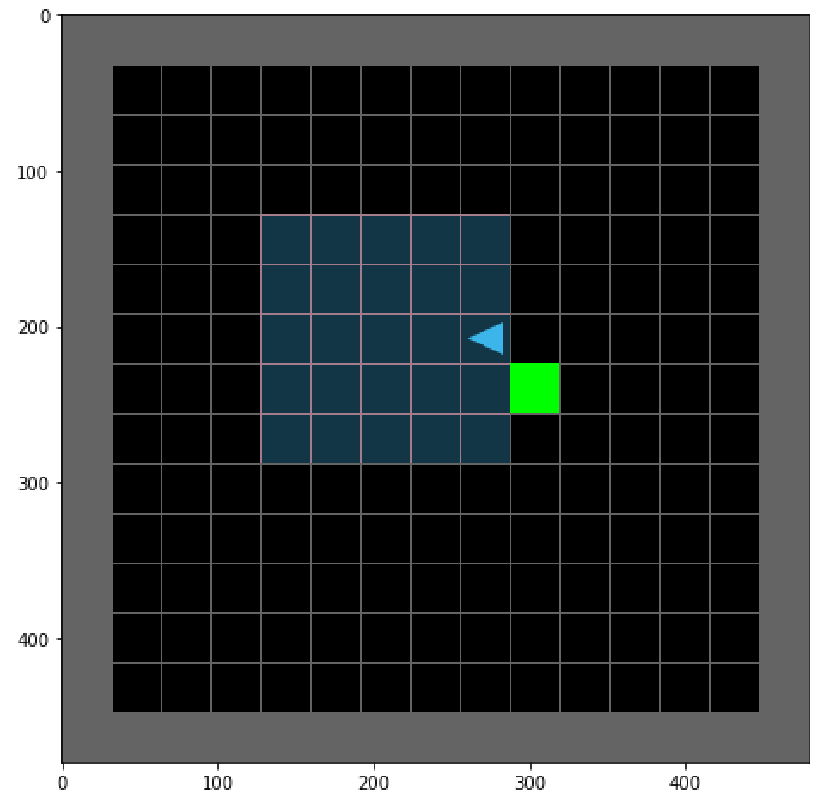
(a) Number of blocks (b) Distance to goal (c) Passable path length (d) Solved path length

- Shortest path length of generated & solved mazes increases although agents were never trained on this objective

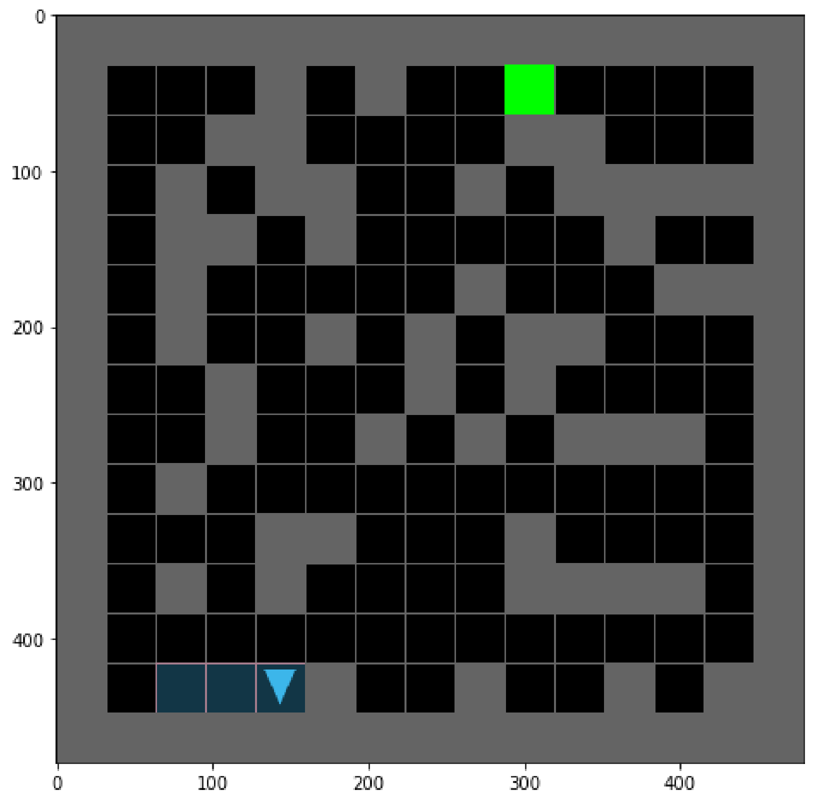
Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design. Dennis*, Jaques*, Vinitsky, Bayen, Russell, Critch, Levine (2020). *Equal contribution. Neural Information Processing Systems (NeurIPS) Oral (top 1%).

Zero-shot transfer

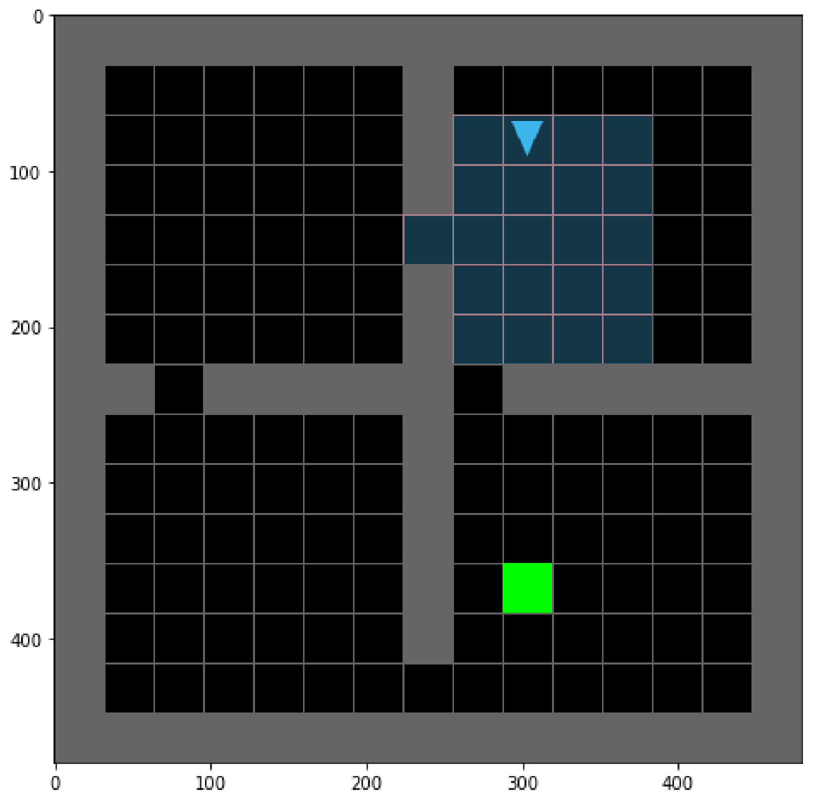
Empty



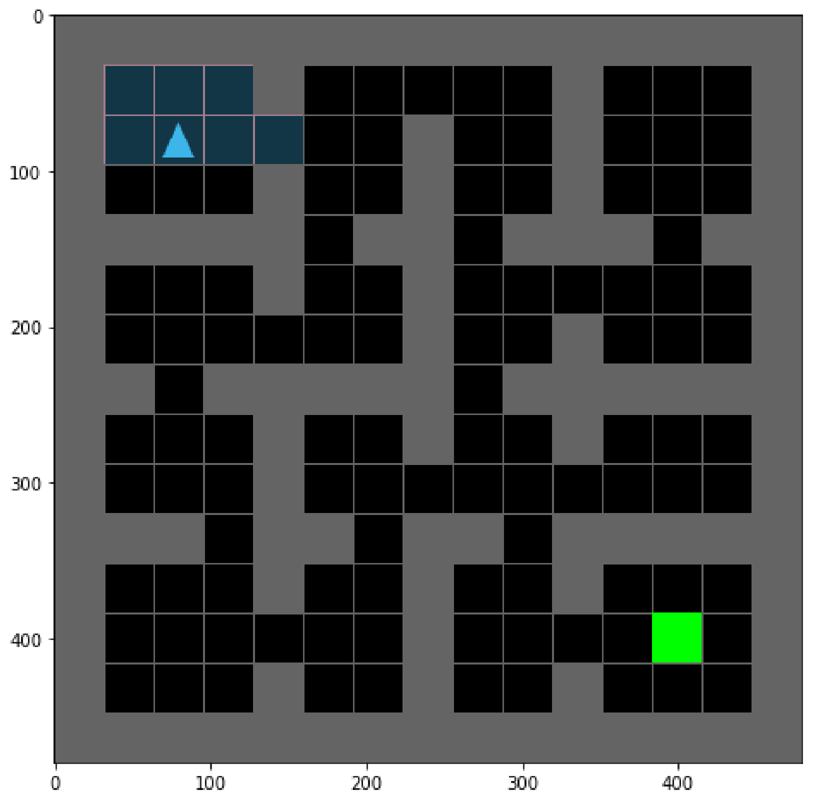
50 Blocks



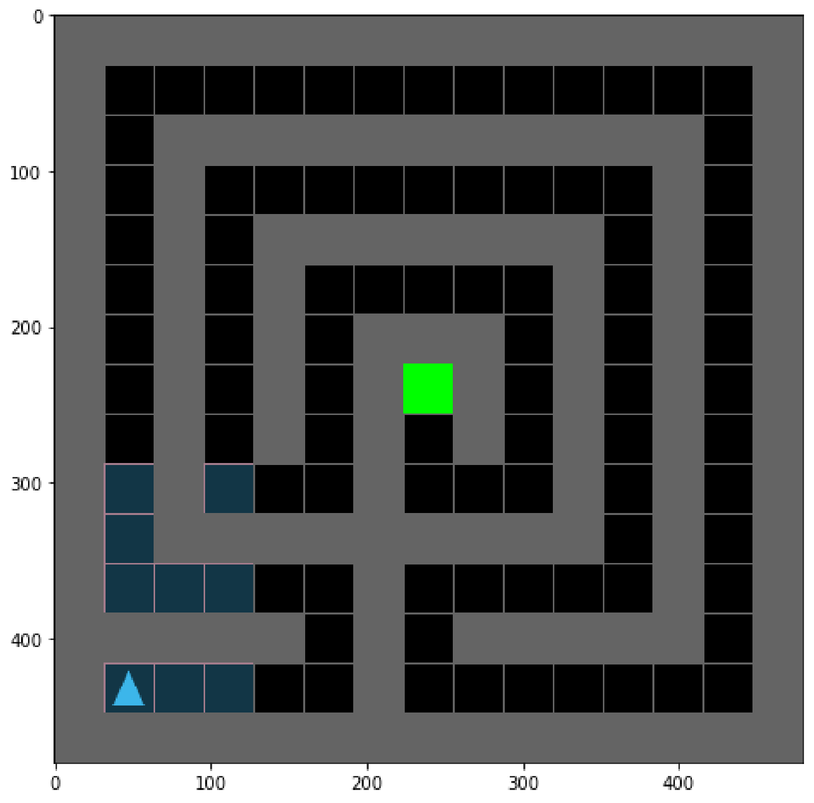
4 Rooms



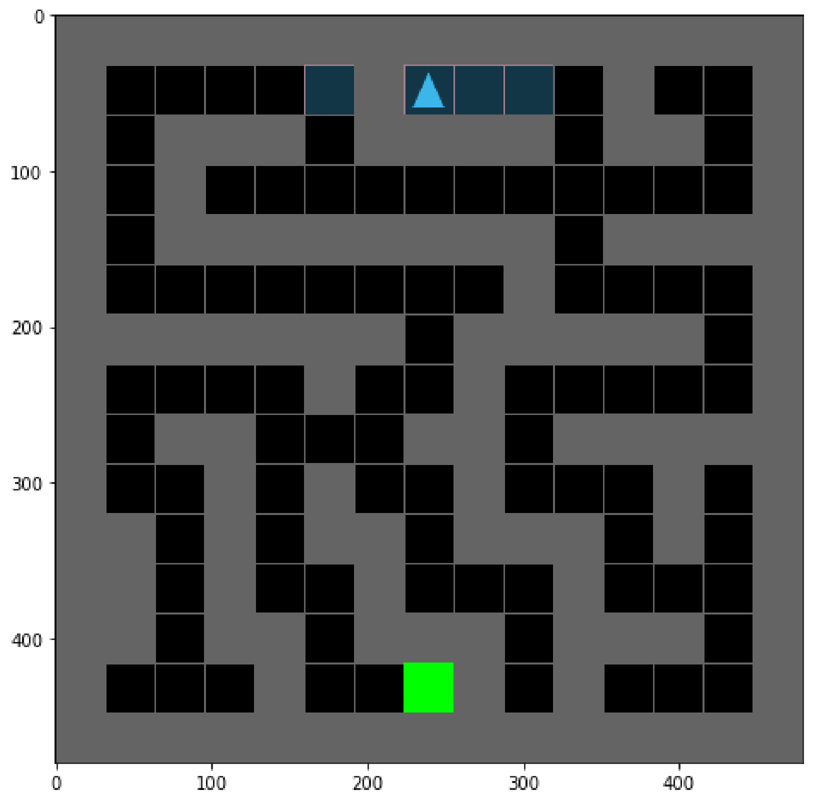
16 Rooms



Labyrinth

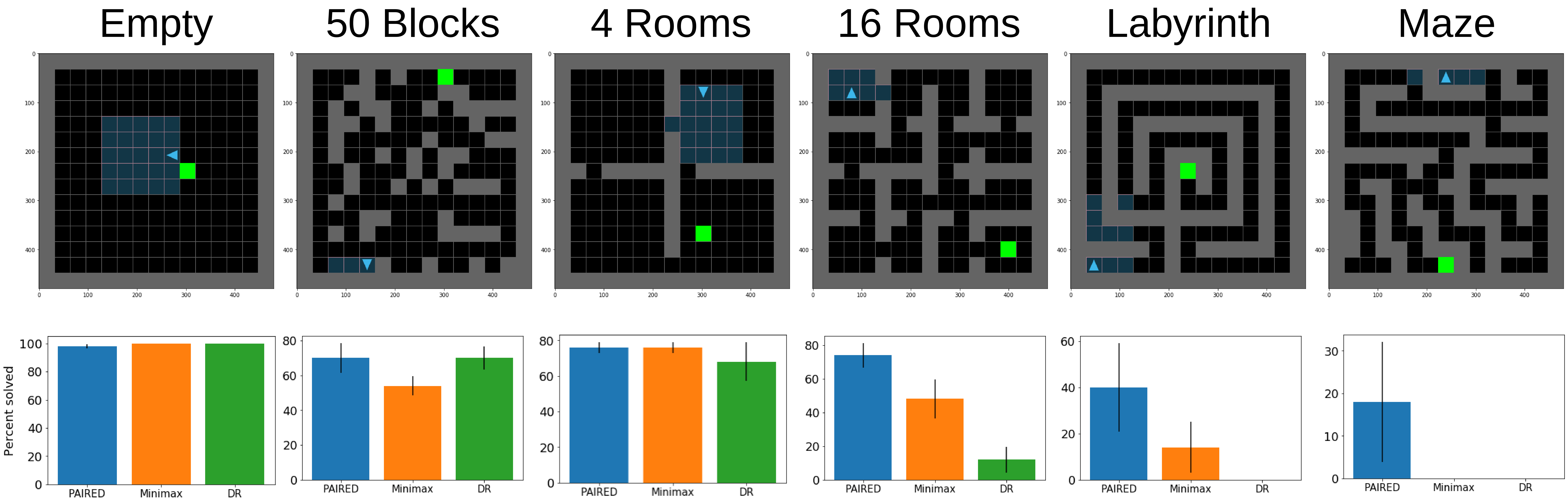


Maze



Emergent Complexity and Zero-shot Transfer via Unsupervised Environment Design. Dennis*, Jaques*, Vinitsky, Bayen, Russell, Critch, Levine (2020). *Equal contribution. Neural Information Processing Systems (NeurIPS) Oral (top 1%).

Zero-shot transfer

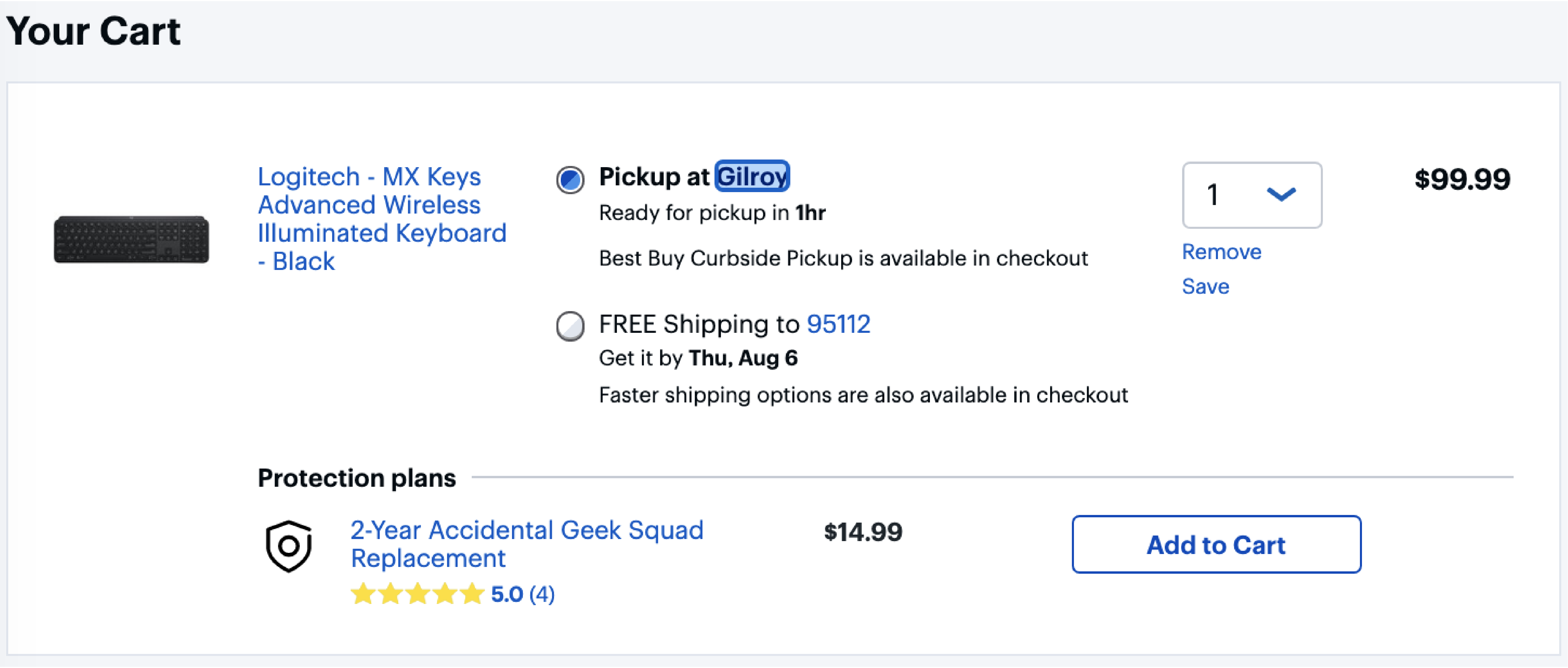


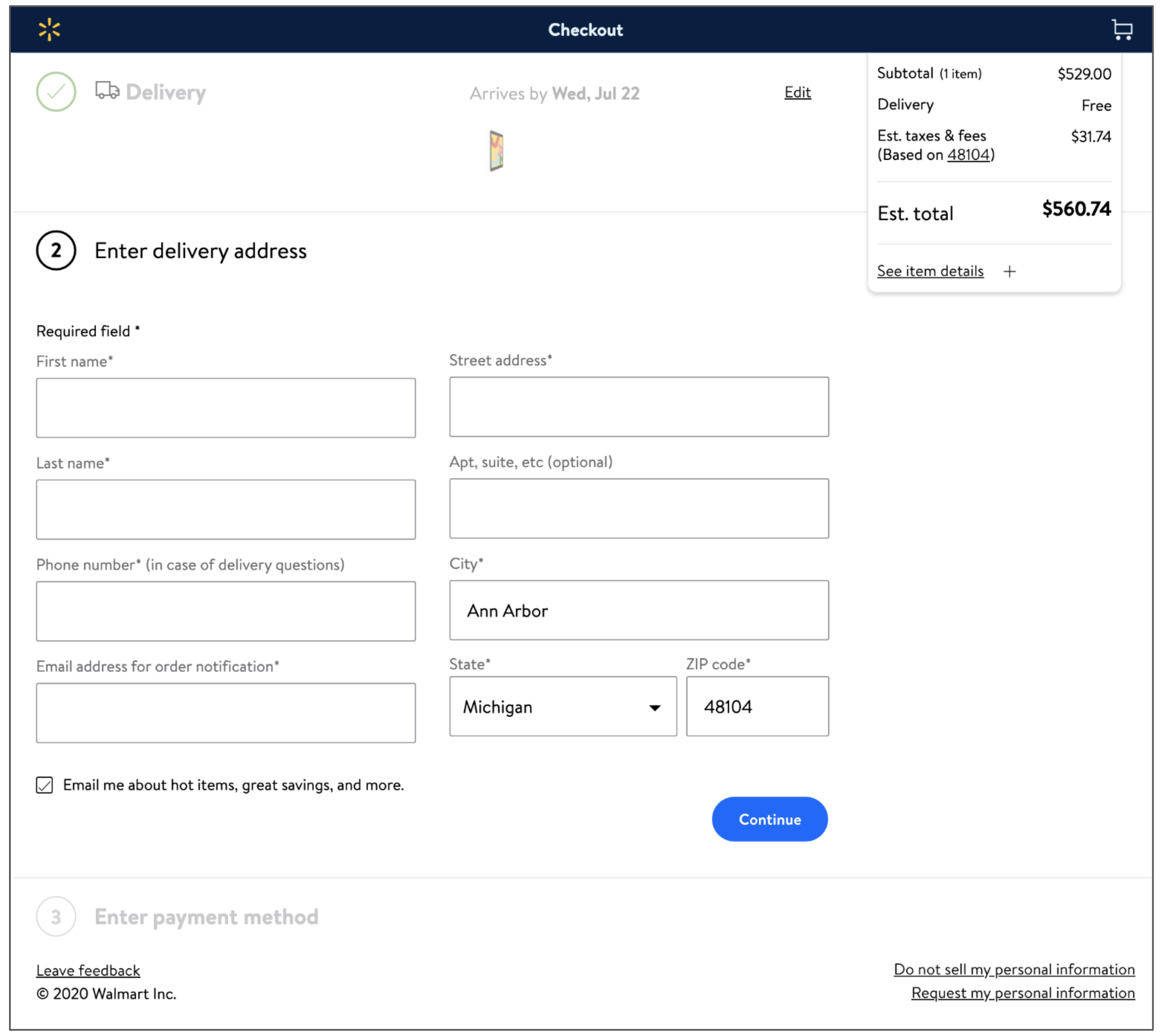
PAIRED
Minimax
Domain Randomization

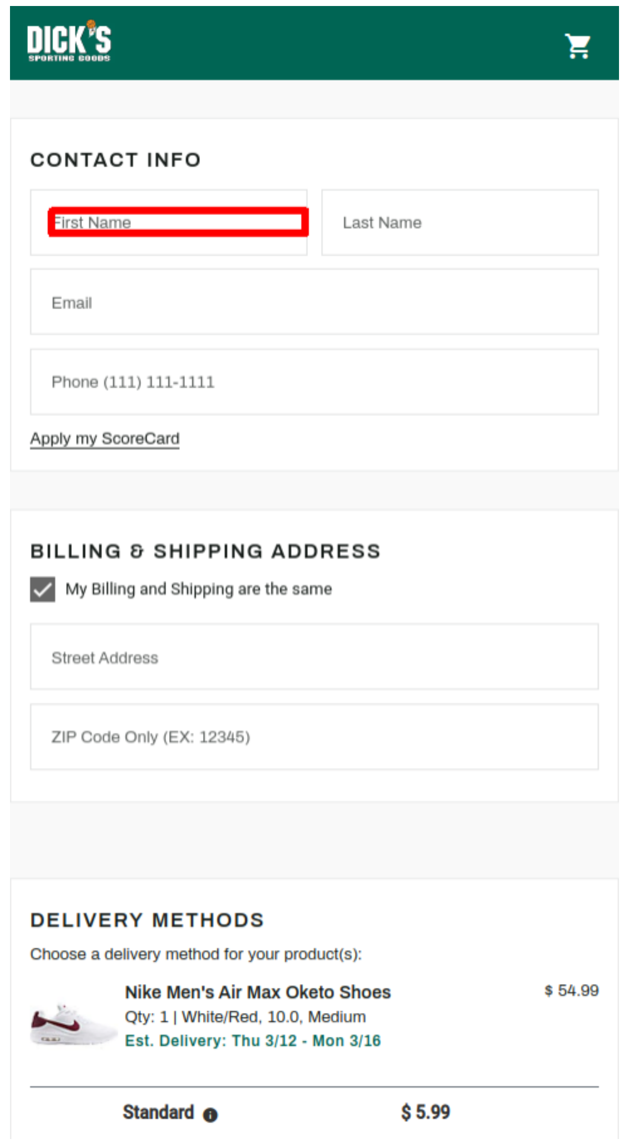
See videos at http://bit.ly/pairedvids!

Environment Generation for Web Navigation

Goal: agent that can automatically **navigate real web pages** to complete tasks for users

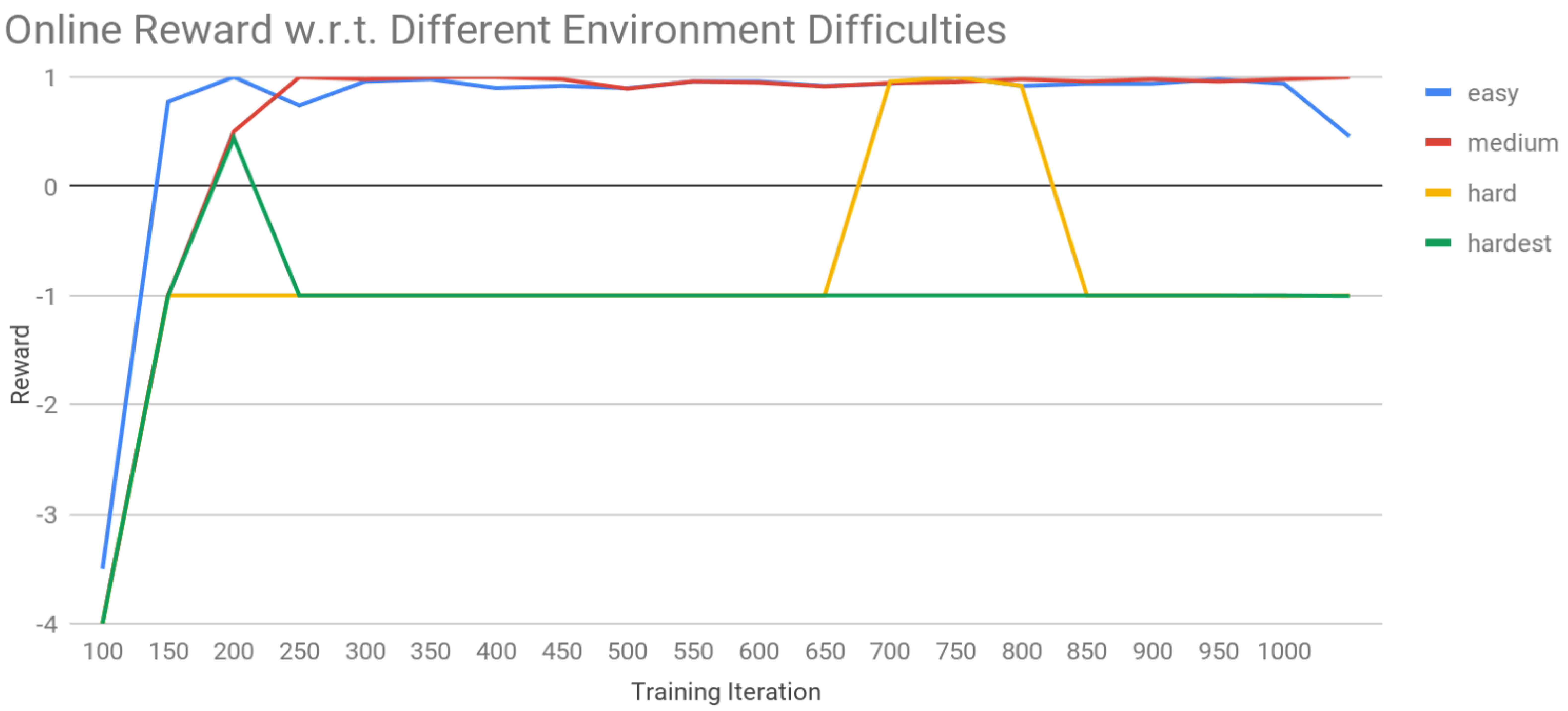






Environment Generation for Zero-Shot Compositional Reinforcement Learning. Gur, Jaques, Miao, Choi, Malta, Tiwari, Lee, Faust (2021). Neural Information Processing Systems (NeurIPS).

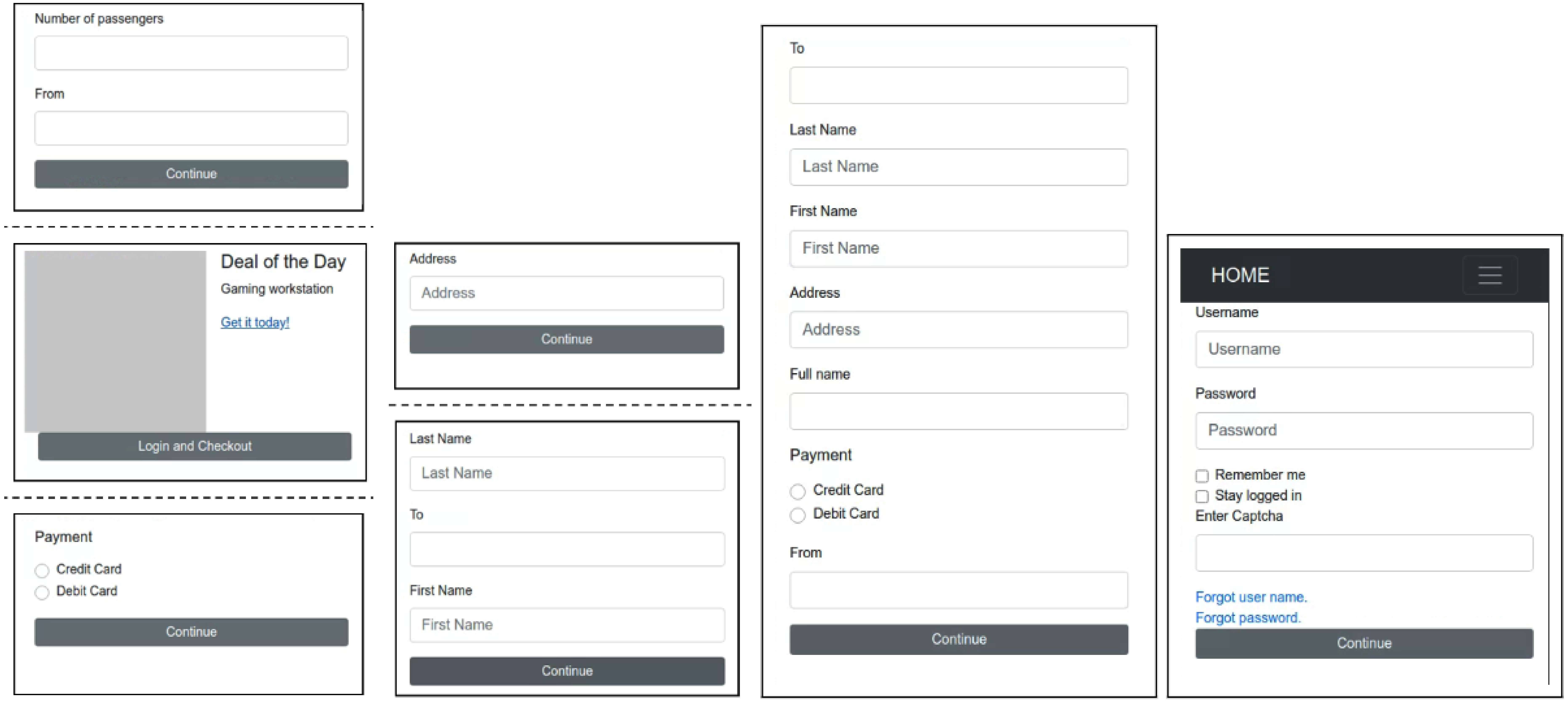
Web navigation



- **Issue:** initial approach involved **hand-programming** curriculum of easy, medium, hard sites
- **Idea:** PAIRED adversary generates a curriculum of websites

Environment Generation for Zero-Shot Compositional Reinforcement Learning. Gur, Jaques, Miao, Choi, Malta, Tiwari, Lee, Faust (2021). Neural Information Processing Systems (NeurIPS).

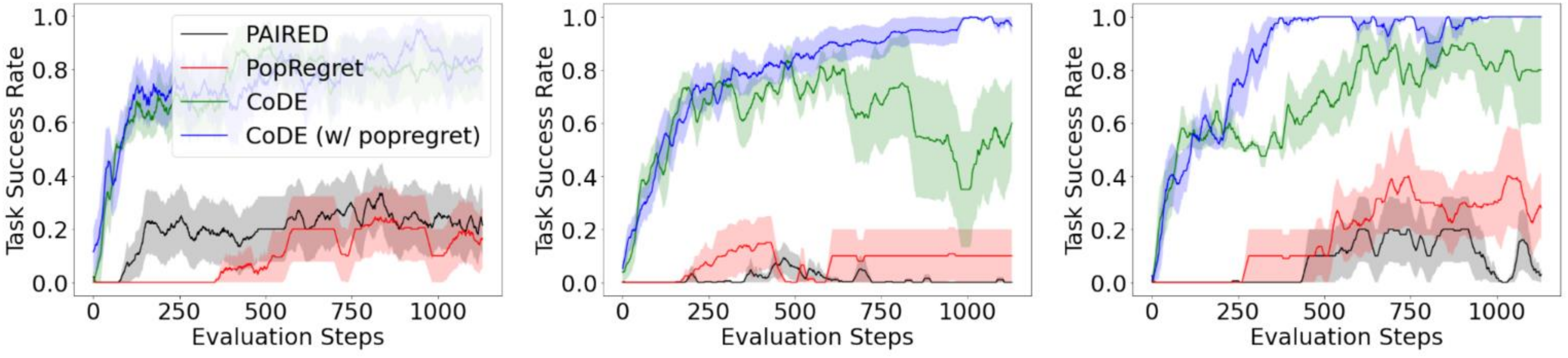
Web navigation - generated environments



(a) Early training (b) Mid training (c) Late training (d) Test

Environment Generation for Zero-Shot Compositional Reinforcement Learning. Gur, Jaques, Miao, Choi, Malta, Tiwari, Lee, Faust (2021). Neural Information Processing Systems (NeurIPS).

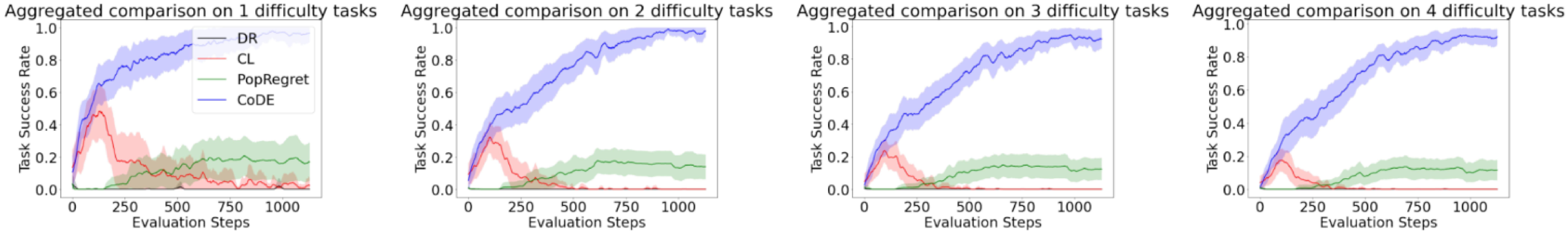
Web navigation - task success



(a) Payment (b) Shopping (c) Flight booking

Deployed to **real websites**

- **93%+ accuracy on initial tasks**



(a) Difficulty level 1 (b) Difficulty level 2 (c) Difficulty level 3 (d) Difficulty level 4

- **4x more successful than SOTA prior work**
- Reaches more than **95% task success** across all difficulty levels

Environment Generation for Zero-Shot Compositional Reinforcement Learning. Gur, Jaques, Miao, Choi, Malta, Tiwari, Lee, Faust (2021). Neural Information Processing Systems (NeurIPS).

Conclusions

- **Multi-agent training** can be an effective tool for improving **learning** and **generalization** *even for single-agent settings*

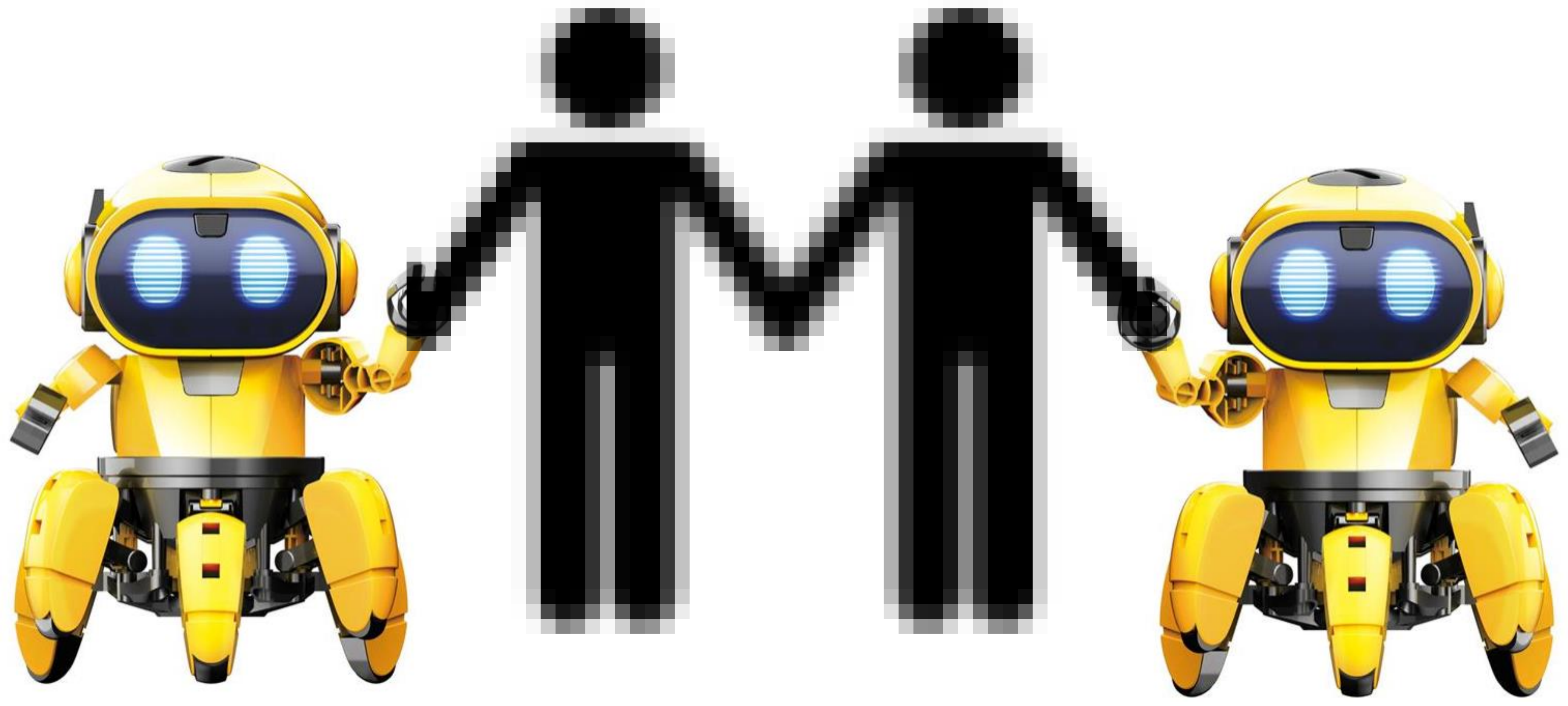
- Advantages of **PAIRED:**
  - The adversary builds **feasible yet challenging** environments that make the protagonist more robust
  - **Regret** minimization creates an **automatic curriculum**

Outline

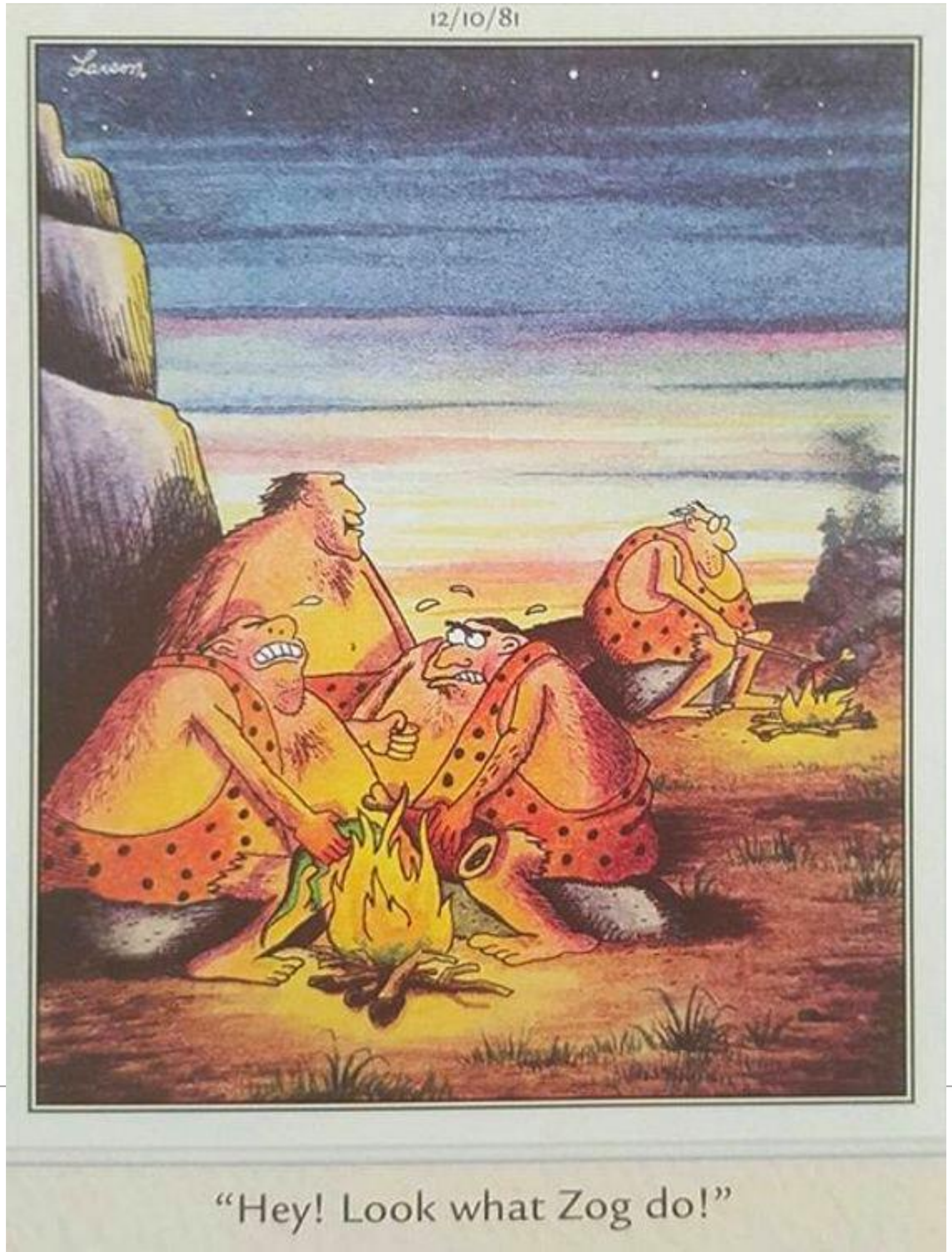
Multi-agent…

1. Emergent complexity

2. **Social Learning**



How social learning can accelerate learning



Social learning enables you to “stand on the shoulders of giants”

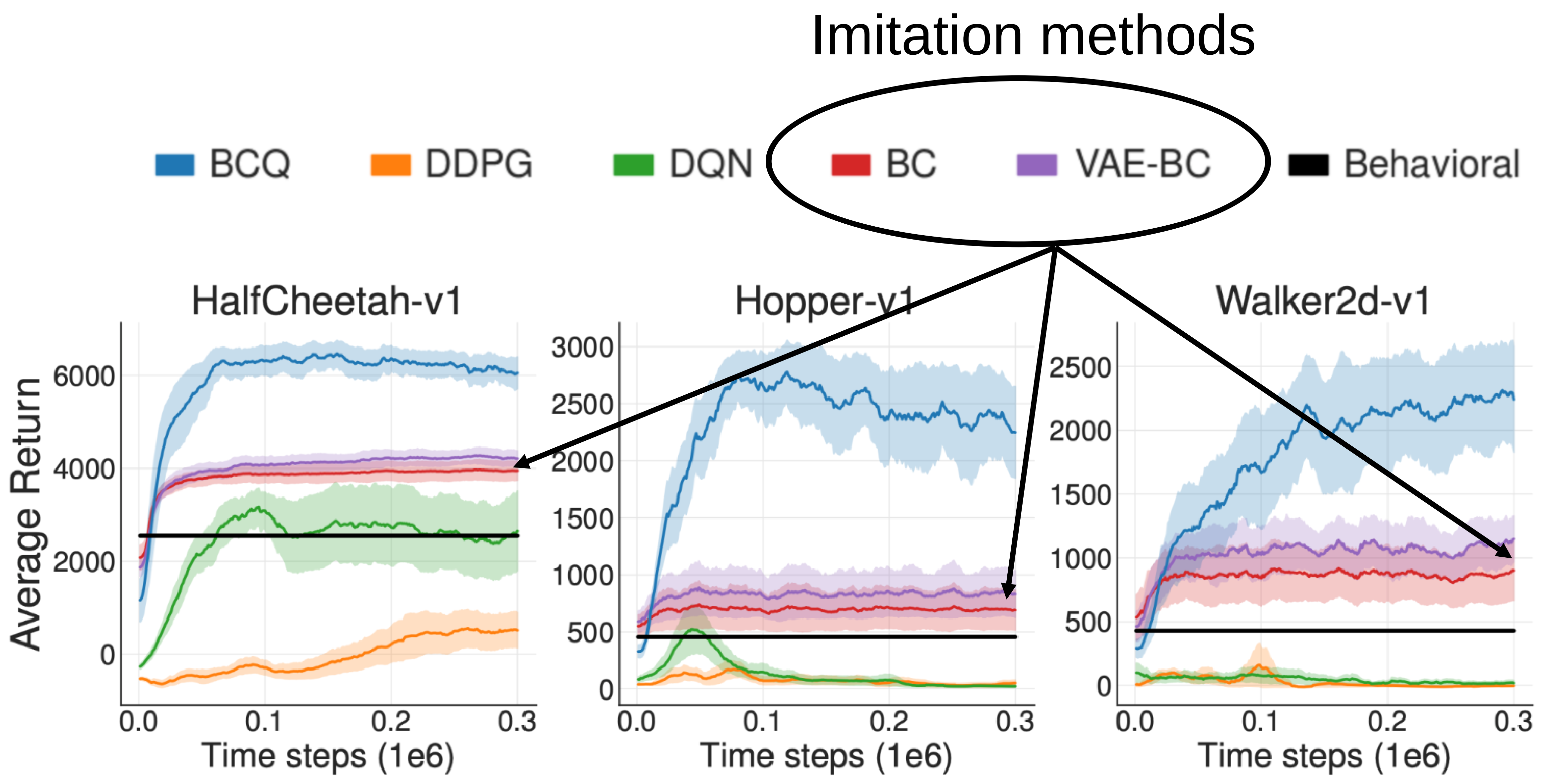
Why not just do imitation learning?

= supervised learning on <s,a> expert dataset

- Requires access to **specially curated** expert trajectories

- **Copy data exactly**; perform poorly if data comes from imperfect experts

- Resulting policies are brittle, **do not generalize**



(d) Imperfect demonstrations performance

From Fujimoto, Scott, David Meger, and Doina Precup. "Off-policy deep reinforcement learning without exploration." International Conference on Machine Learning. PMLR, 2019.

Social Learning with Multi-Agent RL

- Works in more **naturalistic** settings. When other agents:
  - Are **not motivated to teach** you
  - May or **may not have relevant expertise**

- **Do not have to copy** exactly

- Social learners can **learn how to acquire** info from other agents
  - **Generalize/adapt** to new environments

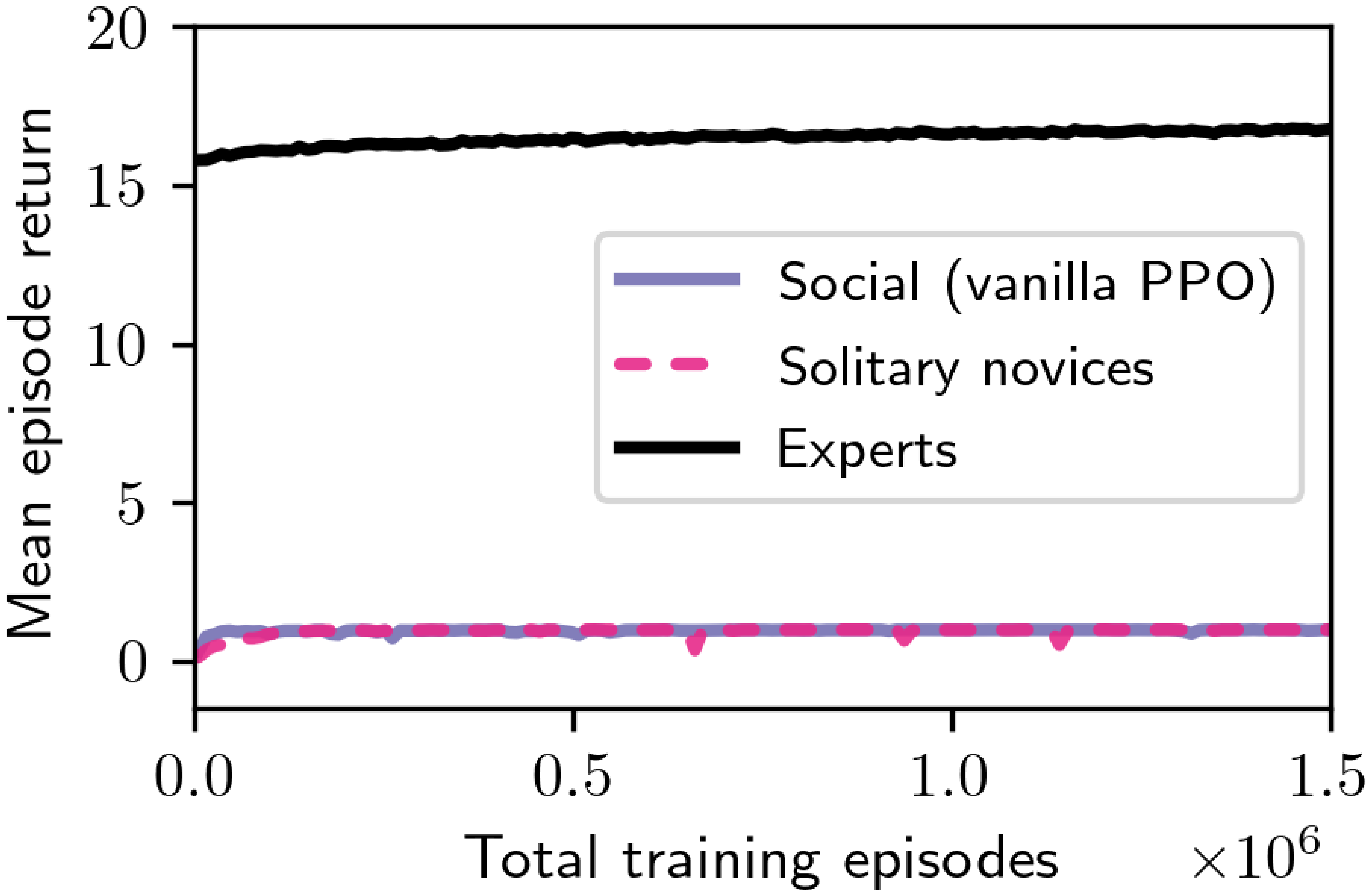
https://www.sfgate.com/traffic/

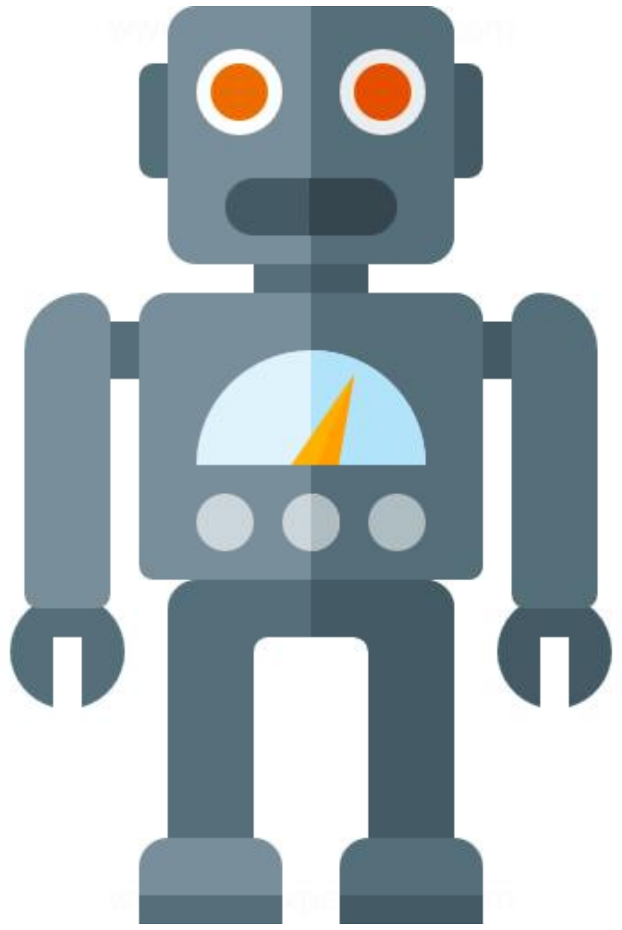
Landolfi, N. C., & Dragan, A. D. (2018, October). Social cohesion in autonomous driving. In 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS) (pp. 8118-8125). IEEE.

Learning Social Learning

Kamal Ndousse

Multi-agent env with **partial observability** and **no privileged access** to other agents' states or actions (unlike imitation learning)



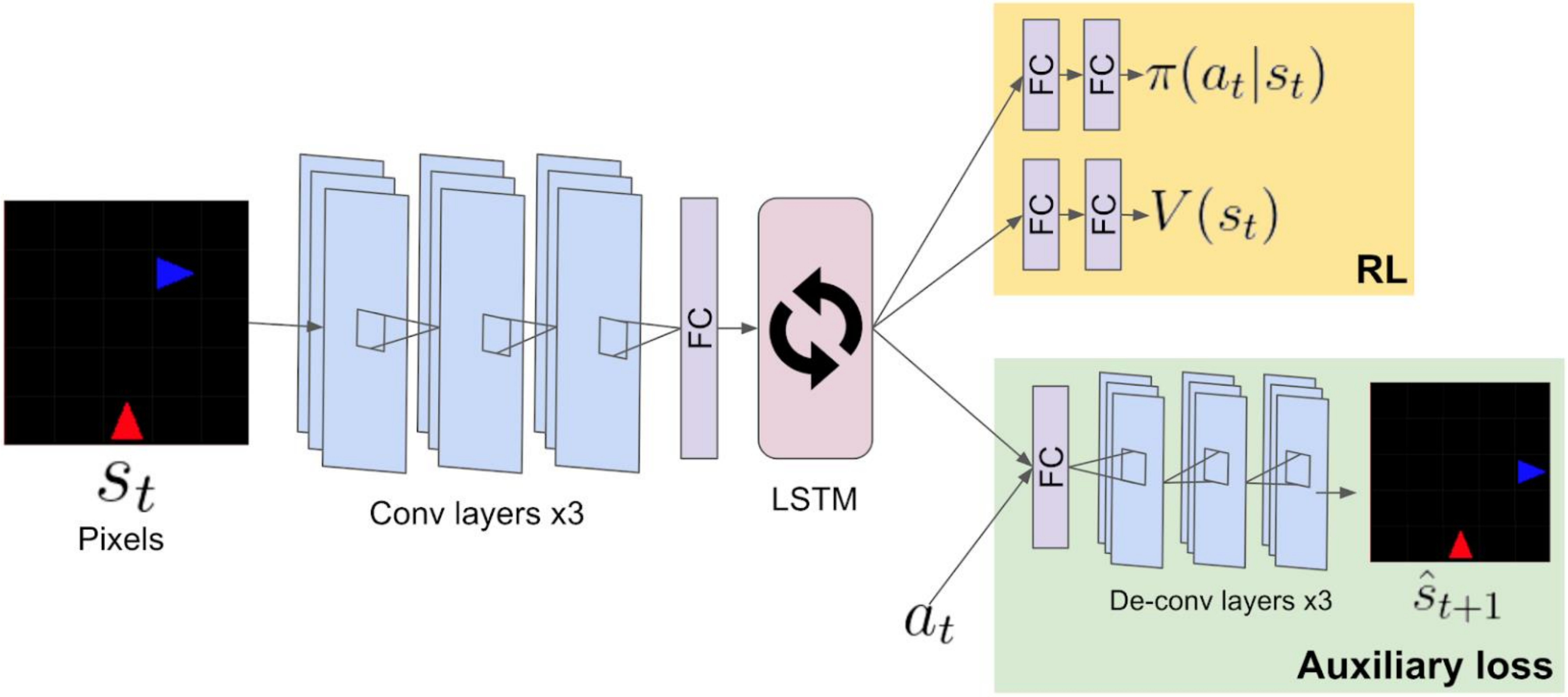


<<

Vanilla RL agents fail to learn from experts in their environment

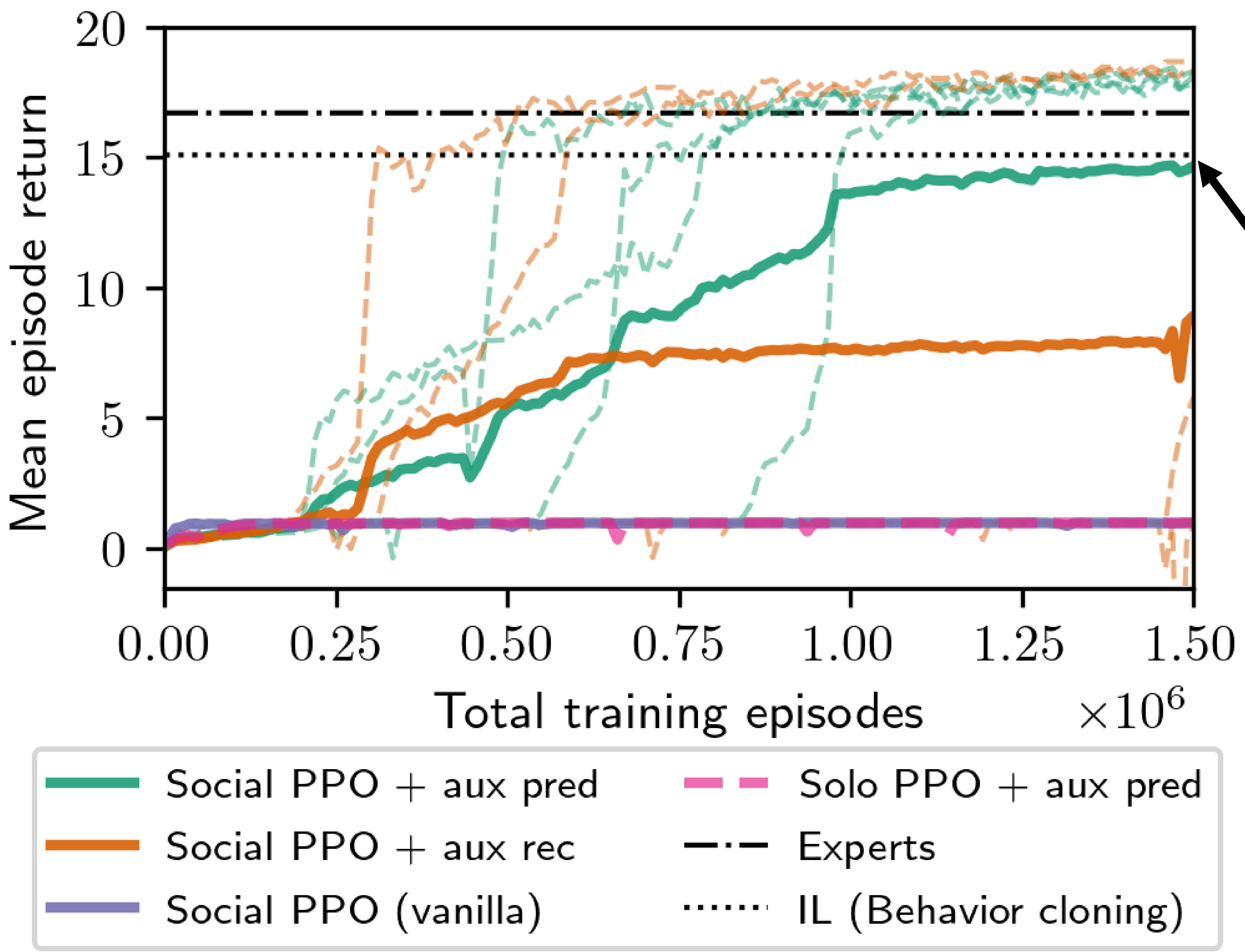
Learning Social Learning. Ndousse, Eck, Levine, Jaques (2020). Best Paper at the Neural Information Processing Systems (NeurIPS) workshop on Cooperative AI; ICML 2021.

Augment agents with auxiliary prediction loss



Learning Social Learning. Ndousse, Eck, Levine, Jaques (2020). Best Paper at the Neural Information Processing Systems (NeurIPS) workshop on Cooperative AI; ICML 2021.

Agents are able to use social learning

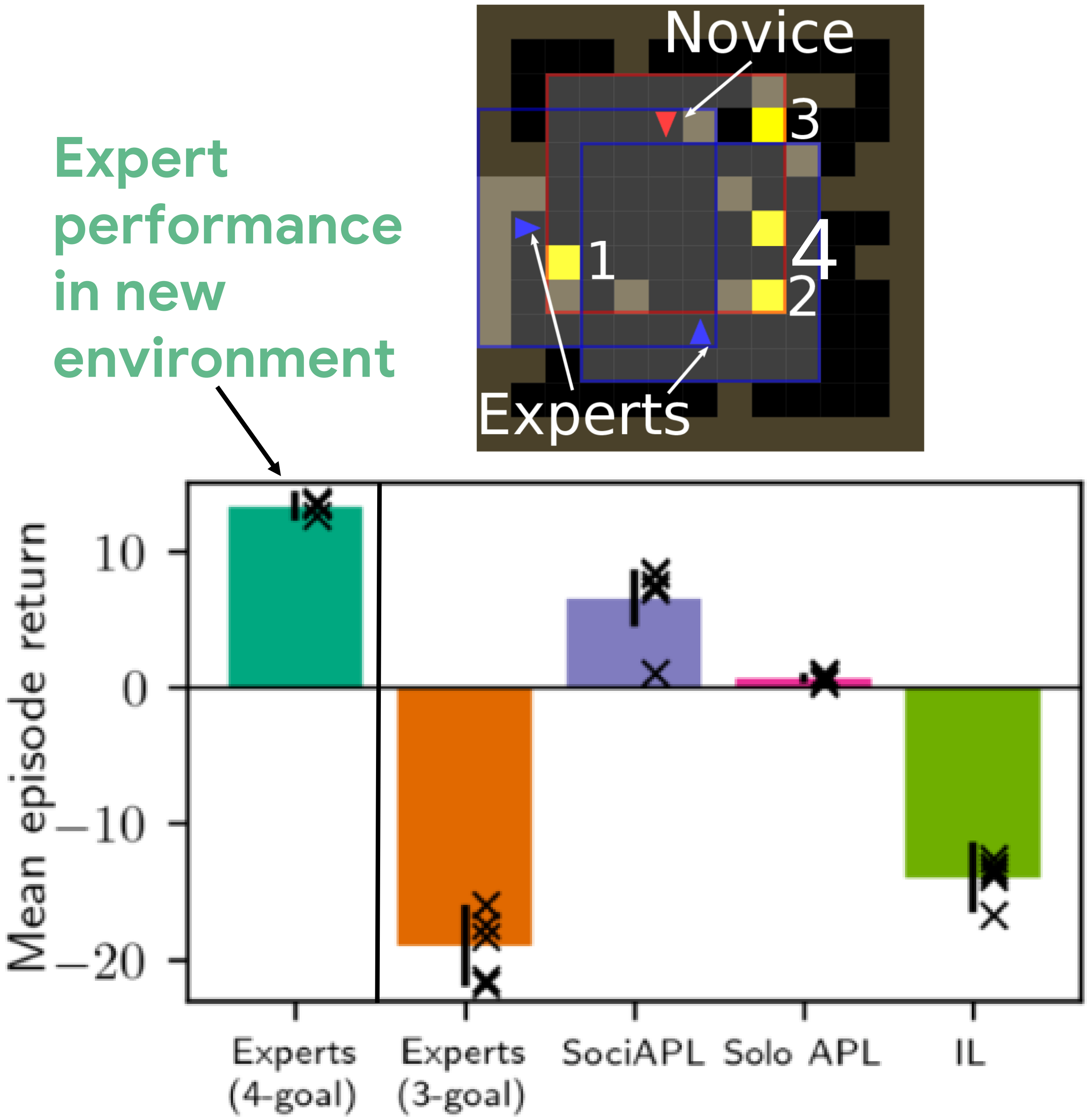


Same as imitation learning (IL) but no access to <s,a,s'> of other agents

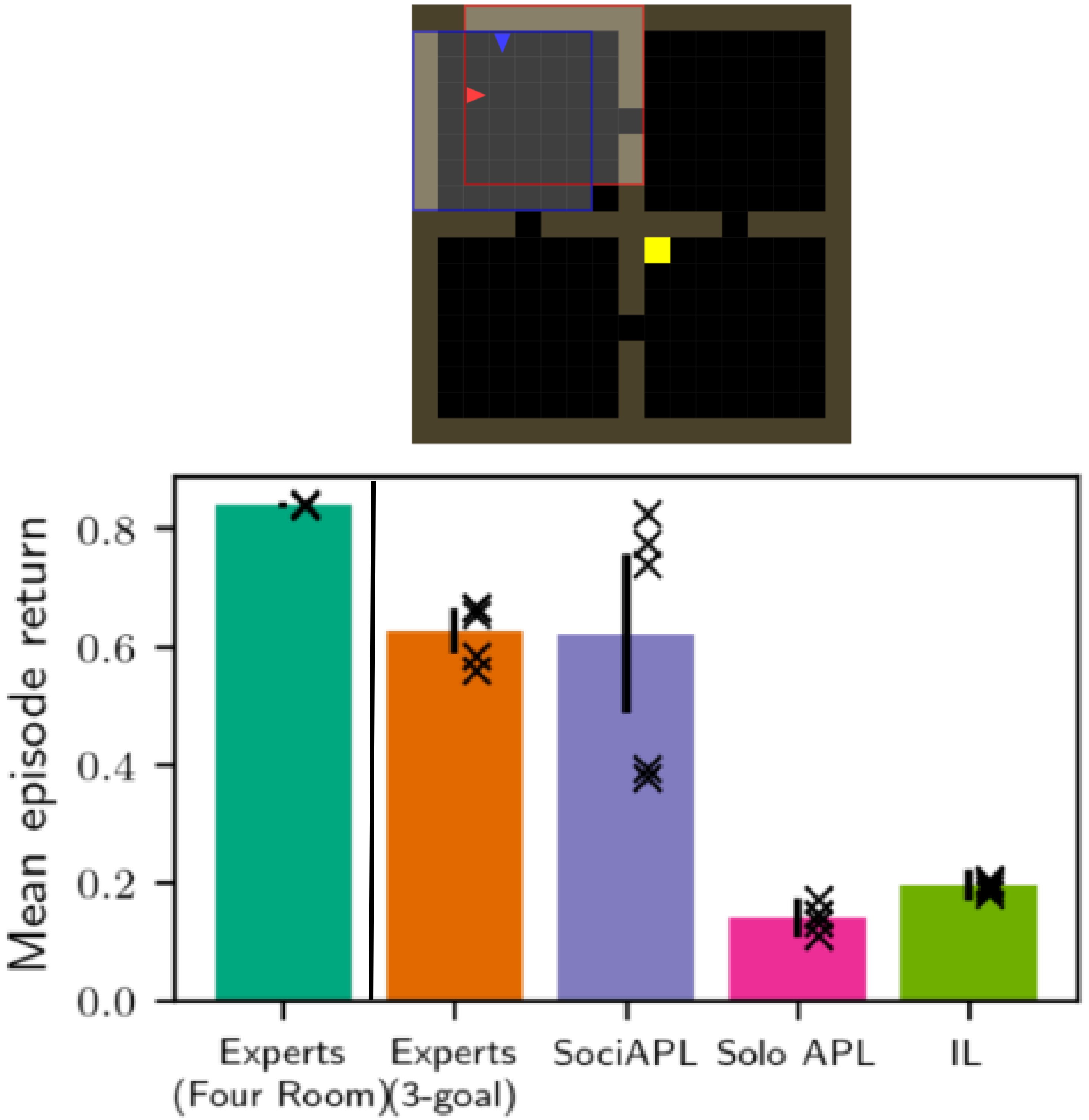
Learning Social Learning. Ndousse, Eck, Levine, Jaques (2020). Best Paper at the Neural Information Processing Systems (NeurIPS) workshop on Cooperative AI; ICML 2021.

Social Learners generalize to new envs

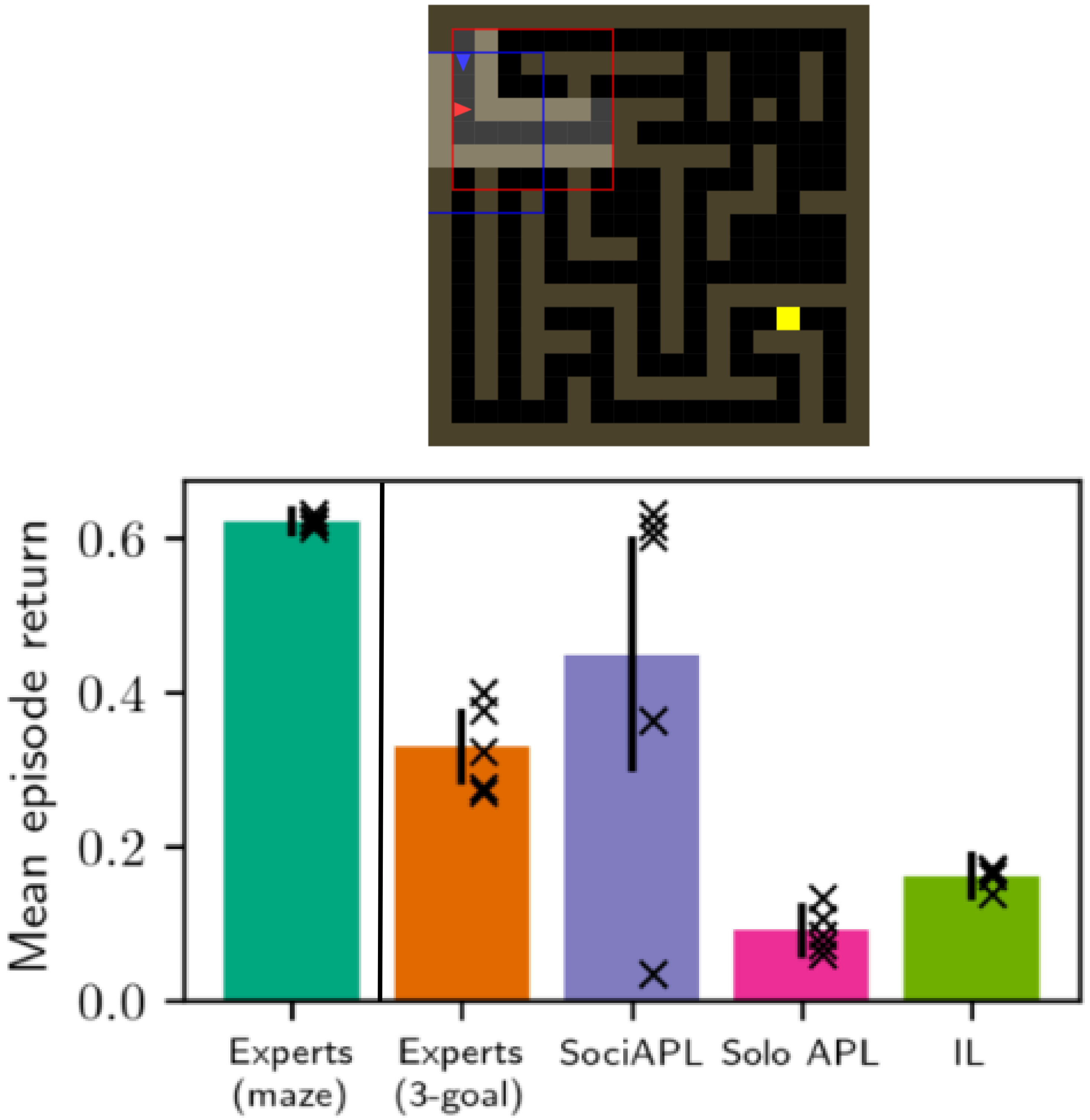
Social learners generalize better than RL experts or imitation learners



(a) Goal Cycle (4 goals, with experts)



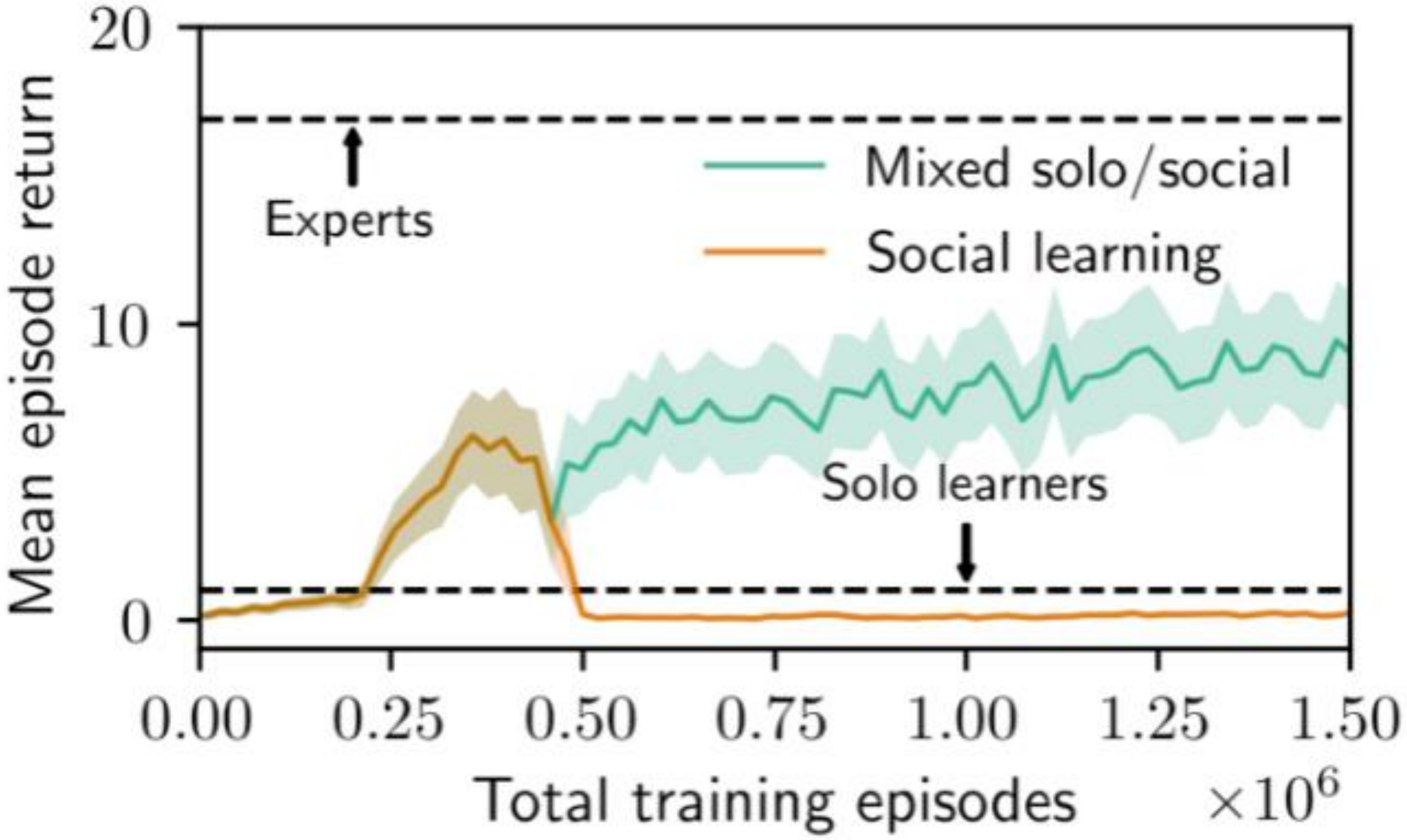
(b) Four Rooms (with expert)



(c) Maze (with expert)

Learning Social Learning. Ndousse, Eck, Levine, Jaques (2020). Best Paper at the Neural Information Processing Systems (NeurIPS) workshop on Cooperative AI; ICML 2021.

Social Learning can benefit performance when alone



(b) Transfer to solo 3-Goal

- Social learners discover skills that enable them to perform better when **alone**
  - If trained with a mix of solo & social episodes
- Outperform agents that were always trained alone

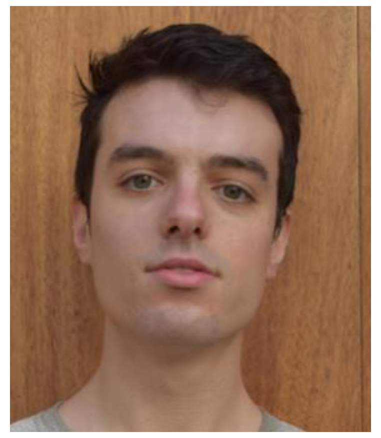
Learning Social Learning. Ndousse, Eck, Levine, Jaques (2020). Best Paper at the Neural Information Processing Systems (NeurIPS) workshop on Cooperative AI; ICML 2021.

Learning social learning helps with

Rapidly adapting
to new
environments

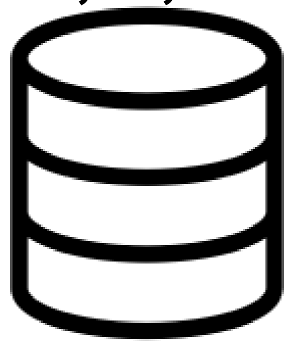
Learning complex behavior

Improving Social Learning with PsiPhi-Learning



Angelos Filos

No-reward demos
<s,a,s'>



$$\mathbb{E}^{\pi}[\sum_{k=0}^{\infty} \gamma^k r_{t+k} | s_t, a_t] = Q^{\pi}(s_t, a_t) \ = \Psi^{\pi}(s_t, a_t)^T \mathbf{w}$$

$$\Psi^{\pi}(s_t, a_t) = \mathbb{E}^{\pi}[\sum_{k=0}^{\infty} \gamma^k \Phi(s_{t+k}, a_{t+k})]$$

Future accumulated state features

Preference vector

$$\begin{bmatrix} 15 \\ 2 \end{bmatrix}$$

$$\begin{bmatrix} 1 & -1 \end{bmatrix} = \begin{bmatrix} 15(1) + 2(-1) \end{bmatrix} = 13$$

Successor Features[1,2]

[1]Peter Dayan. Improving generalization for temporal difference learning: The successor representation. Neural Computation, 5(4):613–624, 1993.
[2]Barreto, A., Dabney, W., Munos, R., Hunt, J. J., Schaul, T., Van Hasselt, H., & Silver, D. (2016). Successor features for transfer in reinforcement learning. arXiv preprint arXiv:1606.05312.

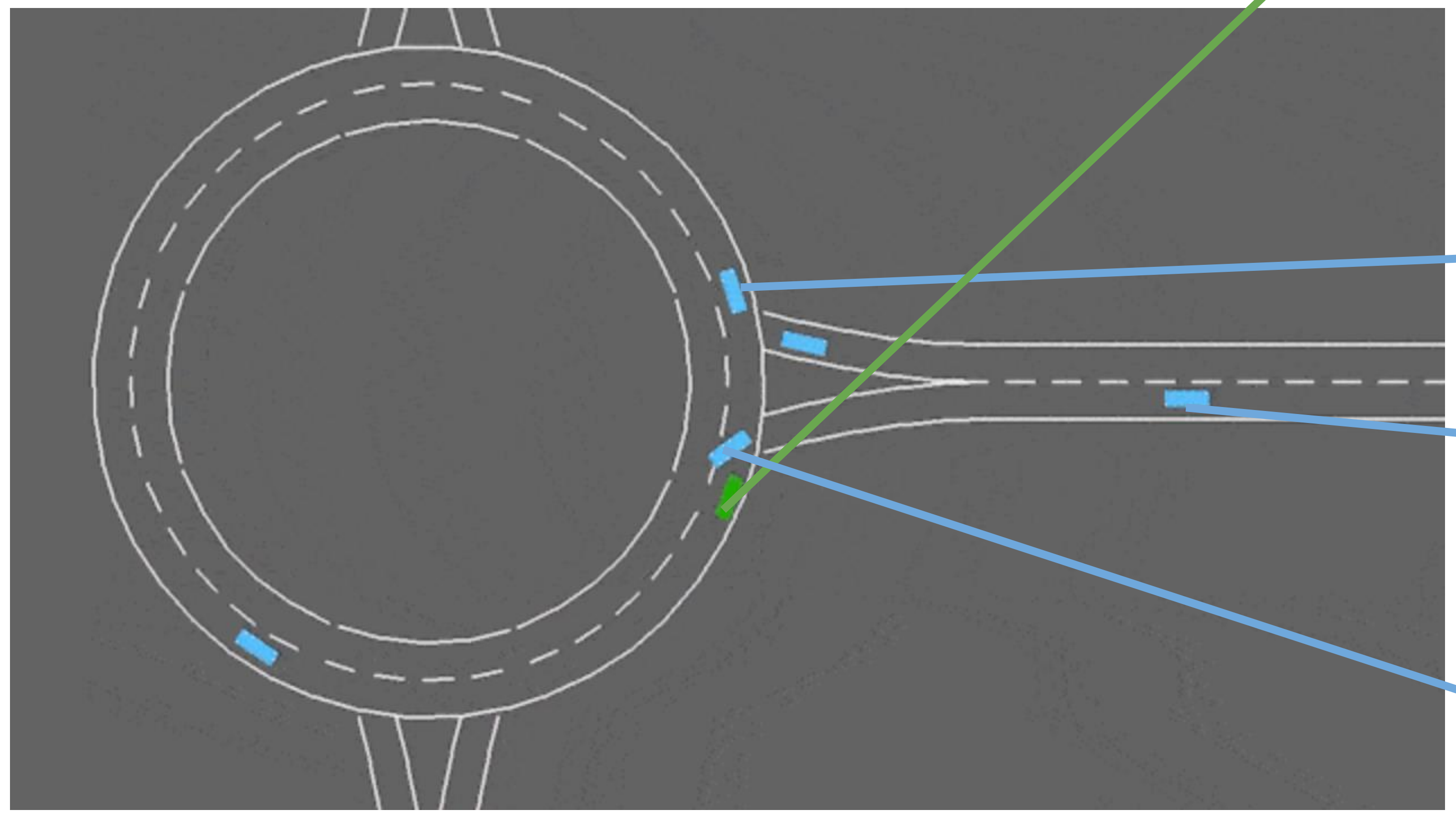
PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). *Equal Contribution. International Conference on Machine Learning (ICML) *(submitted)*.

Combine other agents' policies with ego-agent's preferences

- Use Generalized Policy Improvement (GPI)[2]

Ego-agent

$$Q^{ego}(s,a) = \Psi^{ego}(s,a)^T \mathbf{w^{ego}}$$



Multi-agent environment
(autonomous driving)

Other agents

$$Q^1(s,a) = \Psi^1(s,a)^T \cancel{\mathbf{w^1}}$$

$$Q^2(s,a) = \Psi^2(s,a)^T \cancel{\mathbf{w^2}}$$

...

$$Q^k(s,a) = \Psi^k(s,a)^T \cancel{\mathbf{w^k}}$$

$$\mathbf{w^{ego}}$$

$$\pi^{ego}(s) = \arg\max_a \max_k \Psi^k(s,a)^T \mathbf{w}^{ego}$$

Problem: why should other agents' $\mathbf{\Psi}$ be compatible with my **w?**

[2]Barreto, A., Dabney, W., Munos, R., Hunt, J. J., Schaul, T., Van Hasselt, H., & Silver, D. (2016). Successor features for transfer in reinforcement learning. arXiv preprint arXiv:1606.05312.

PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). *Equal Contribution. International Conference on Machine Learning (ICML) *(submitted)*.

ΨΦ-Learning

$$\begin{array}{c c} 🍌 & \\ 🍎 & \end{array}\begin{bmatrix} 15 \\ 2 \end{bmatrix} \qquad \Psi^{\pi}(s_t, a_t) = \mathbb{E}^{\pi}\Big[\sum_{k=0}^{\infty} \gamma^k \Phi(s_{t+k}, a_{t+k})\Big] \qquad \begin{array}{c c} 🍌 & \\ 🍎 & \end{array}\begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Policy-specific **Agent-agnostic**

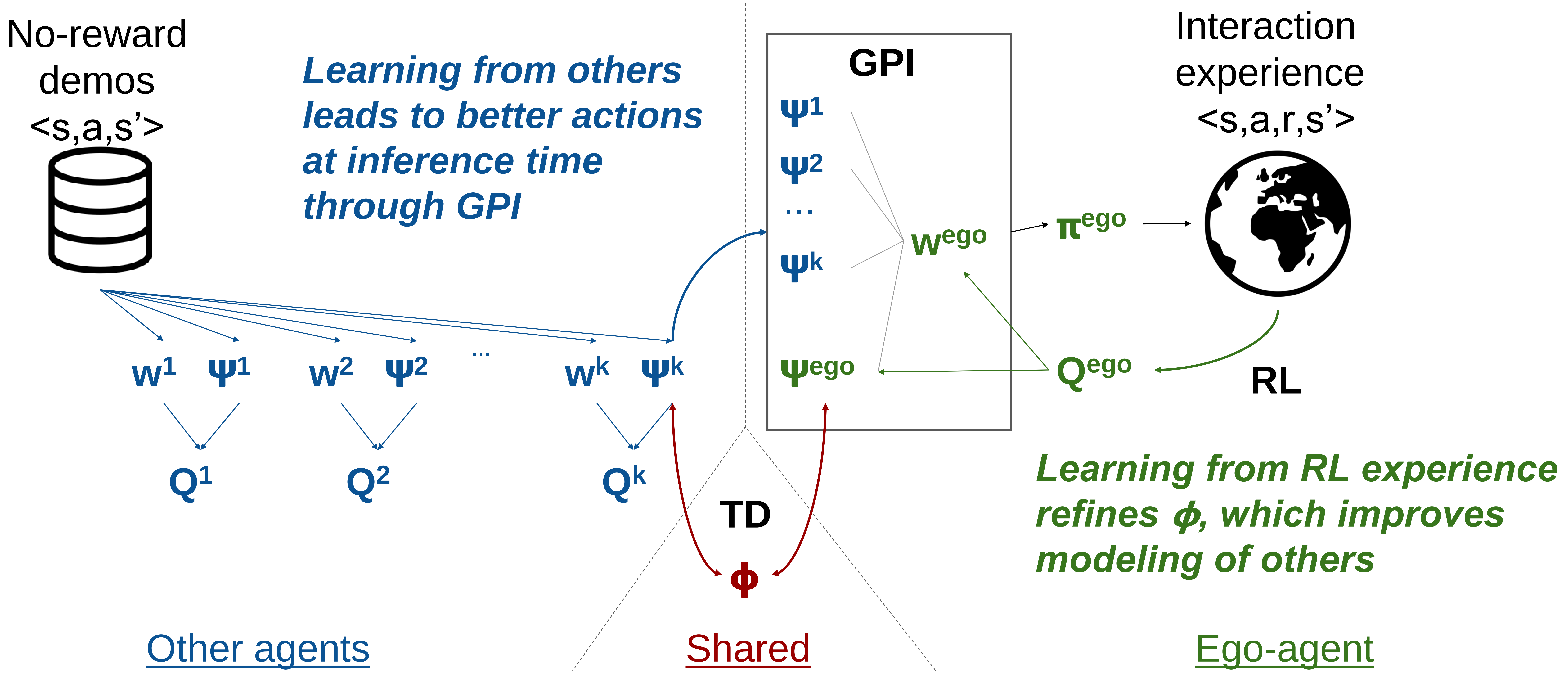
Solution: learn a shared feature representation **Φ** that explains all agents' behavior

- Each agent's $\mathbf{\Psi^k}$ should be built from the same **Φ**
- Learn **Φ** with **Inverse Temporal Difference** loss:

$$L_{ITD} = ||\Phi(s_t, a_t) + \gamma \Psi(s_{t+1}, a_{t+1}) - \Psi(s_t, a_t)||$$

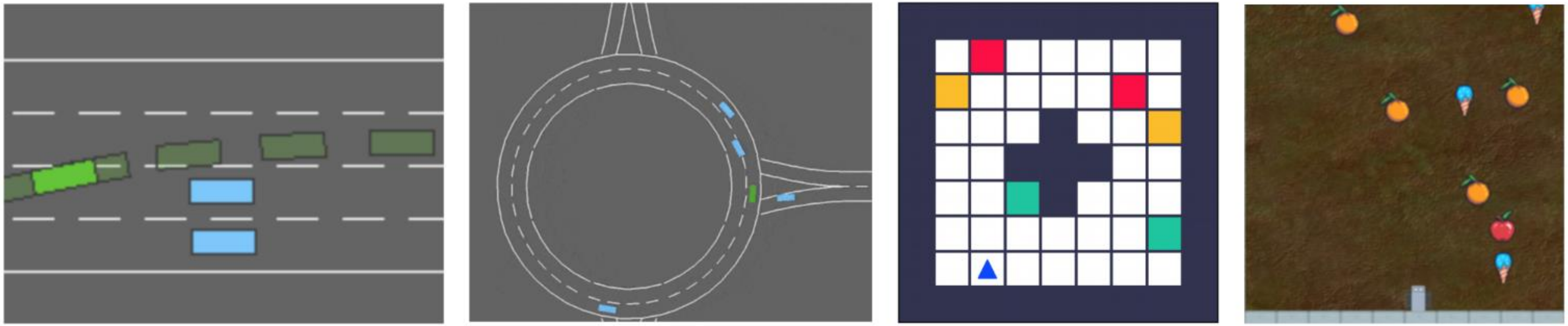
- Given **Φ**, learn each agents' $\mathbf{\Psi^k}$ with supervised behavior cloning loss
- Update **Φ** with rewards observed by ego-agent: $r^{ego}(s, a) = \Phi(s, a)^T \mathbf{w}^{\mathbf{ego}}$

ΨΦ-Learning



PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). *Equal Contribution. International Conference on Machine Learning (ICML) *(submitted)*.

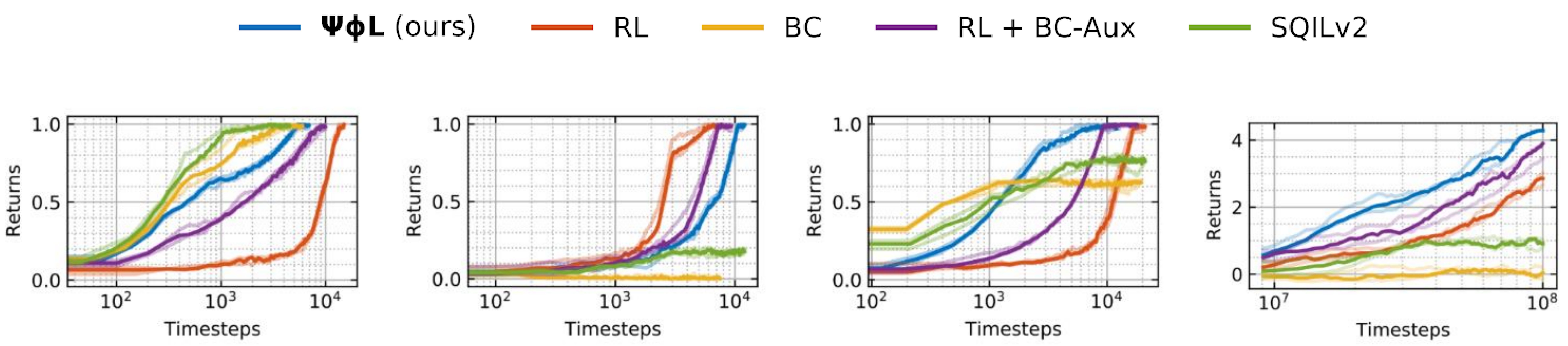
ΨΦ-Learning experiments



(a) Highway (b) Roundabout (c) CoinGrid (d) Fruit Bot

PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). *Equal Contribution. International Conference on Machine Learning (ICML) *(submitted)*.

ΨΦ-Learning improves on RL tasks



(a) Highway: Single-task

As good as IL when all other agents relevant

(b) Highway: Adversarial

As good as RL when all other agents irrelevant

(c) Highway: Multi-task

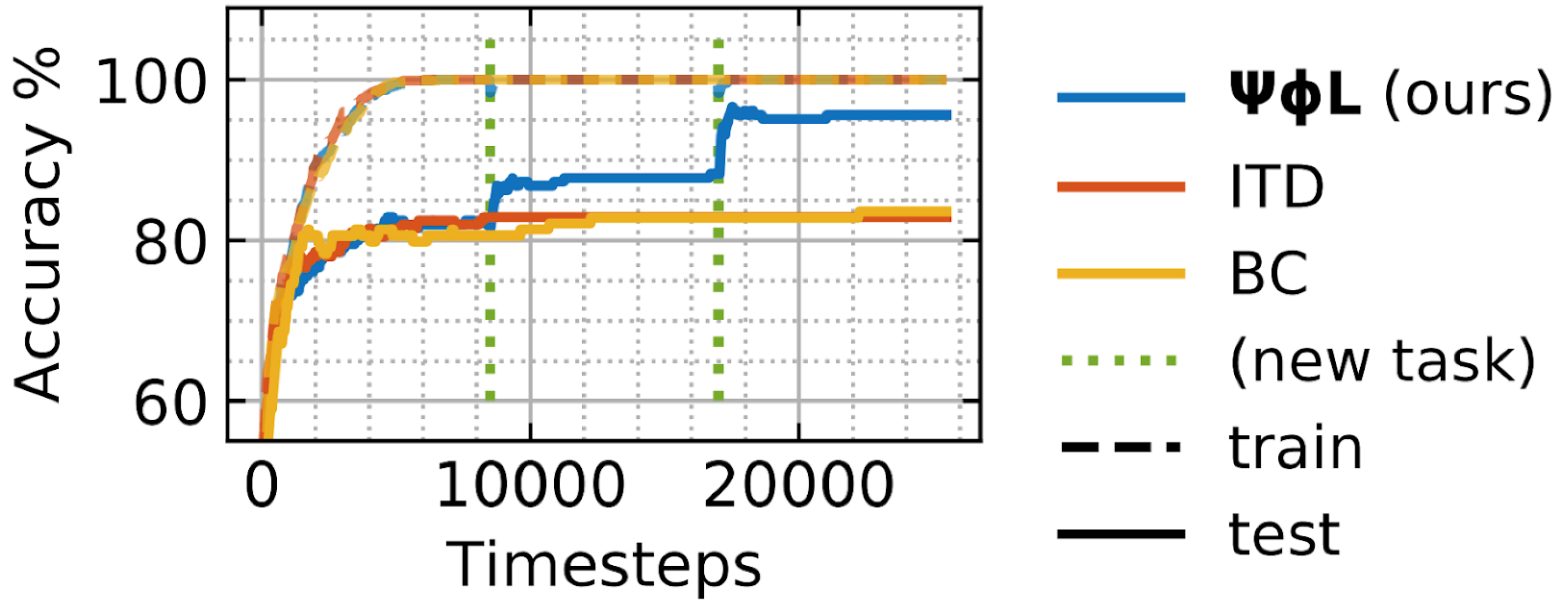
(d) FruitBot

Better than both in multi-task setting

PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). *Equal Contribution. International Conference on Machine Learning (ICML) *(submitted)*.

ΨΦ-Learning improves modeling of other agents

Predicting actions:



Inverse RL:

(predict rewards, train on them)

Methods	Roundabout^{DQN}	CoinGrid^{DQN}	FruitBot^{PPO}
BC$^{\dagger\clubsuit}$ (Pomerleau, 1989)	0.81±0.02	0.69±0.06	**0.37**±0.02
SQIL$^{\dagger\clubsuit}$ (Reddy et al., 2019)	0.85±0.02	0.64±0.05	**0.35**±0.03
GAIL$^{\dagger\diamondsuit}$ (Ho & Ermon, 2016)	0.77±0.07	0.73±0.02	0.31±0.02
ITD$^{\diamondsuit}$ (ours, cf. Section 3.1)	**0.92**±0.01	**0.77**±0.03	**0.35**±0.04

PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). *Equal Contribution. International Conference on Machine Learning (ICML) *(submitted)*.

ΨΦ-Learning improves generalization

	0-shot				1-shot				100-shot			
Methods	R+G	R−G	−R+G	−R−G	R+G	R−G	−R+G	−R−G	R+G	R−G	−R+G	−R−G
SQILv2♣ (Reddy et al., 2019)	1.0±0.0	0.0±0.0	0.0±0.0	−1.0±0.0	**1.0**±0.0	0.0±0.0	0.0±0.0	−1.0±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0
ΨΦ-learning ◇ (ours, cf. Section 3.2)	1.0±0.0	**0.2**±0.1	**0.2**±0.1	−**0.4**±0.2	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0

Few-shot transfer in CoinGrid

PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). *Equal Contribution. International Conference on Machine Learning (ICML) *(submitted)*.

Conclusion

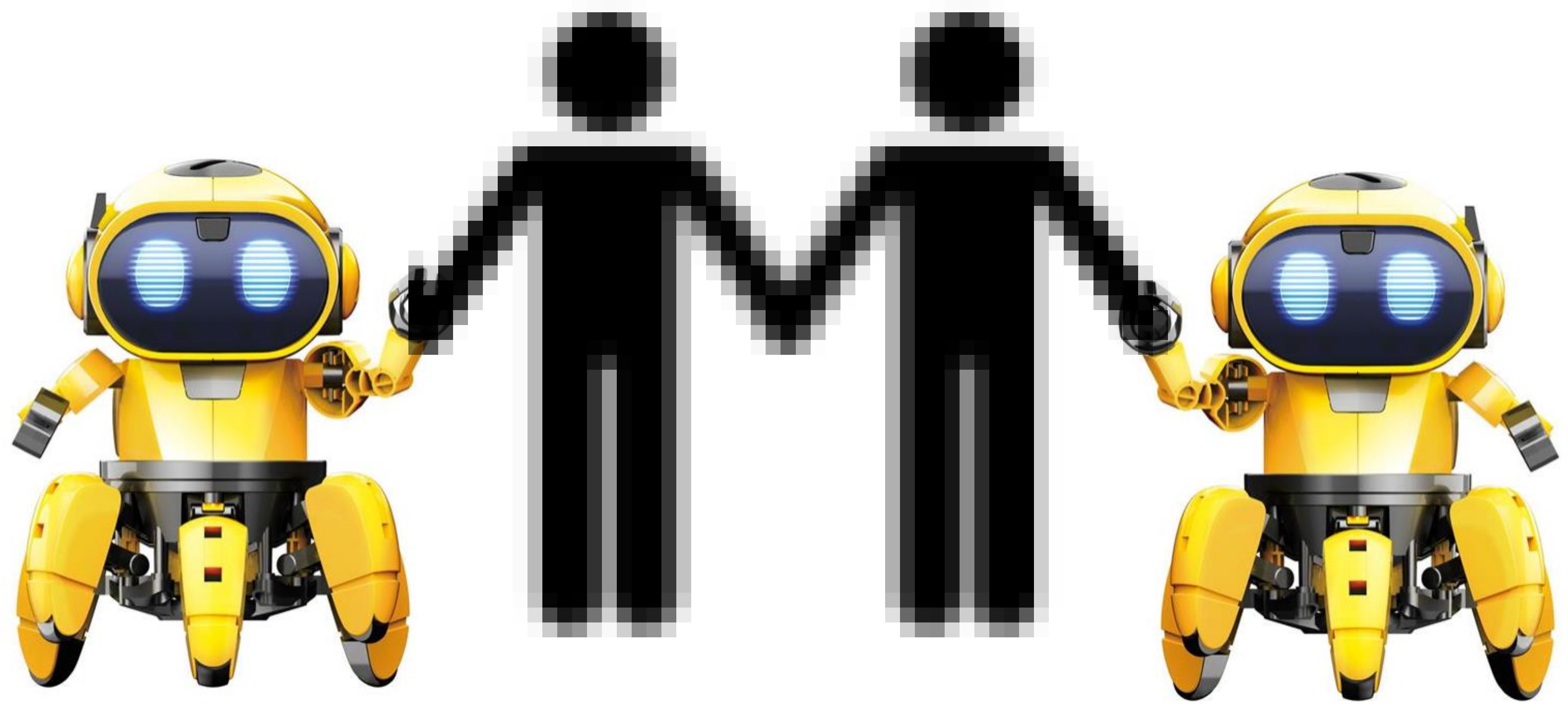
Social learning is a powerful mechanism that can help RL agents:

Rapidly adapt to new environments

Learn complex behavior

Social Reinforcement Learning improves...

- **Coordination** with other agents
- **Human-AI interaction**
- **Learning** complex behavior
- **Generalization** to new environments

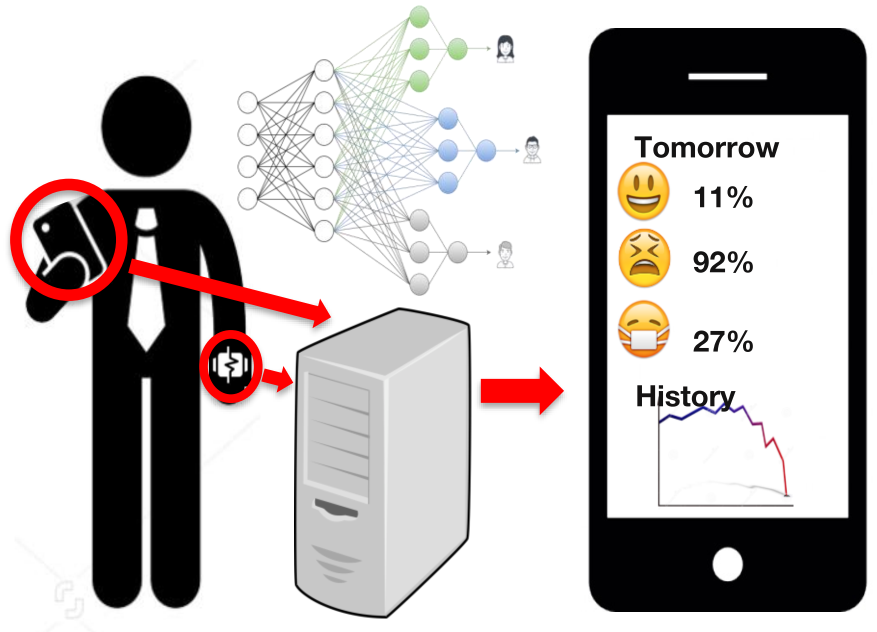


Extra slides

Questions?

Human-AI ⟷ Multi-agent

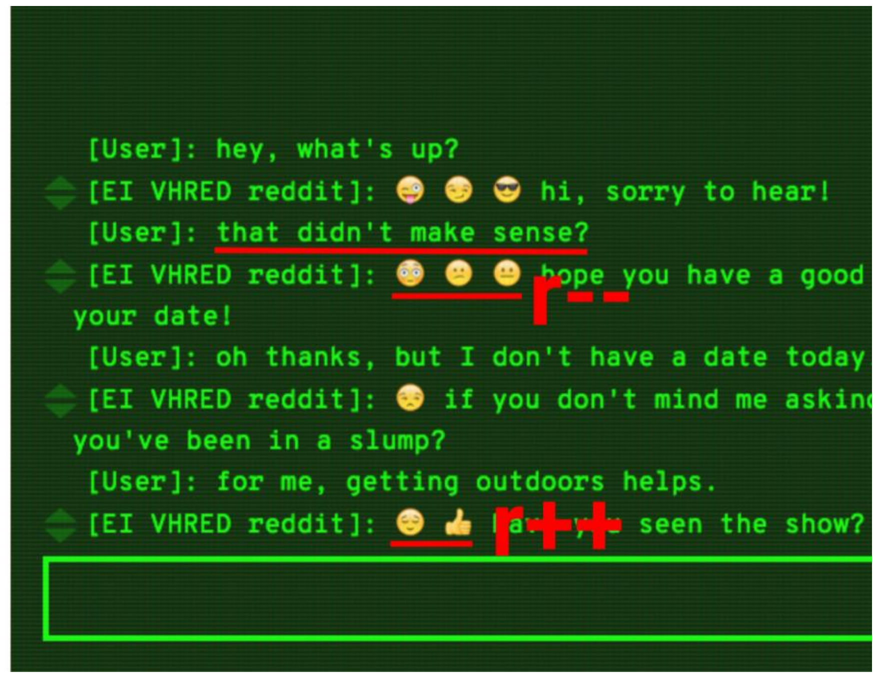
Detecting social & affective cues



[Personalized Multitask Learning.](#) Taylor*, Jaques*, et al. **IEEE Transactions on Affective Computing (TAC) 2020 Best Paper**.

[Predicting bonding from facial expressions and body language](#). Jaques et al. **IVA 2016**

Learning from human social cues



[Human-centric Dialog Training via Offline RL](#). Jaques*, Shen*, et al. **EMNLP 2021.**

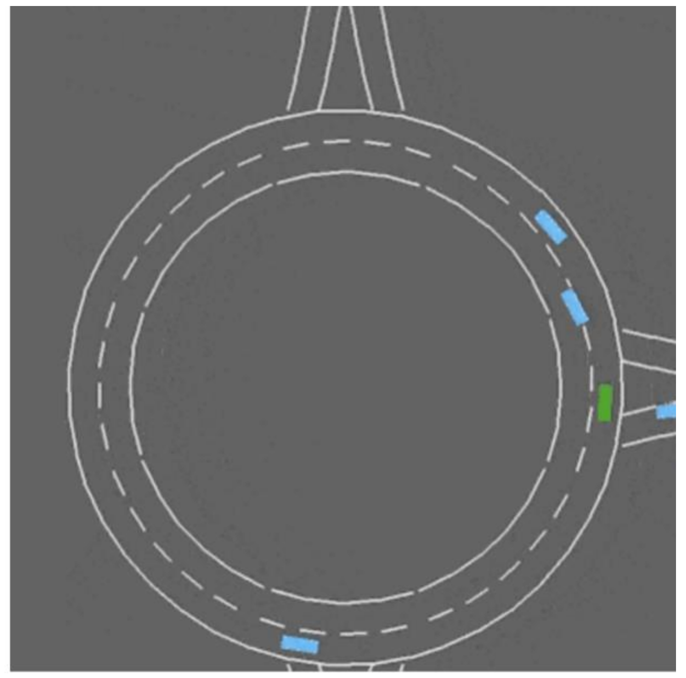
[Human Evaluation of Dialog Systems.](#) Ghandeharioun*, Shen*, Jaques*, et al. **NeurIPS 2020.**

[Hierarchical RL for Dialog](#). Saleh*, Jaques*, et al. **AAAI Oral (top 8%)**

Coordination

[Social influence](#). Jaques et al. **ICML 2019 Best Paper Honourable mention (top 0.26%)**.

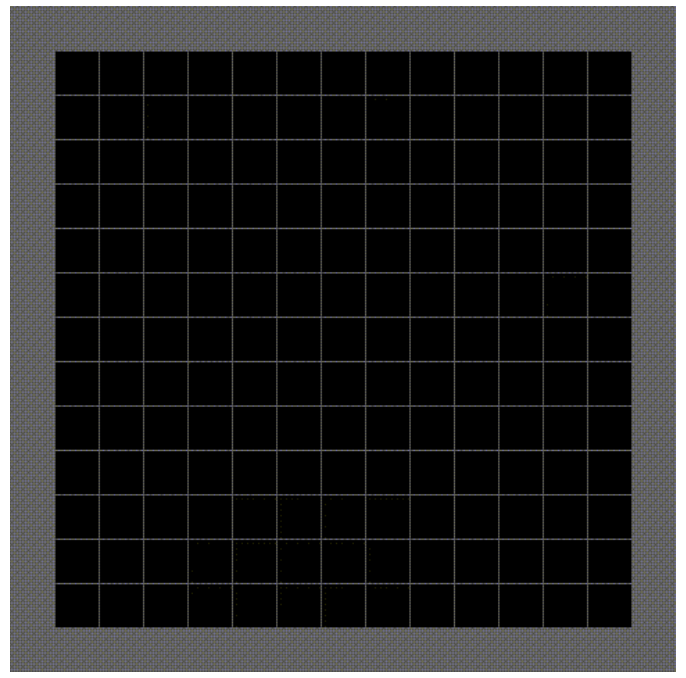
Multi-agent social learning



[Emergent Social Learning from MARL](#). Ndousse, ..., Jaques . **ICML 2021 & NeurIPS CoopAI workshop Best Paper 2020**

[PsiPhi-Learning](#). Filos, ... Jaques*, Farquhar*. **ICML 2021 oral (top 3%)**.

Emergent Complexity

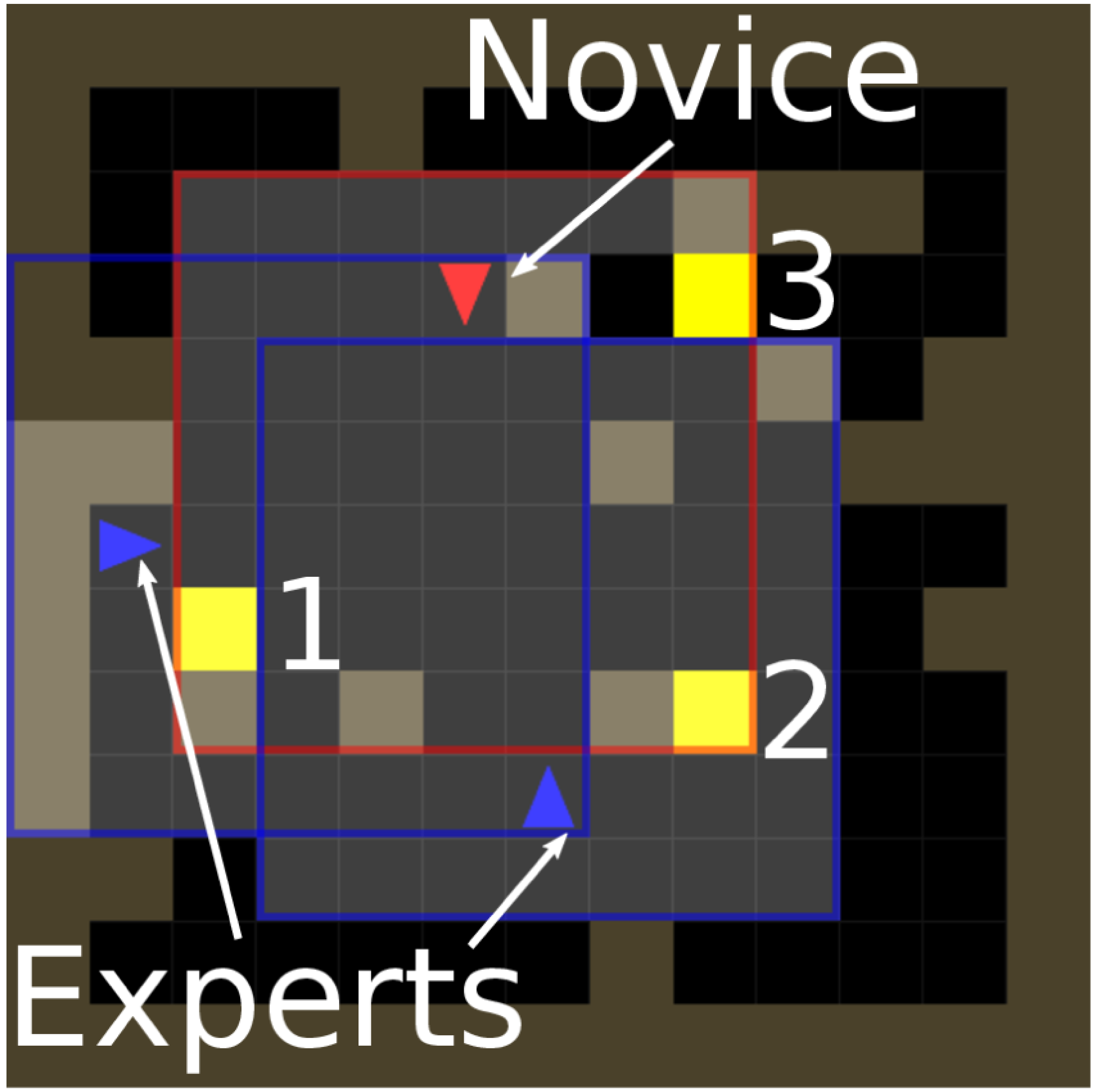


[PAIRED](#). Dennis*, Jaques*, et al. **NeurIPS 2020 oral (top 1%).**

[Environment Generation for Zero-Shot Compositional RL](#). Gur, Jaques, et al. **NeurIPS 2021**

Social learning environment

Kamal Ndousse



(a) Goal Cycle

- **Partial observability** and **no privileged access** to other agents states or actions

- Information (correct goal ordering) is **hidden**[1]. Incorrect **exploration expensive**.

- Agents change colour with the recent average of their rewards (**prestige cue**)

[1]Hiding information to evoke social learning proposed in Borsa, D., Piot, B., Munos, R., & Pietquin, O. (2017). Observational learning by reinforcement learning.

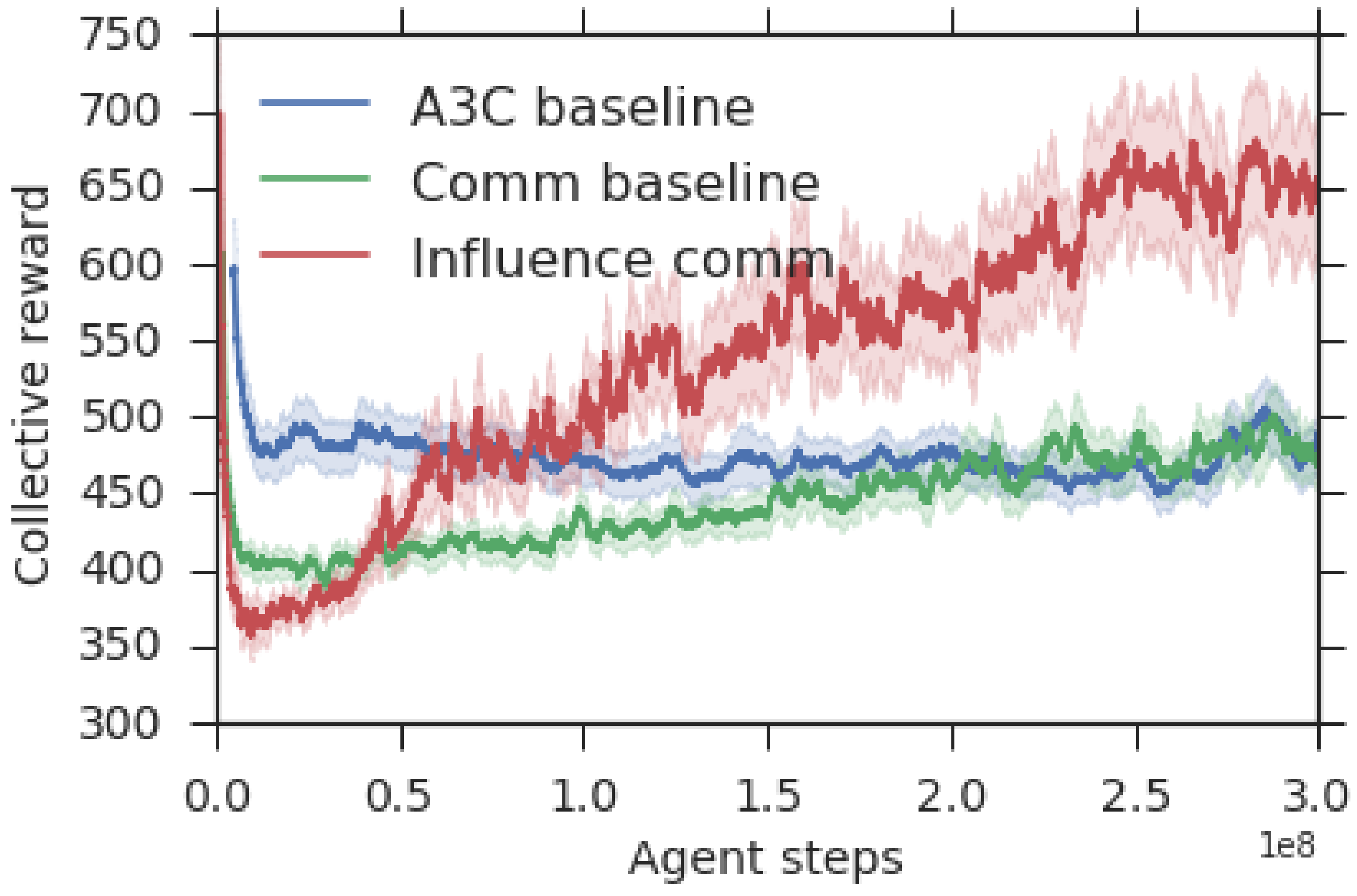
Learning Social Learning. Ndousse, Eck, Levine, Jaques (2020). Best Paper at the Neural Information Processing Systems (NeurIPS) workshop on Cooperative AI.

Coordinate with other agents… via social influence

Give agents an intrinsic social reward for having a **causal influence** on other agents' actions. Like **social empowerment.**

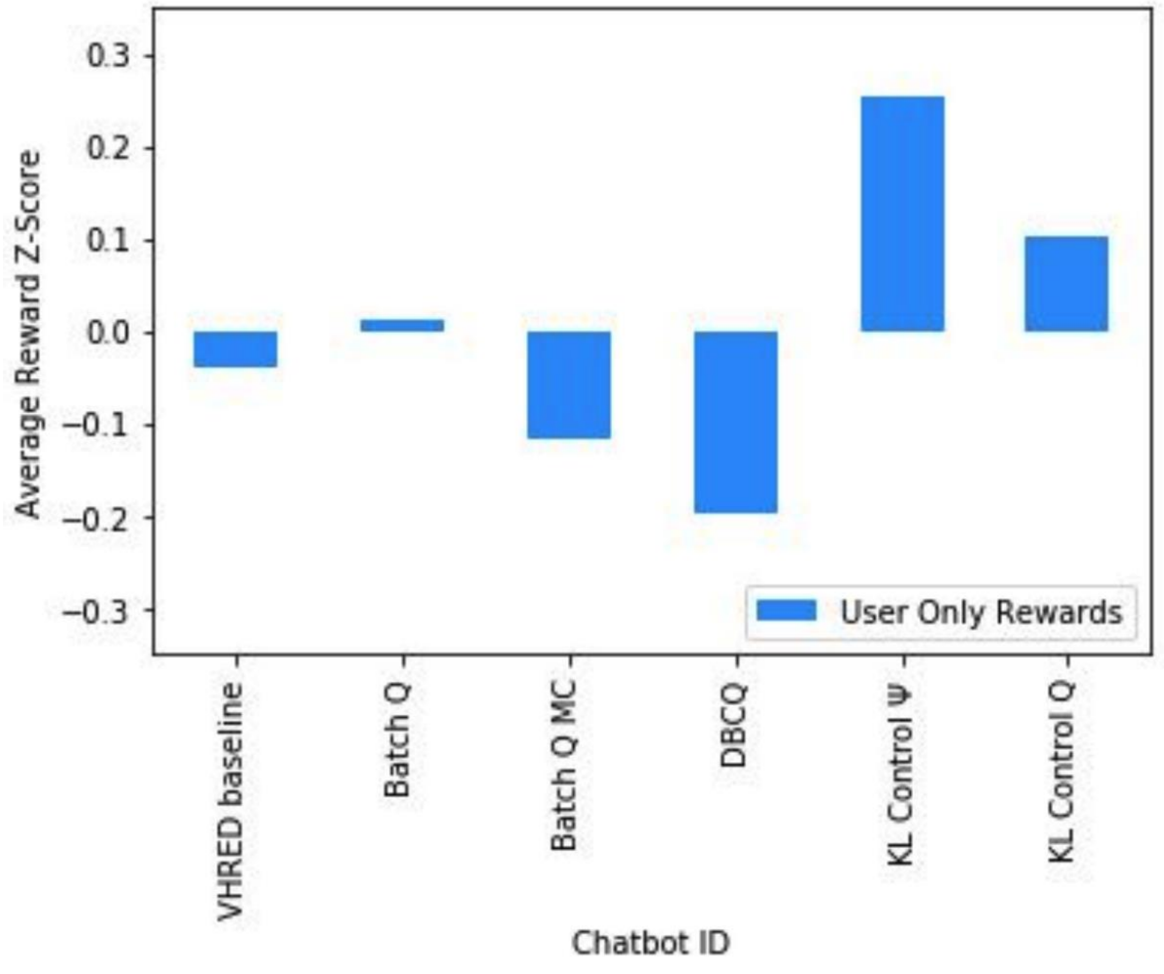
$$D_{KL}[p(a_{t+1}^B | s_t^A, a_t^A) || p(a_{t+1}^B | s_t^A)]$$

- Enhances **cooperation** in social dilemmas
- Leads to **emergent communication**
- Works without centralized control or privileged access to other agent's states/rewards



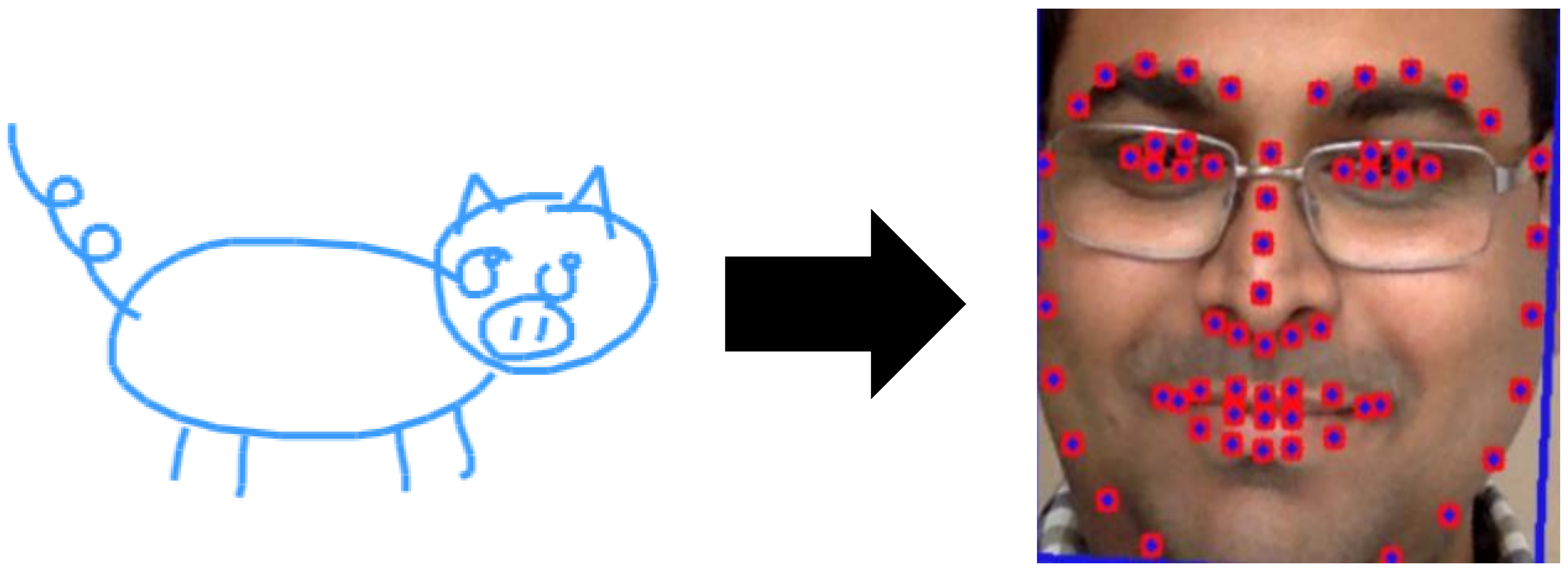
Social Influence as Intrinsic Motivation for Multi-Agent Deep Reinforcement Learning. Jaques, Lazaridou, Hughes, Gulcehre, Ortega, Strouse, Leibo, de Freitas (2019). International Conference on Machine Learning (ICML) *Best Paper Honourable Mention.*

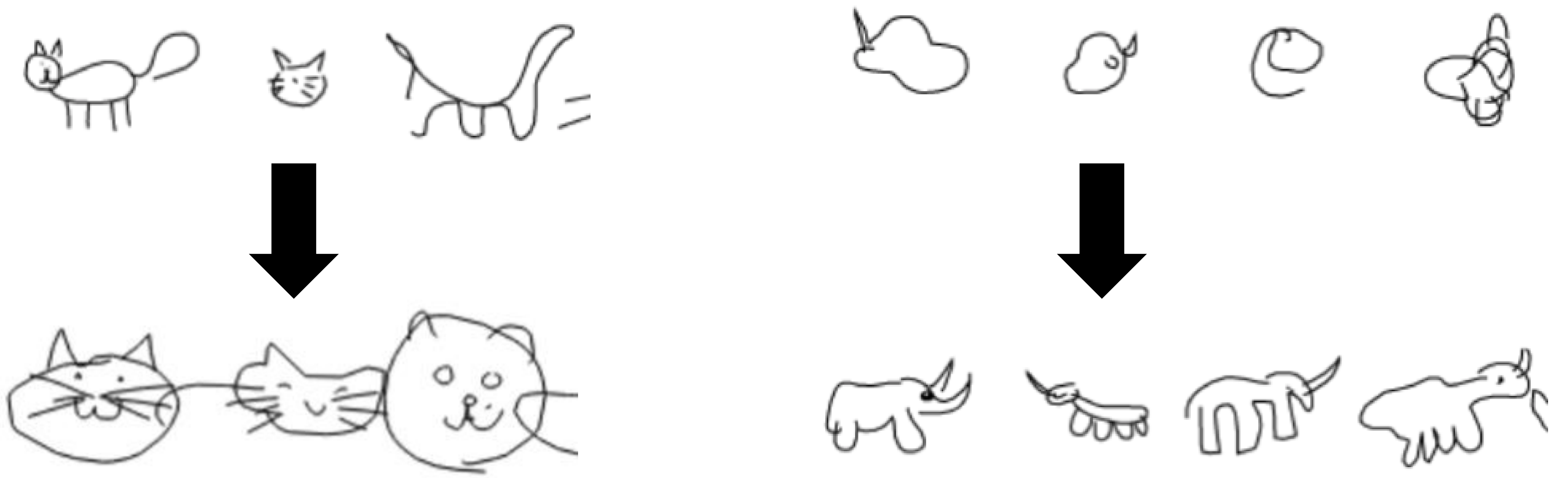
Improve human-AI interaction… in dialog



Human-centric Dialog Training via Offline Reinforcement Learning. Jaques*, Shen*, Ghandeharioun, Ferguson, Lapedriza, Jones, Gu, Picard (2020). Empirical Methods on Natural Language Processing (EMNLP).

Learning from human-AI interaction… with faces





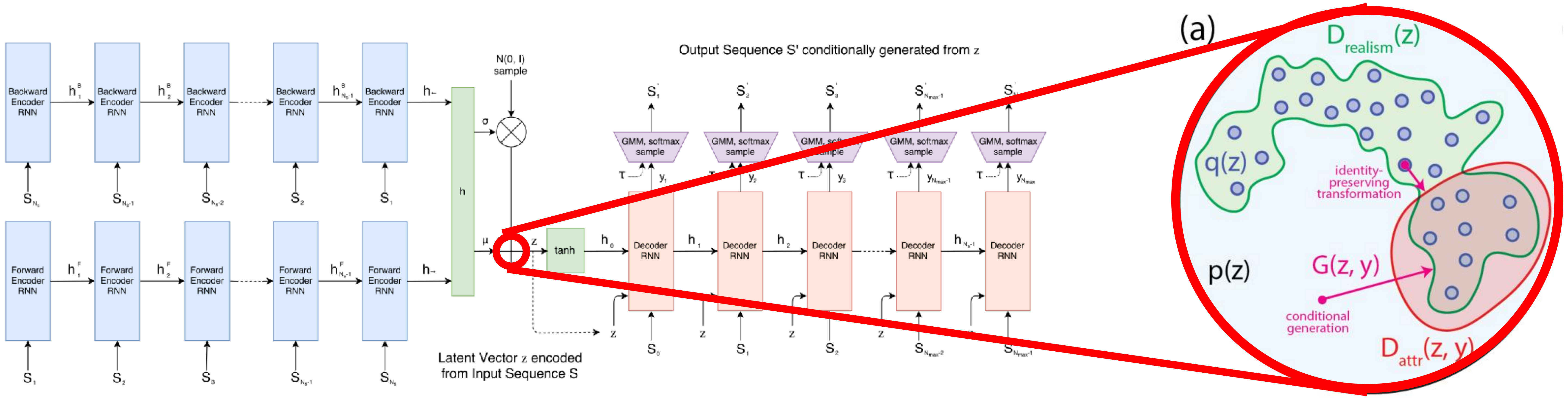
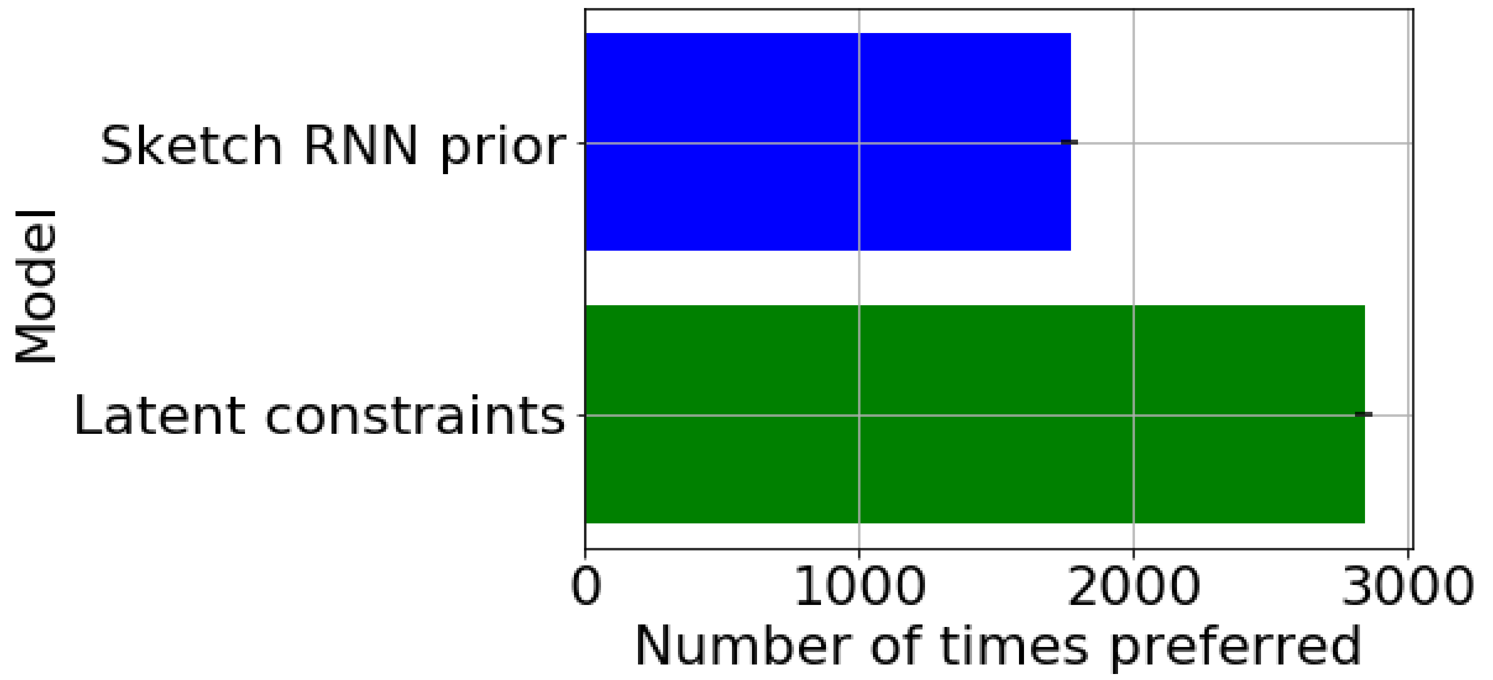


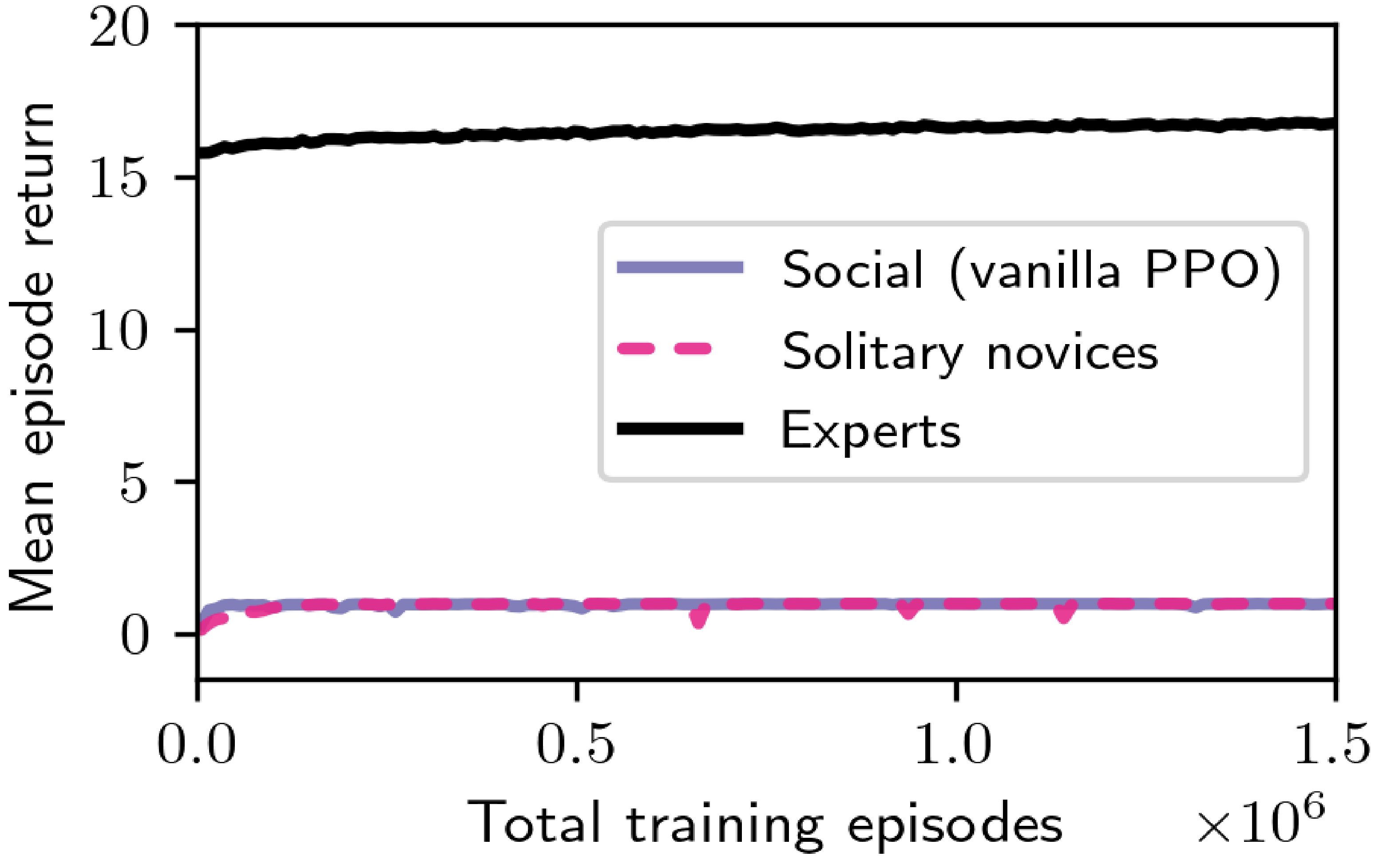
Figure 2: Schematic diagram of sketch-rnn.



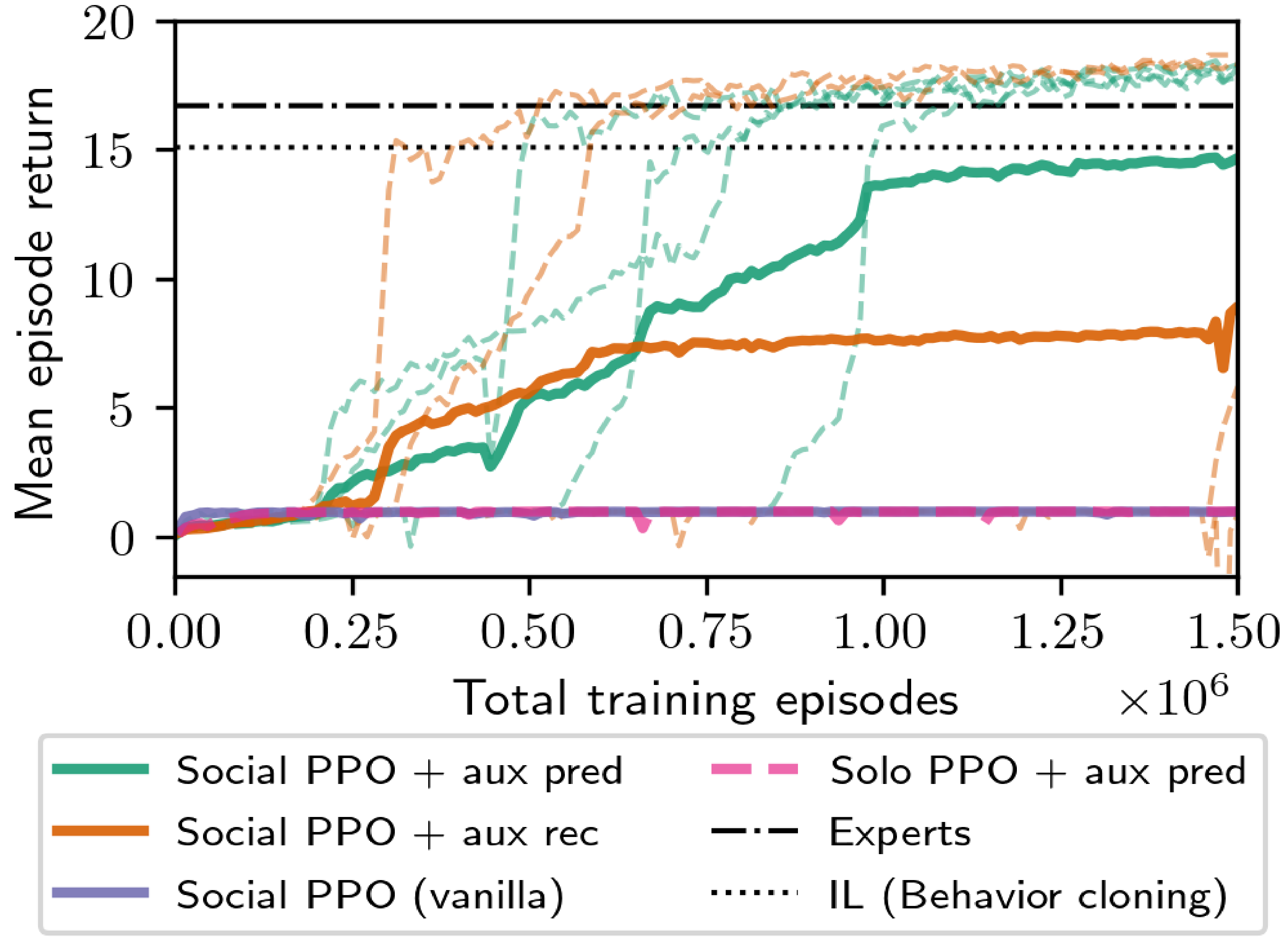
Learning via social awareness: Improving a deep generative sketching model with facial feedback. Jaques, McCleary, Engel, Ha, Bertsch, Picard, Eck (2018). International Conference on Representation Learning (ICLR) workshop.

Multi-agent social learning

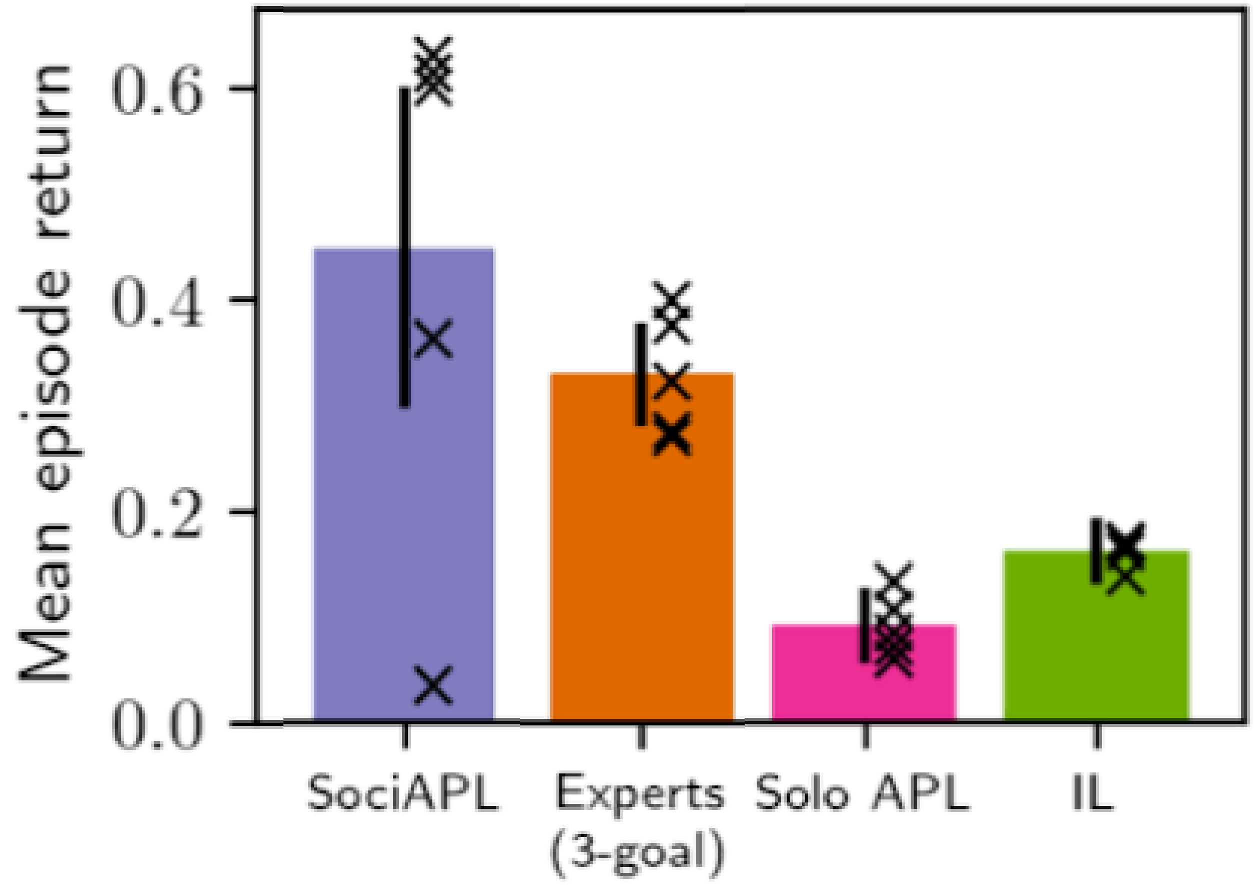
Model-free RL agents do not learn from experts in their environment



Fix with model-based auxiliary loss



Social learners generalize better than **RL experts** or **imitation learners**



Emergent Social Learning from Multi-Agent Reinforcement Learning. Ndousse, Eck, Levine, Jaques (2021). International Conference on Machine Learning (ICML). **Best Paper** at the NeurIPS 2020 Cooperative AI workshop.

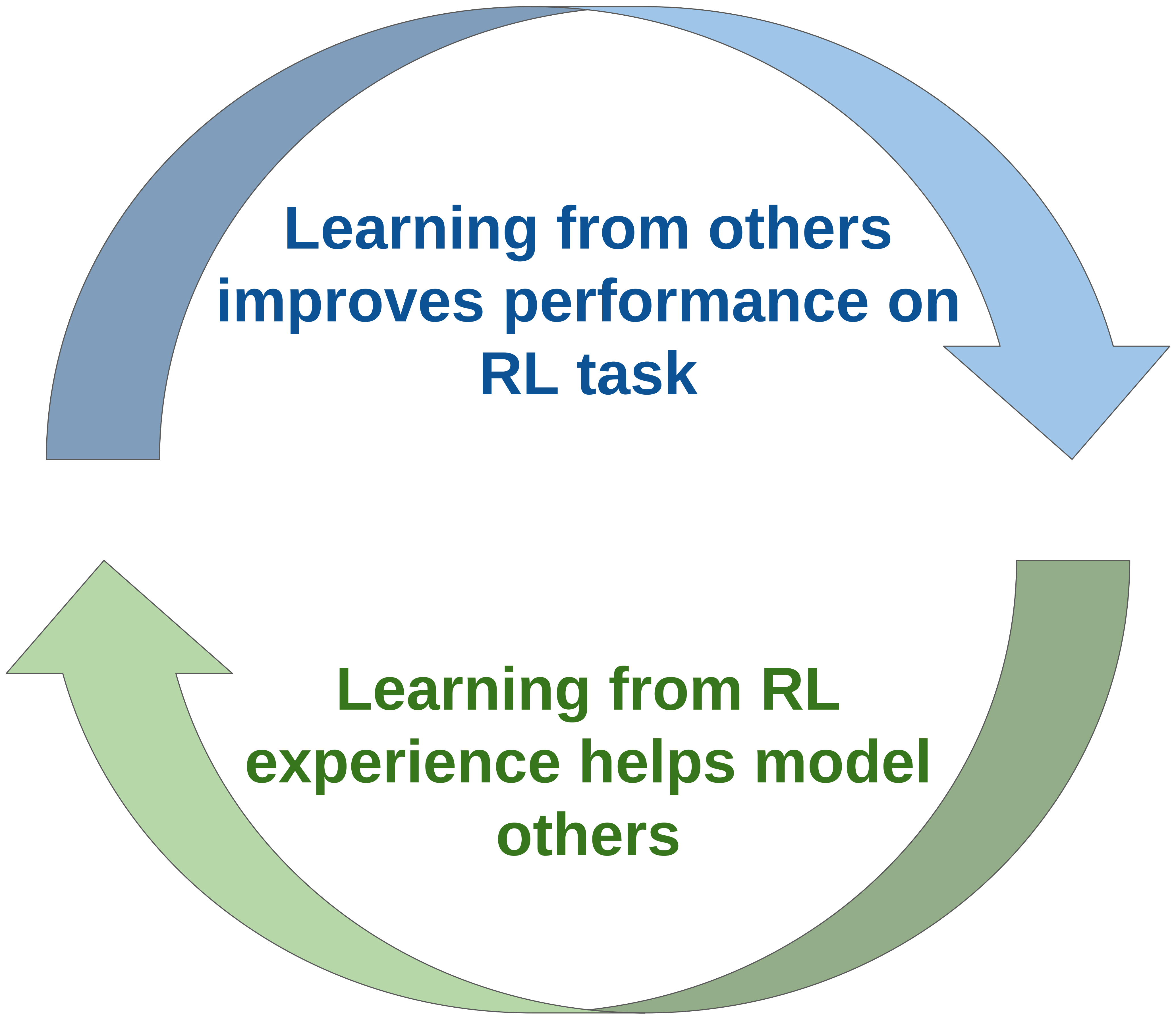
Improving Social Learning with PsiPhi-Learning

- Model both other agents and RL agent with **successor features**

$$Q^k(s,a) = \Psi^k(s,a)^T \mathbf{w^k}$$

- Learned a **shared basis for behavior**

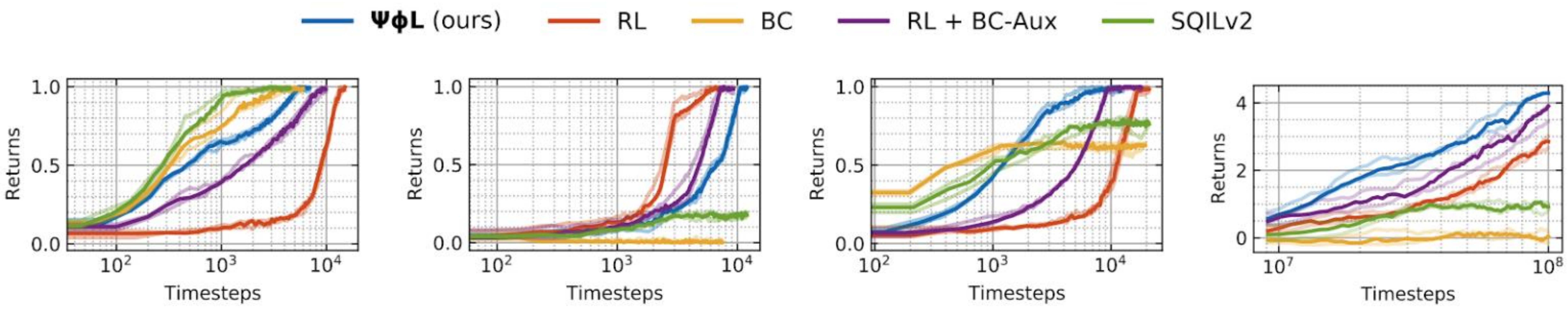
$$\Psi^k(a_t, s_t) = \mathbb{E}^\pi[\sum_{t'=t}^{T} \phi(a_t, s_t)]$$



[PsiPhi-Learning: Reinforcement Learning with Demonstrations using Successor Features and Inverse Temporal Difference Learning]. Filos, Lyle, Gal, Levine, Jaques*, Farquhar* (2021). International Conference on Machine Learning **(ICML) oral**. *Equal Contribution.

PsiPhi Improves:

RL with...

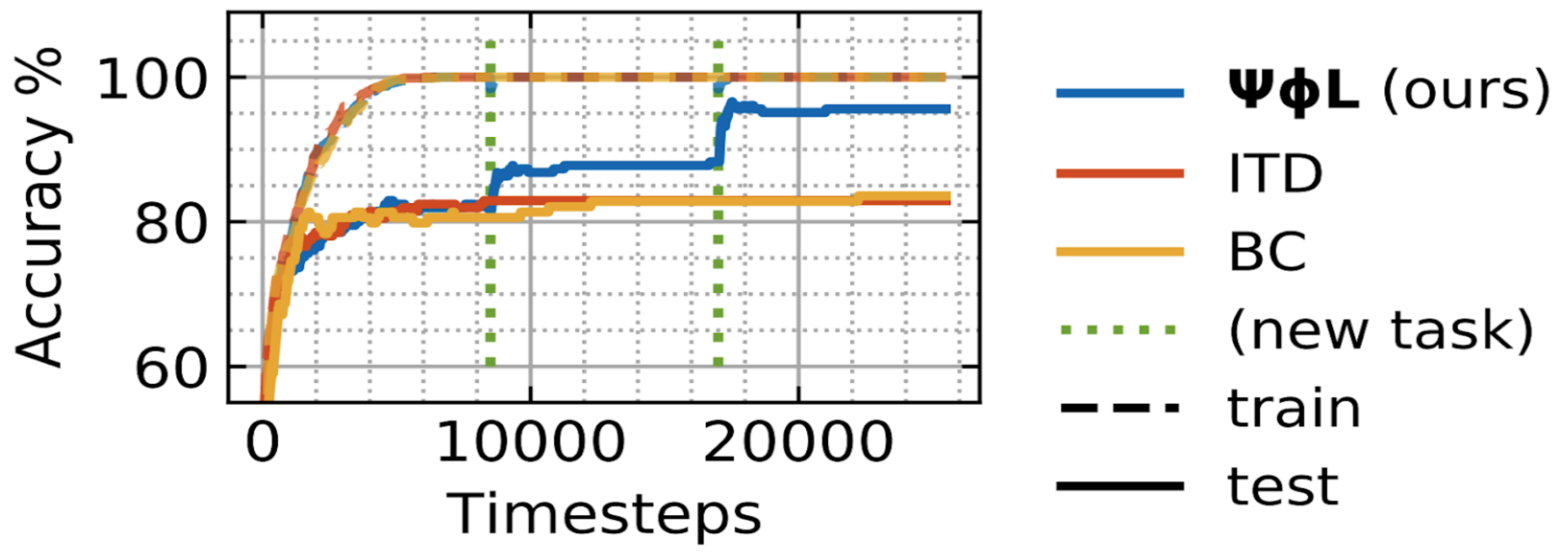


(a) Highway: Single-task (b) Highway: Adversarial (c) Highway: Multi-task (d) FruitBot

...perfect demos ...irrelevant demos ...multi-task demos

Methods	Roundabout^{DQN}	CoinGrid^{DQN}	FruitBot^{PPO}
BC$^{\dagger\clubsuit}$ (Pomerleau, 1989)	0.81±0.02	0.69±0.06	**0.37**±0.02
SQIL$^{\dagger\clubsuit}$ (Reddy et al., 2019)	0.85±0.02	0.64±0.05	**0.35**±0.03
GAIL$^{\dagger\diamond}$ (Ho & Ermon, 2016)	0.77±0.07	0.73±0.02	0.31±0.02
ITD$^{\diamond}$ (ours, cf. Section 3.1)	**0.92**±0.01	**0.77**±0.03	**0.35**±0.04

IRL



Imitation

1-shot transfer

	0-shot				1-shot				100-shot			
Methods	R+G	R−G	−R+G	−R−G	R+G	R−G	−R+G	−R−G	R+G	R−G	−R+G	−R−G
SQILv2$^{\clubsuit}$ (Reddy et al., 2019)	1.0±0.0	0.0±0.0	0.0±0.0	−1.0±0.0	**1.0**±0.0	0.0±0.0	0.0±0.0	−1.0±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0
ΨΦ-learning$^{\diamond}$ (ours, cf. Section 3.2)	1.0±0.0	**0.2**±0.1	**0.2**±0.1	**−0.4**±0.2	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0	**1.0**±0.0

Social learning

Learning from other intelligent agents in your environment

Interaction, feedback

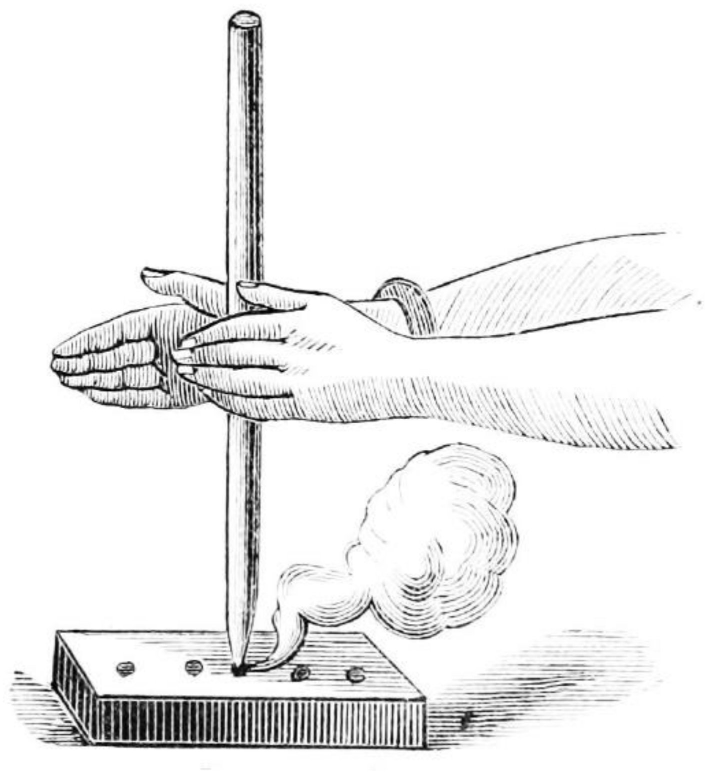
Cooperation

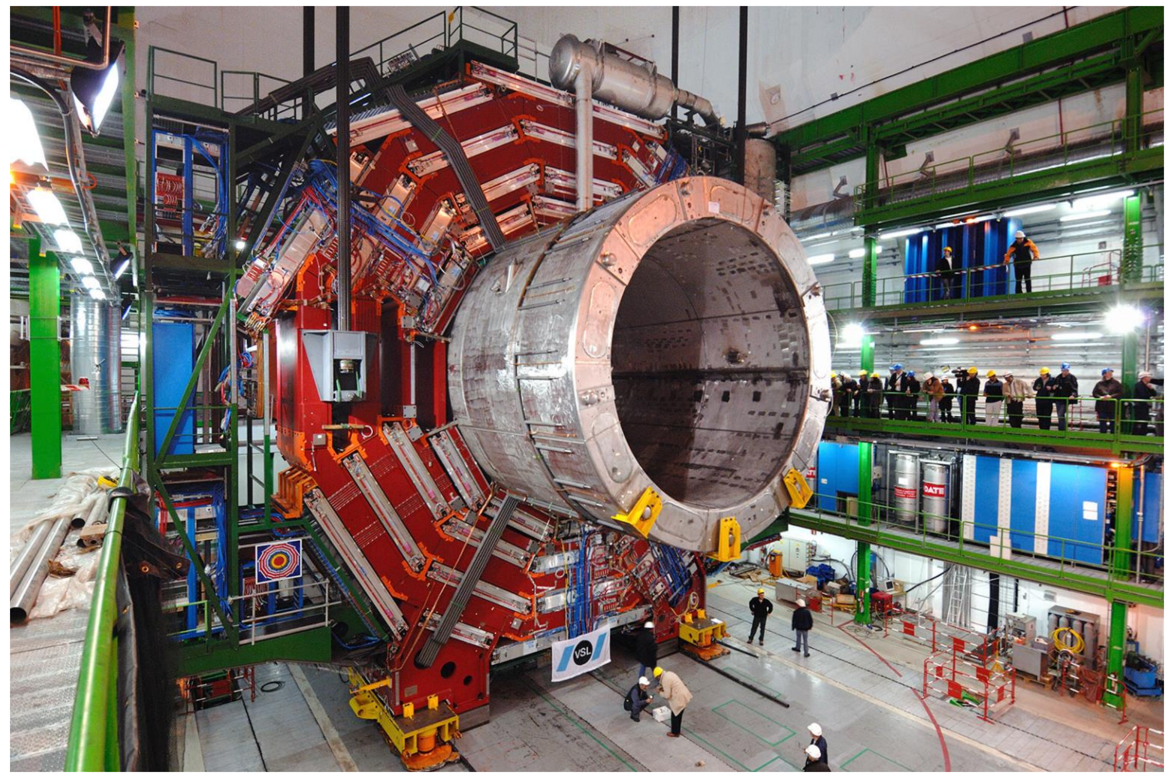
Competition

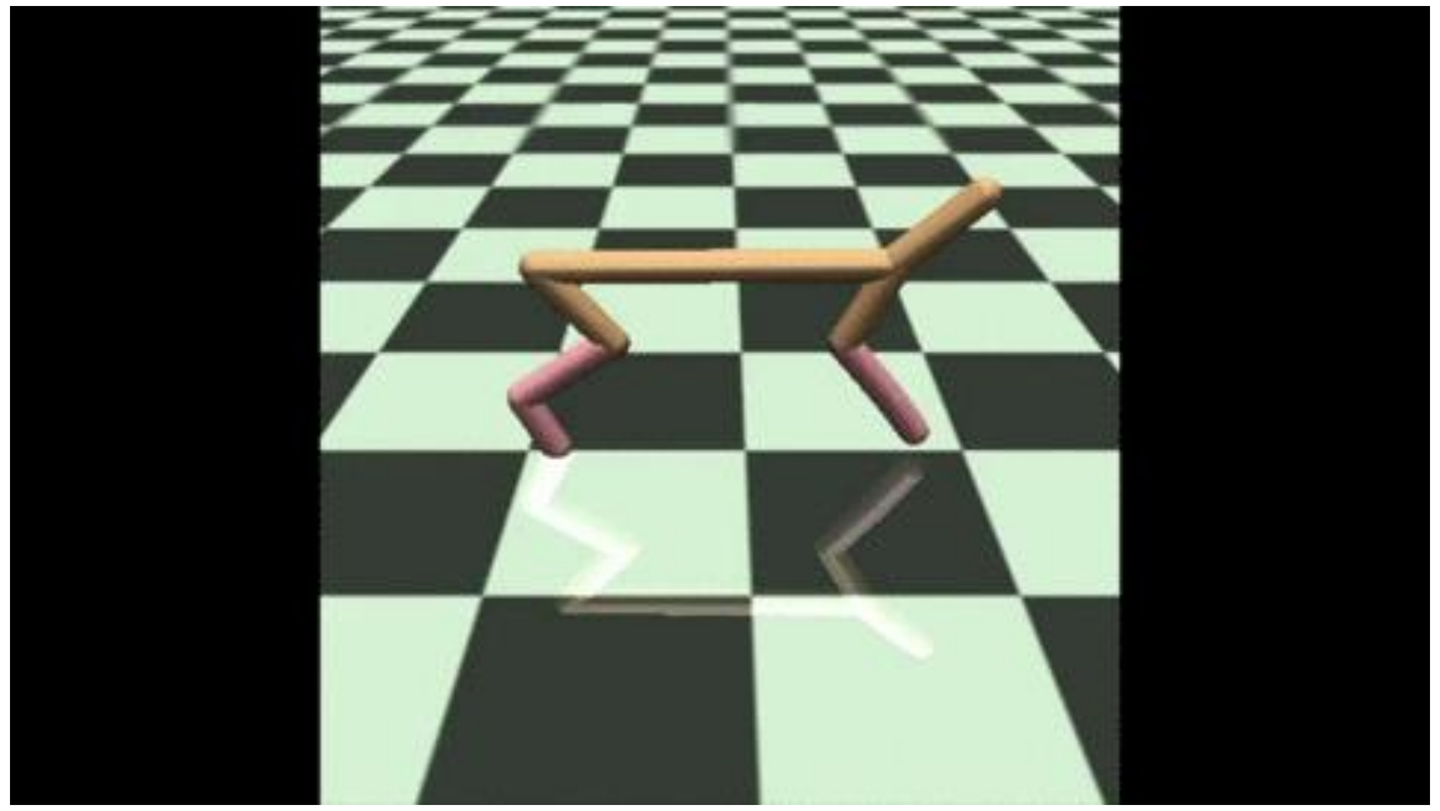
Co-existence

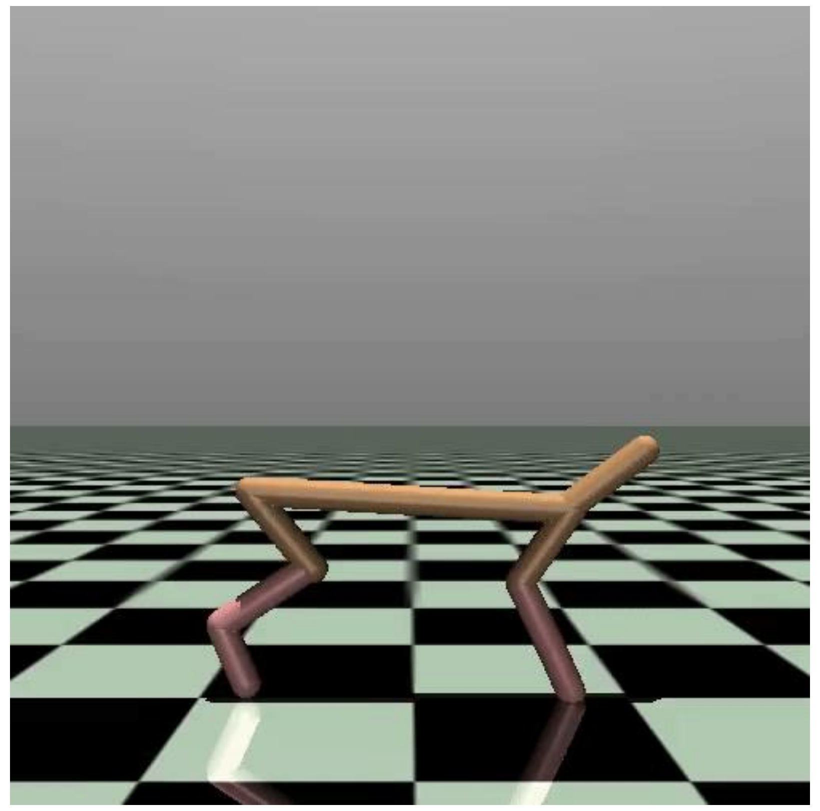
Benefits of social learning for RL

- Individual exploration is expensive, time consuming









- Social learning lets you “stand on the shoulders of giants”

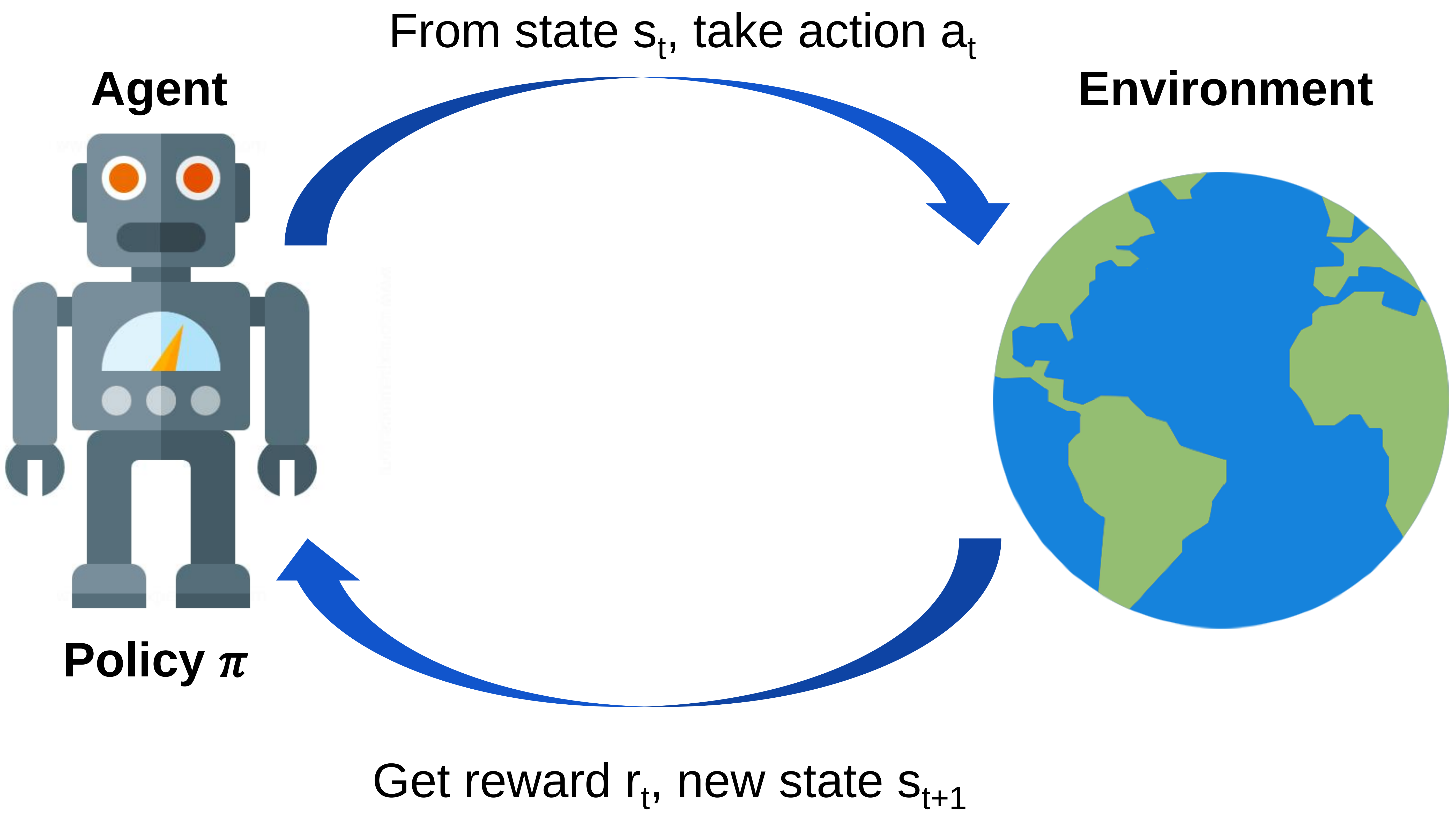
Benefits of social learning for RL

- Individual exploration is dangerous

[1]https://www.gpsworld.com/autonomous-car-hits-autonomous-robot-in-bizarre-collision/
[2]https://www.sciencealert.com/video-captures-self-driving-tesla-hitting-and-killing-a-robot-in-las-vegas

From Koide, K., & Miura, J. (2016). Identification of a specific person using color, height, and gait features for a person following robot. *Robotics and Autonomous Systems*, *84*, 76-87.

Reinforcement Learning



Goal: maximize discounted future **reward**

$$R(\tau) = \sum_{t=0}^{T} \gamma^t r_t$$

γ = discount factor

- **Sequential decision making**: interact with environment at each timestep t
- **Trial and error** learning. When to **explore** vs. **exploit**