Carnegie Mellon University
Language Technologies Institute

Ask & Explore: Grounded Question Answering for Curiosity-driven exploration

Jivat Neet Kaur, Yiding Jiang, Paul Pu Liang

Exploration in Reinforcement Learning

Exploration in complex environments can be hard without structured priors!

Reward formulation in RL

$$r_t = r_t^e + r_t^i$$

[Krebs et al., 2009, Dayan & Sejnowski, 1996, Sutton, 1990]

Extrinsic reward

Exploration bonus

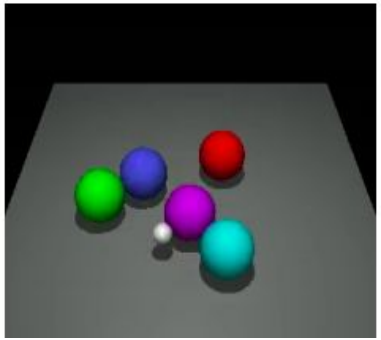
Reward formulation in RL

$$r_t = r_t^e + r_t^i$$

[Krebs et al., 2009, Dayan & Sejnowski, 1996, Sutton, 1990]

Extrinsic reward

Exploration bonus



Goal: “There is a green sphere; are there any rubber cyan balls in front of it?"

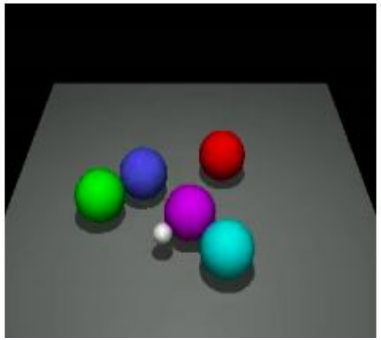
Reward formulation in RL

$$r_t = r_t^e + r_t^i$$

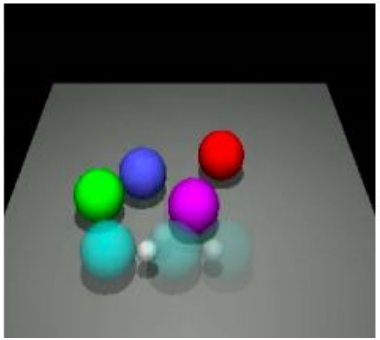
[Krebs et al., 2009, Dayan & Sejnowski, 1996, Sutton, 1990]

Extrinsic reward

Exploration bonus



Goal: “There is a green sphere; are there any rubber cyan balls in front of it?"



Agent performs actions and tries to satisfy goal.

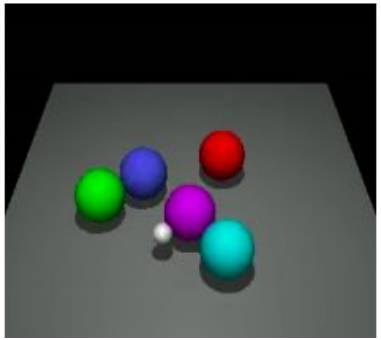
Reward formulation in RL

$$r_t = r_t^e + r_t^i$$

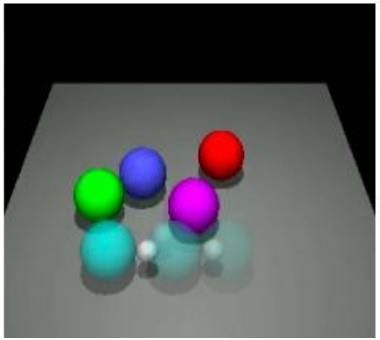
[Krebs et al., 2009, Dayan & Sejnowski, 1996, Sutton, 1990]

Extrinsic reward

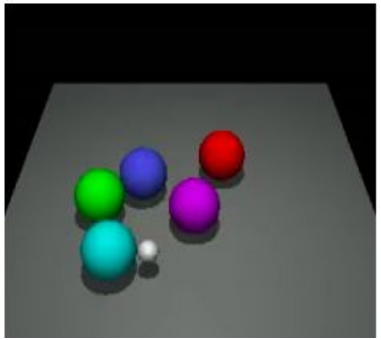
Exploration bonus



Goal: “There is a green sphere; are there any rubber cyan balls in front of it?"



Agent performs actions and tries to satisfy goal.



Resulting state: Agent receives **+1** reward

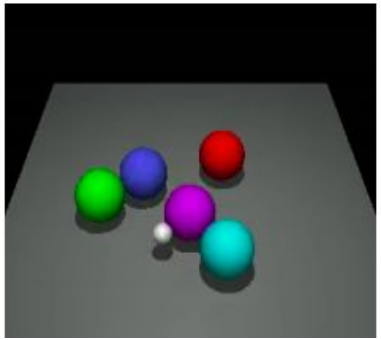
Reward formulation in RL

$$r_t = r_t^e + r_t^i$$

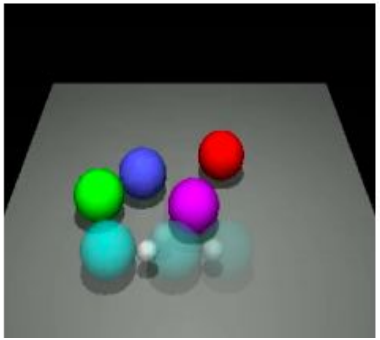
Extrinsic reward

Exploration bonus

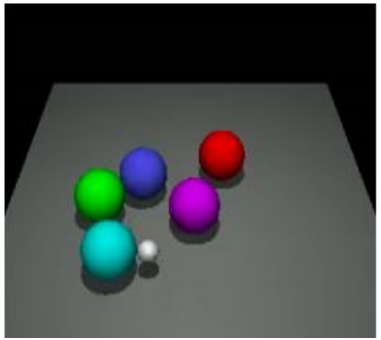
[Krebs et al., 2009, Dayan & Sejnowski, 1996, Sutton, 1990]



Goal: “There is a green sphere; are there any rubber cyan balls in front of it?"



Agent performs actions and tries to satisfy goal.



Resulting state: Agent receives **+1** reward

- Curiosity-driven exploration by self-supervised prediction [Pathak et al., 2017, Burda et al., 2018, Pathak et al., 2019]

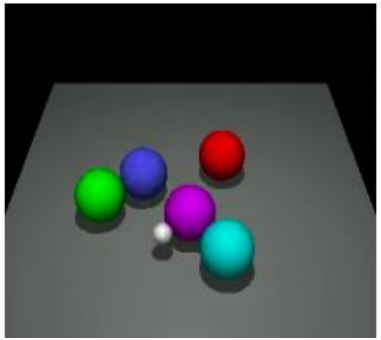
Reward formulation in RL

$$r_t = r_t^e + r_t^i$$

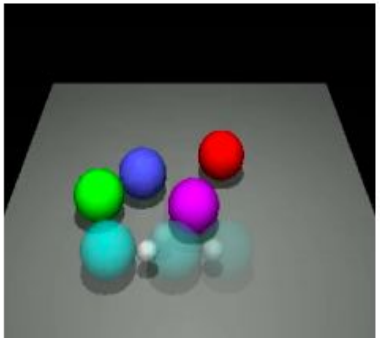
Extrinsic reward

Exploration bonus

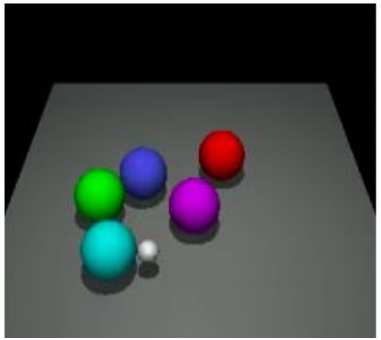
[Krebs et al., 2009, Dayan & Sejnowski, 1996, Sutton, 1990]



Goal: “There is a green sphere; are there any rubber cyan balls in front of it?"



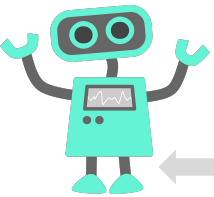
Agent performs actions and tries to satisfy goal.



Resulting state: Agent receives **+1** reward

- Curiosity-driven exploration by self-supervised prediction [Pathak et al., 2017, Burda et al., 2018, Pathak et al., 2019]
- Random Network Distillation (RND) [Burda et al., 2018]

Ask & Explore (ANE)



Agent

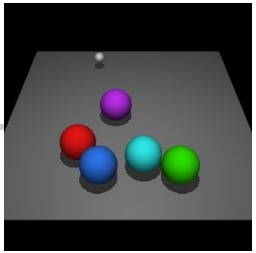
Environment

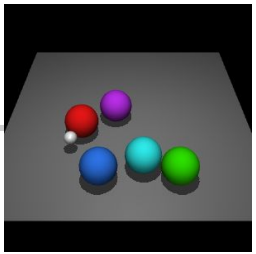
a_t

r_t

s_t

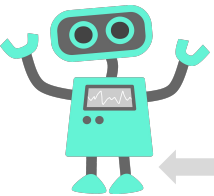
s_{t+1}





q_1	There is a **red** rubber ball; are there any **purple** balls on the right side of it?	False
q_2	There is a **cyan** matte ball, are there any **blue** rubber spheres behind it?	False
q_n		
q_1	There is a **red** rubber ball; are there any **purple** balls on the right side of it?	True
q_2	There is a **cyan** matte ball, are there any **blue** rubber spheres behind it?	False
q_n		

Ask & Explore (ANE)



Agent

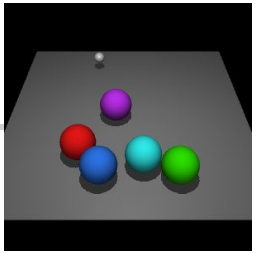
Environment

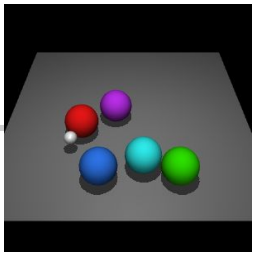
s_t

a_t

r_t

s_{t+1}





q_1 There is a **red** rubber ball; are there any **purple** balls on the right side of it? — False — $A(s_t, q_1)$

q_2 There is a **cyan** matte ball, are there any **blue** rubber spheres behind it? — False — $A(s_t, q_2)$

q_n — $A(s_t, q_n)$

q_1 There is a **red** rubber ball; are there any **purple** balls on the right side of it? — True — $A(s_{t+1}, q_1)$

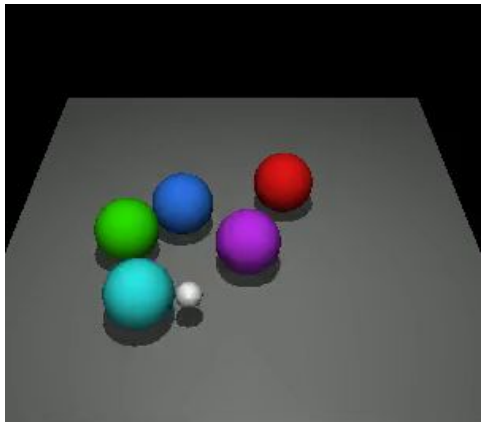
q_2 There is a **cyan** matte ball, are there any **blue** rubber spheres behind it? — False — $A(s_{t+1}, q_2)$

q_n — $A(s_{t+1}, q_n)$

$$r_t^i = \sum_{k=1}^{n} 1[A(s_t, q_k) \neq A(s_{t+1}, q_k)]$$

Dense & Sparse reward setting

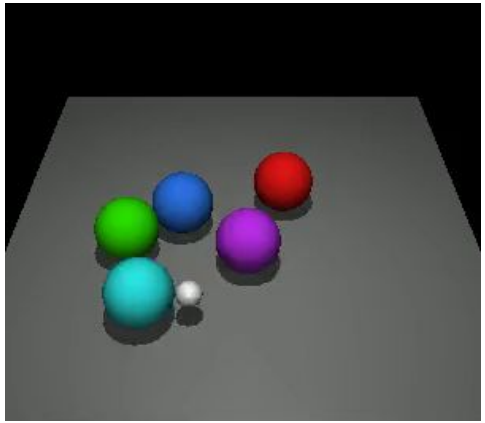
Dense & Sparse reward setting



Goal: “There is a green sphere; are there any rubber cyan balls in front of it?"

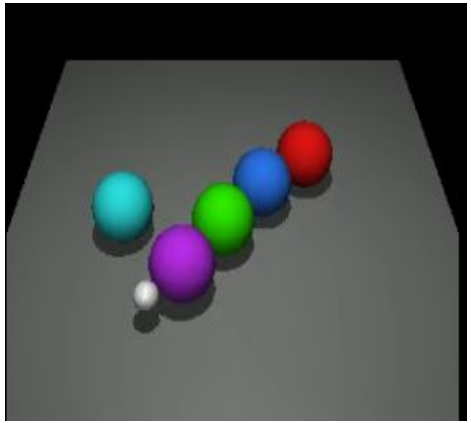
Two object alignment

Dense & Sparse reward setting



Goal: “There is a green sphere; are there any rubber cyan balls in front of it?"

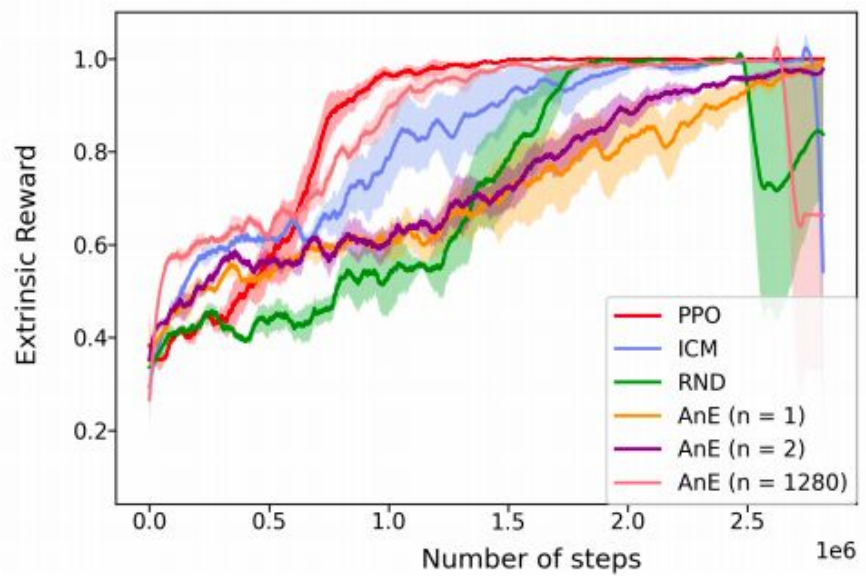
Two object alignment



Goal: “Arrange the objects so that their colors range from blue to green in the horizontal direction, and keep the objects close vertically".

Multiple pairwise object constraints to be mutually satisfied

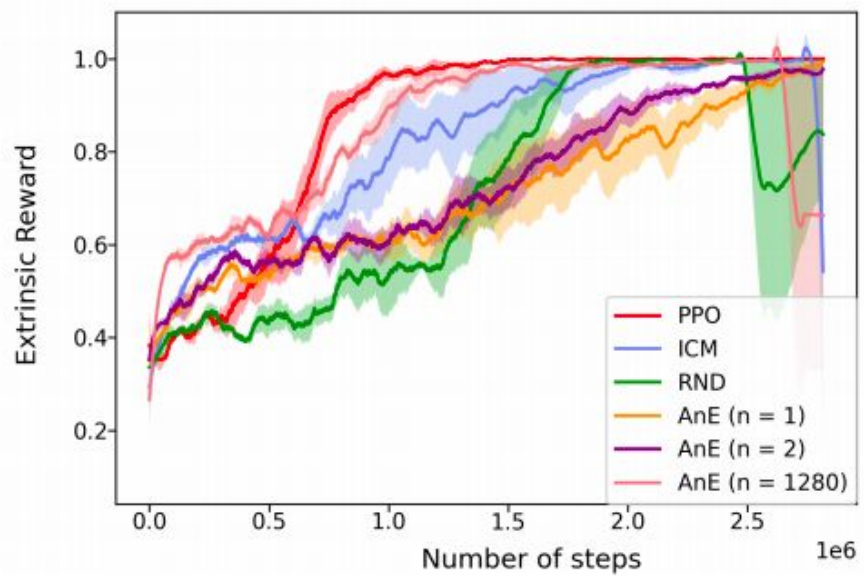
Results



Dense reward setting

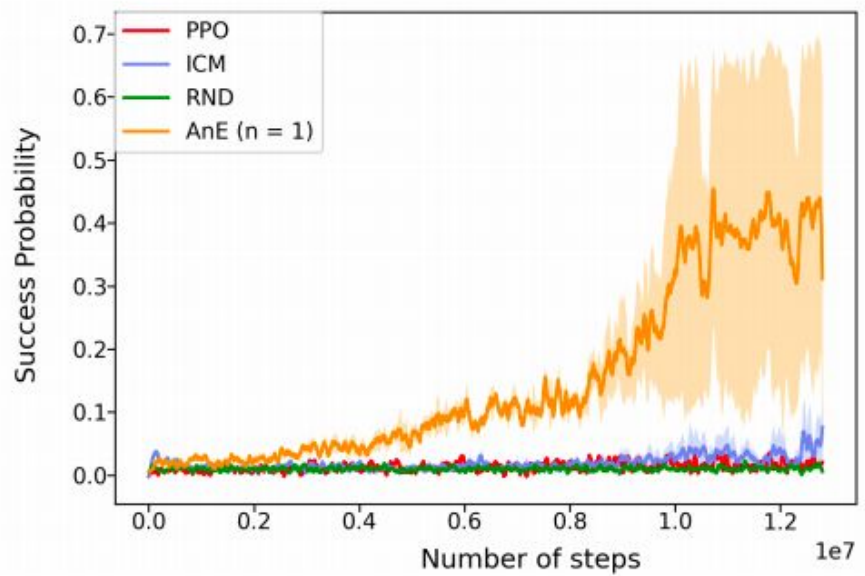
PPO outperforms all curiosity-driven methods

Results



Dense reward setting

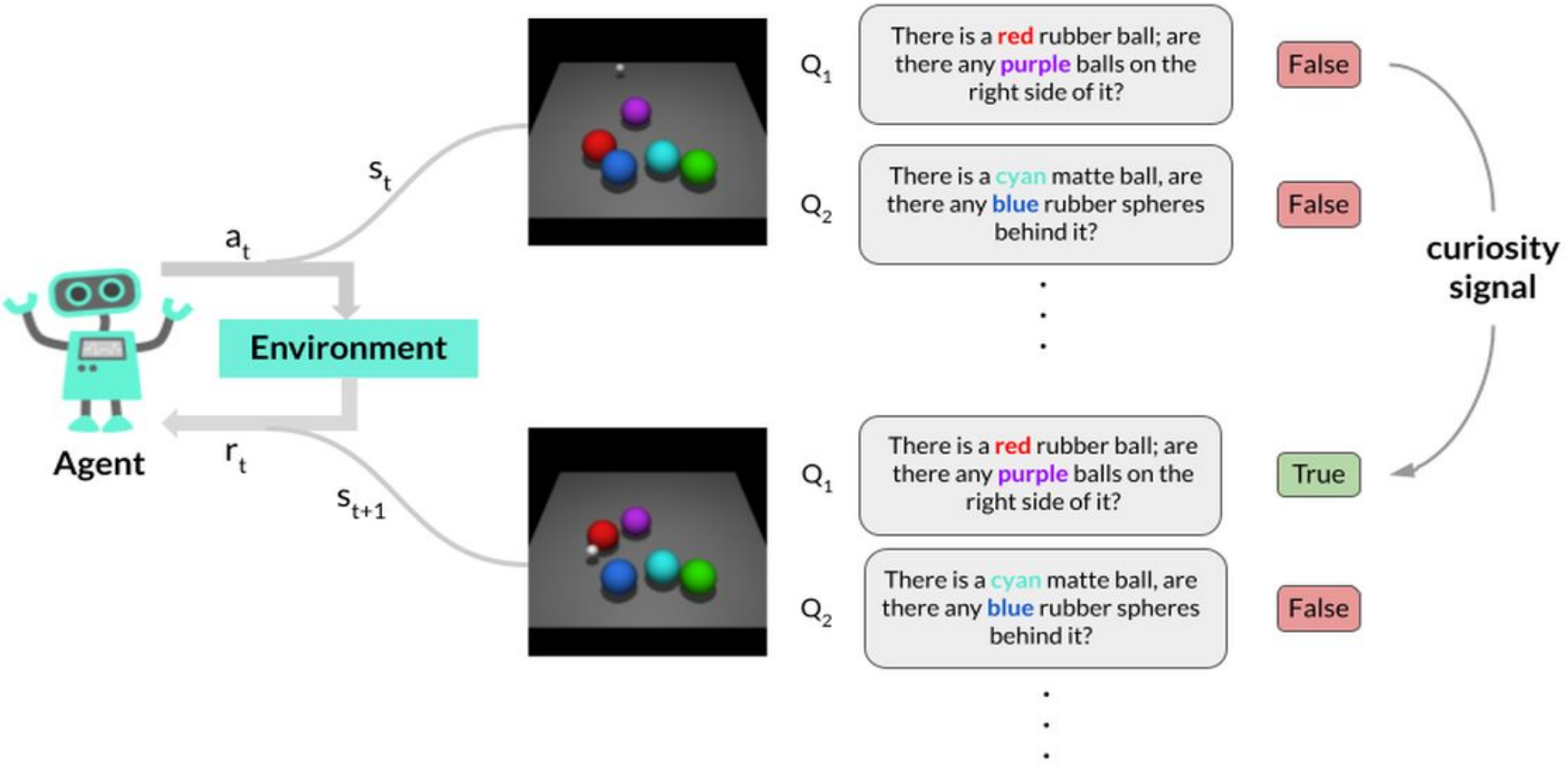
PPO outperforms all curiosity-driven methods



Sparse reward setting

AnE significantly outperforms baselines using single question

Conclusion


Agent
Environment
a_t
s_t
r_t
s_{t+1}
Q_1
There is a red rubber ball; are there any purple balls on the right side of it?
False
Q_2
There is a cyan matte ball, are there any blue rubber spheres behind it?
False
curiosity signal
Q_1
There is a red rubber ball; are there any purple balls on the right side of it?
True
Q_2
There is a cyan matte ball, are there any blue rubber spheres behind it?
False

Thank you!