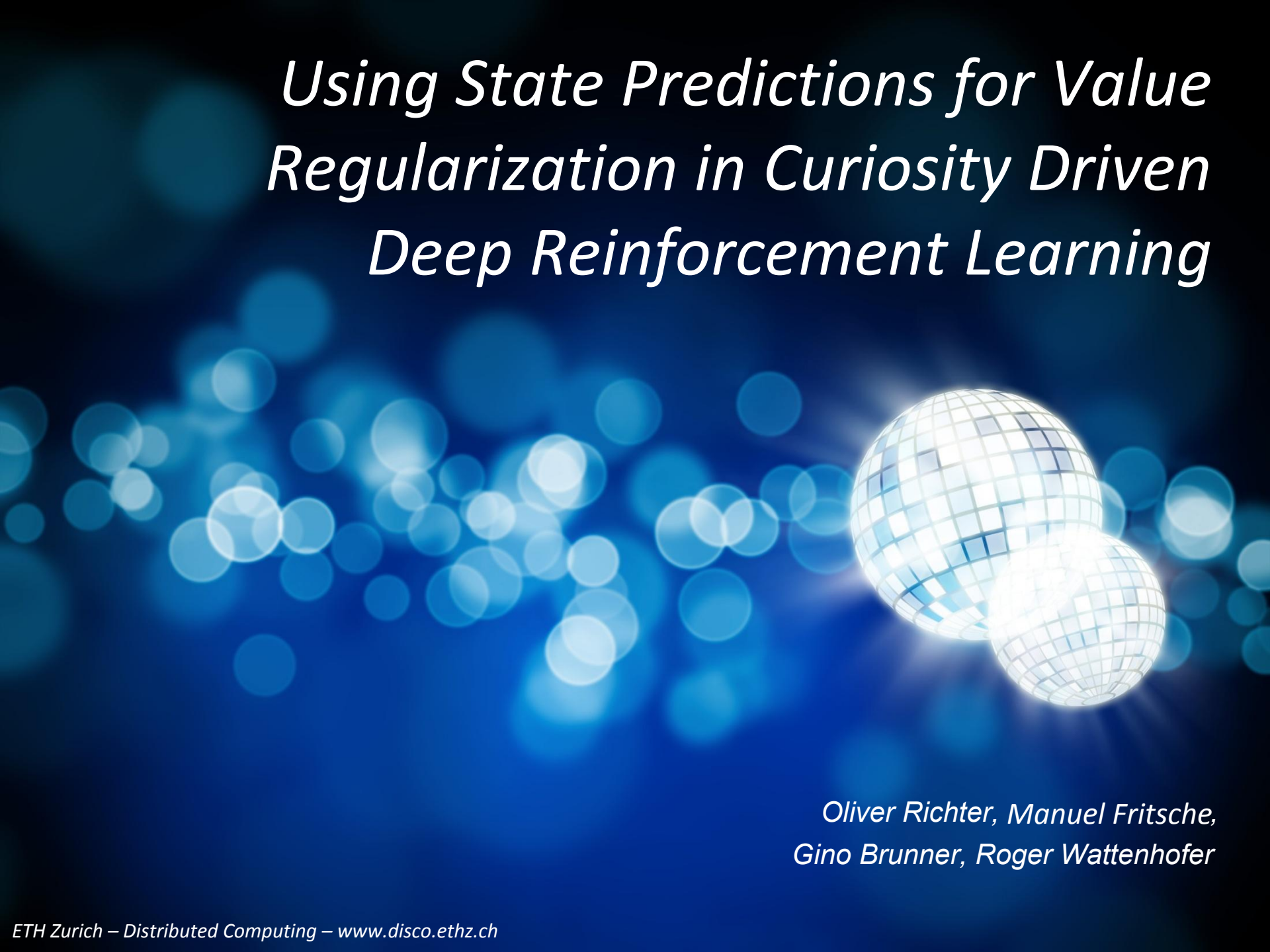
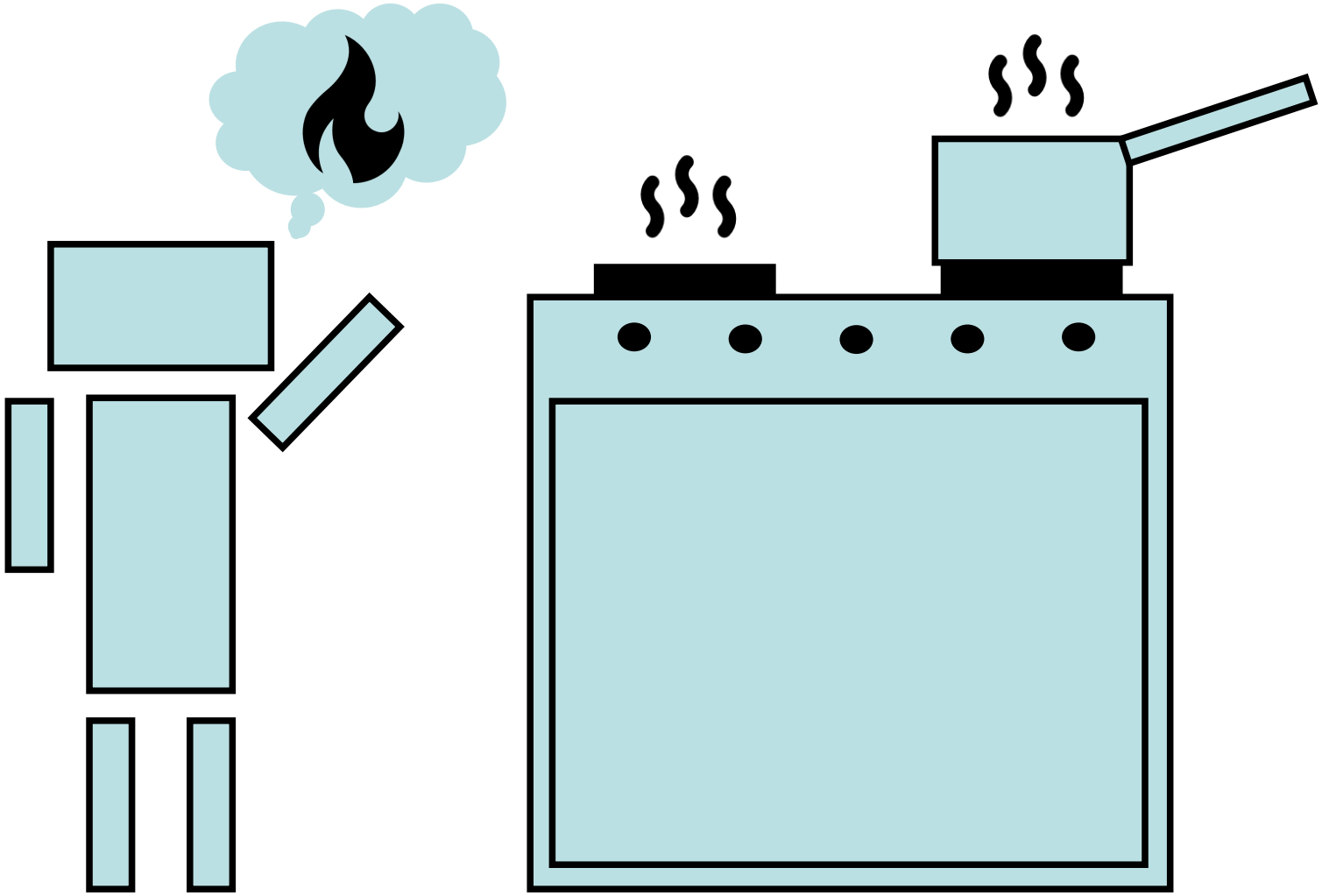
Using State Predictions for Value Regularization in Curiosity Driven Deep Reinforcement Learning

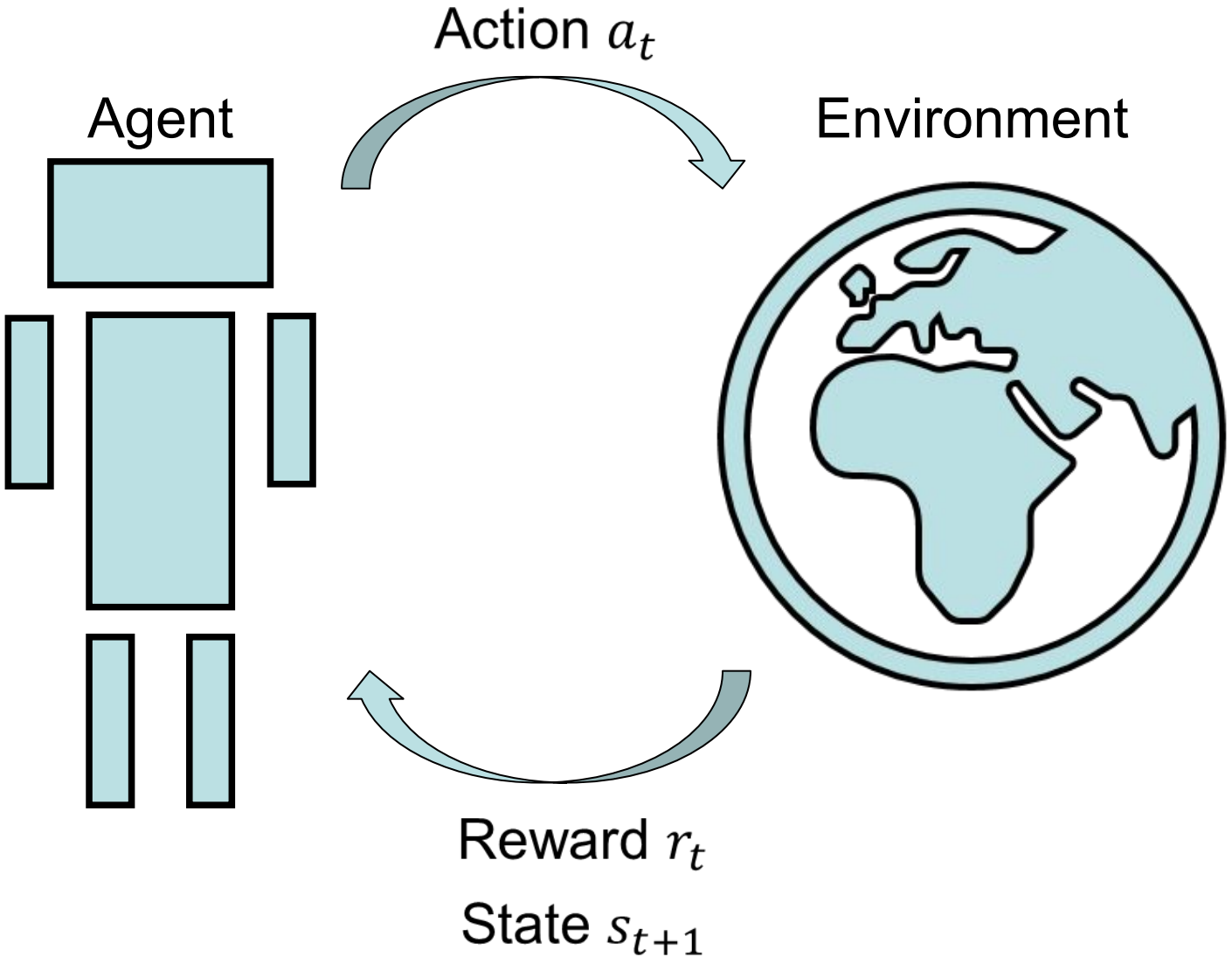


Oliver Richter, Manuel Fritsche,
Gino Brunner, Roger Wattenhofer

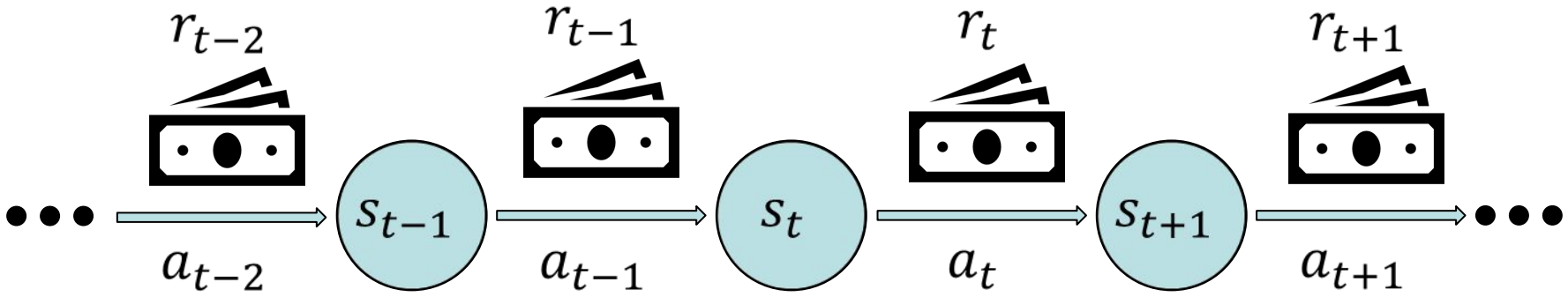
Base actions on predictions



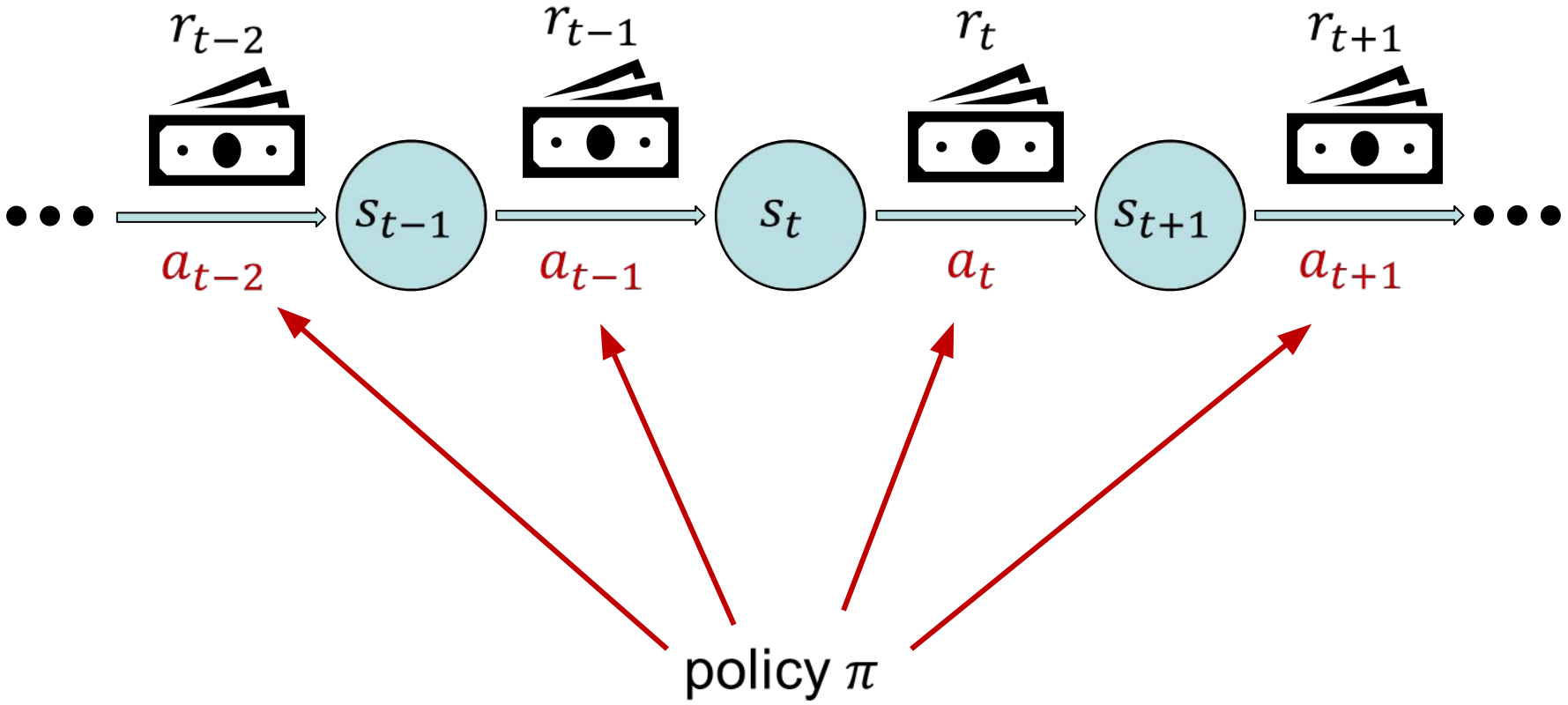
Reinforcement learning



Reinforcement learning

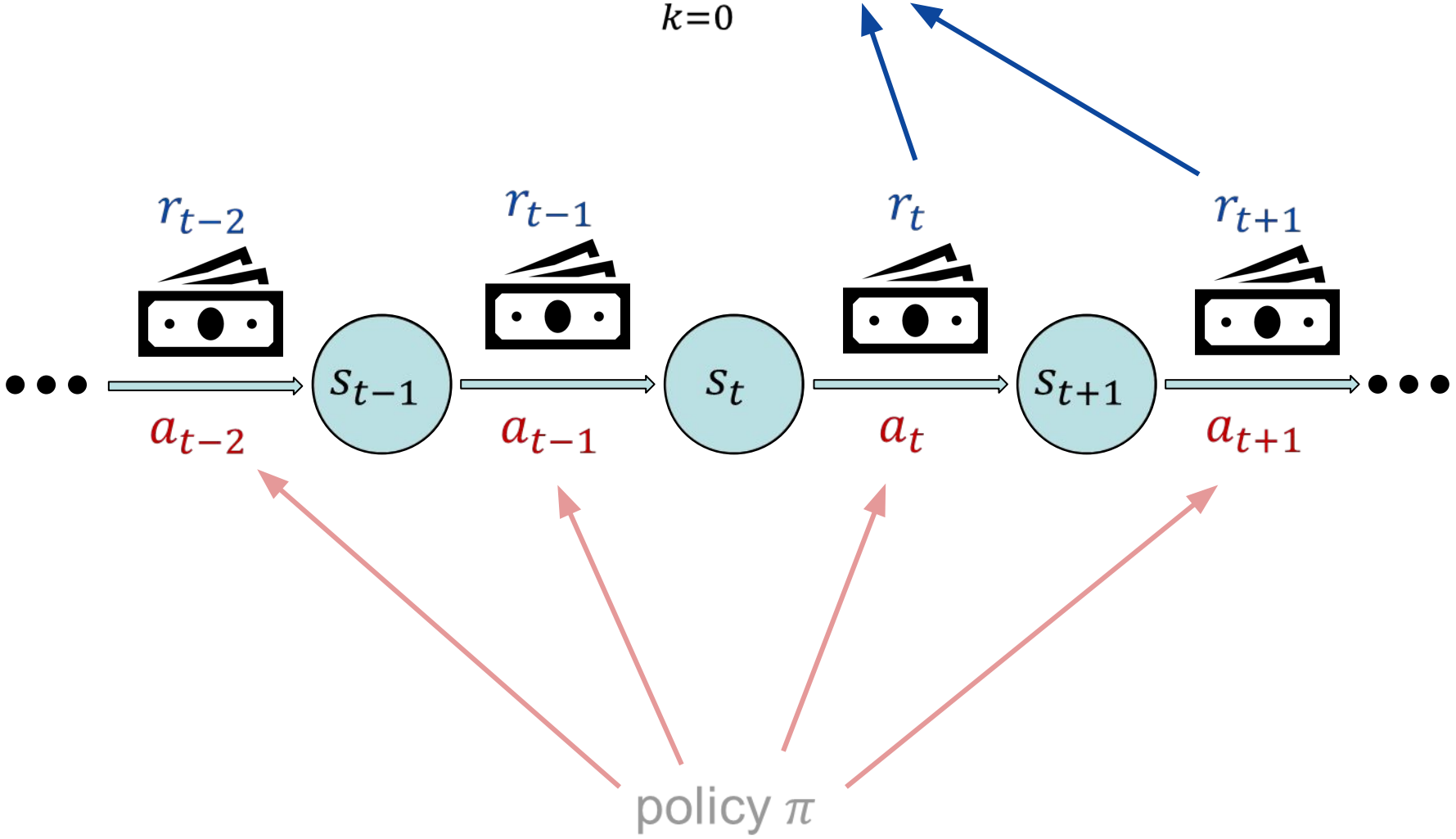


How to choose the action?



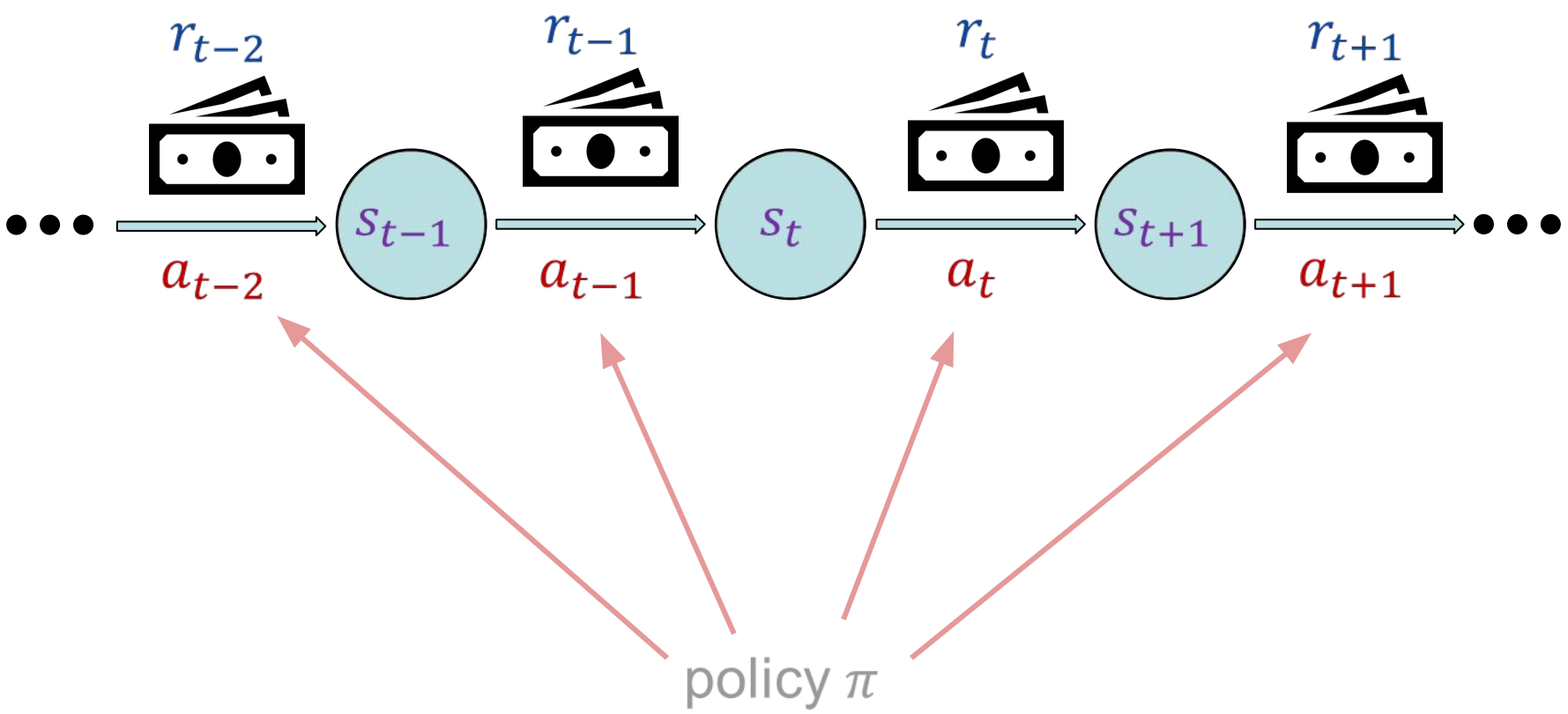
Return value

$$R_t = \sum_{k=0}^{\infty} \gamma^k \, r_{t+k}$$

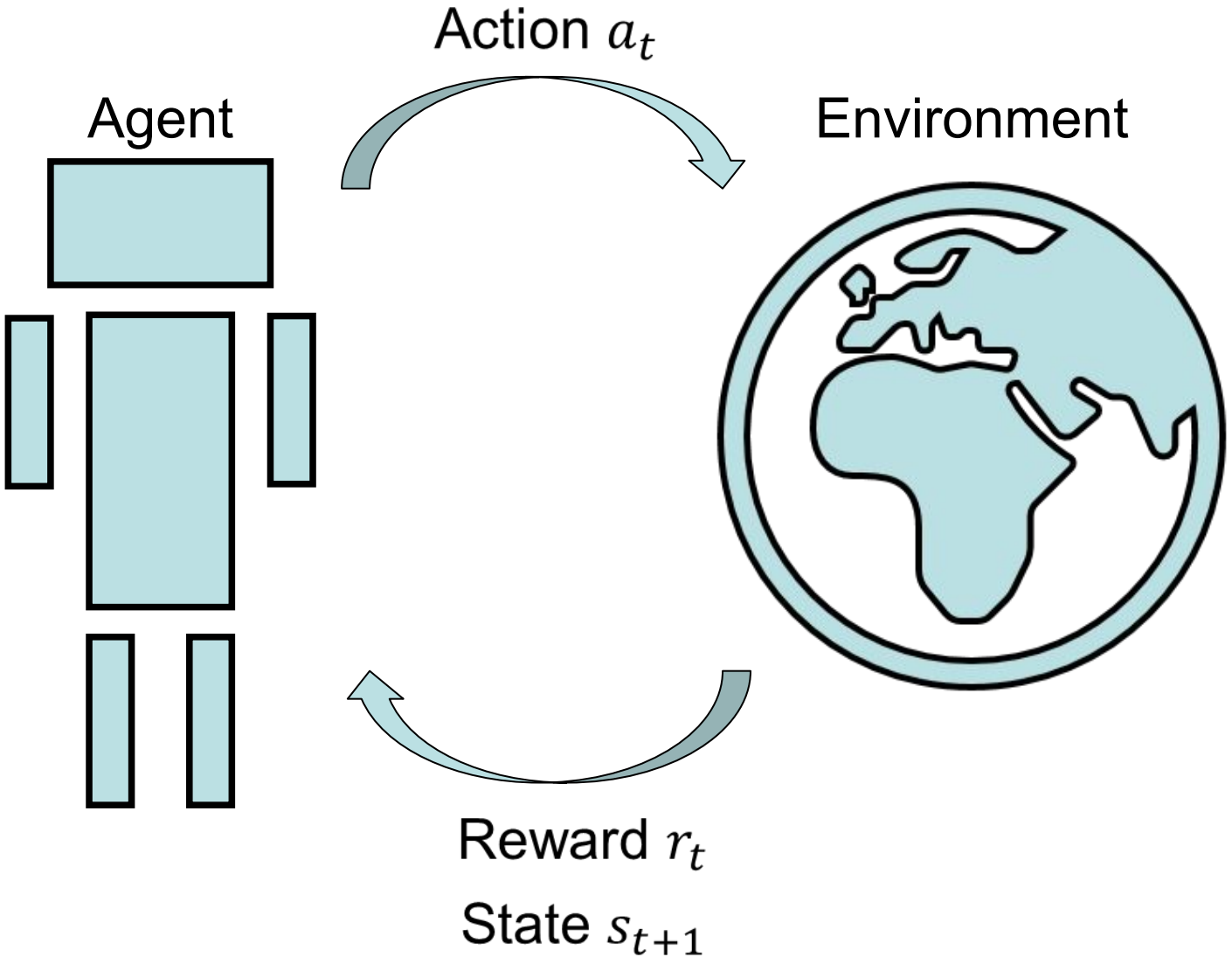


Value function

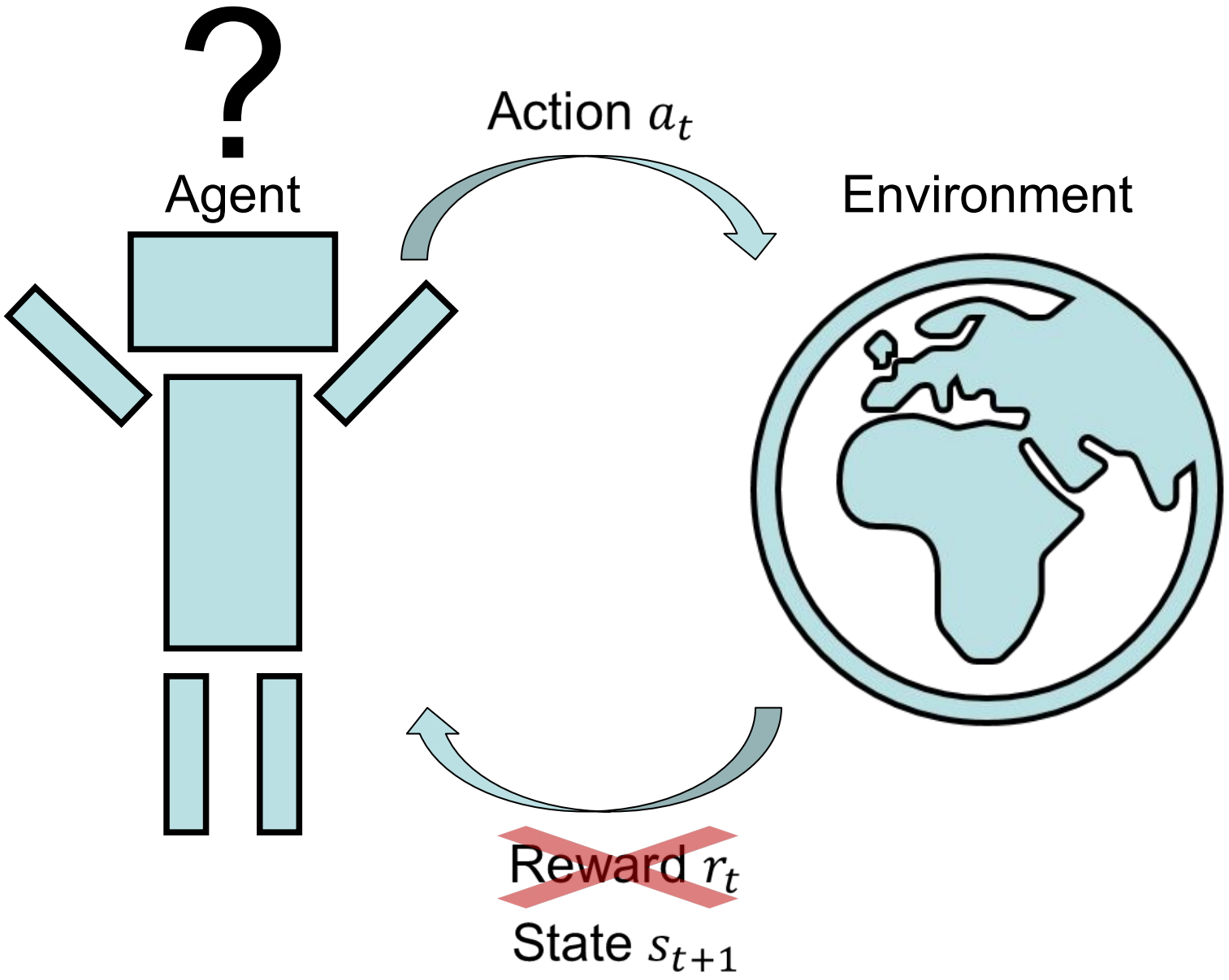
$$V^{\pi}(s_t) = \mathbb{E}_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k \, r_{t+k}\right]$$



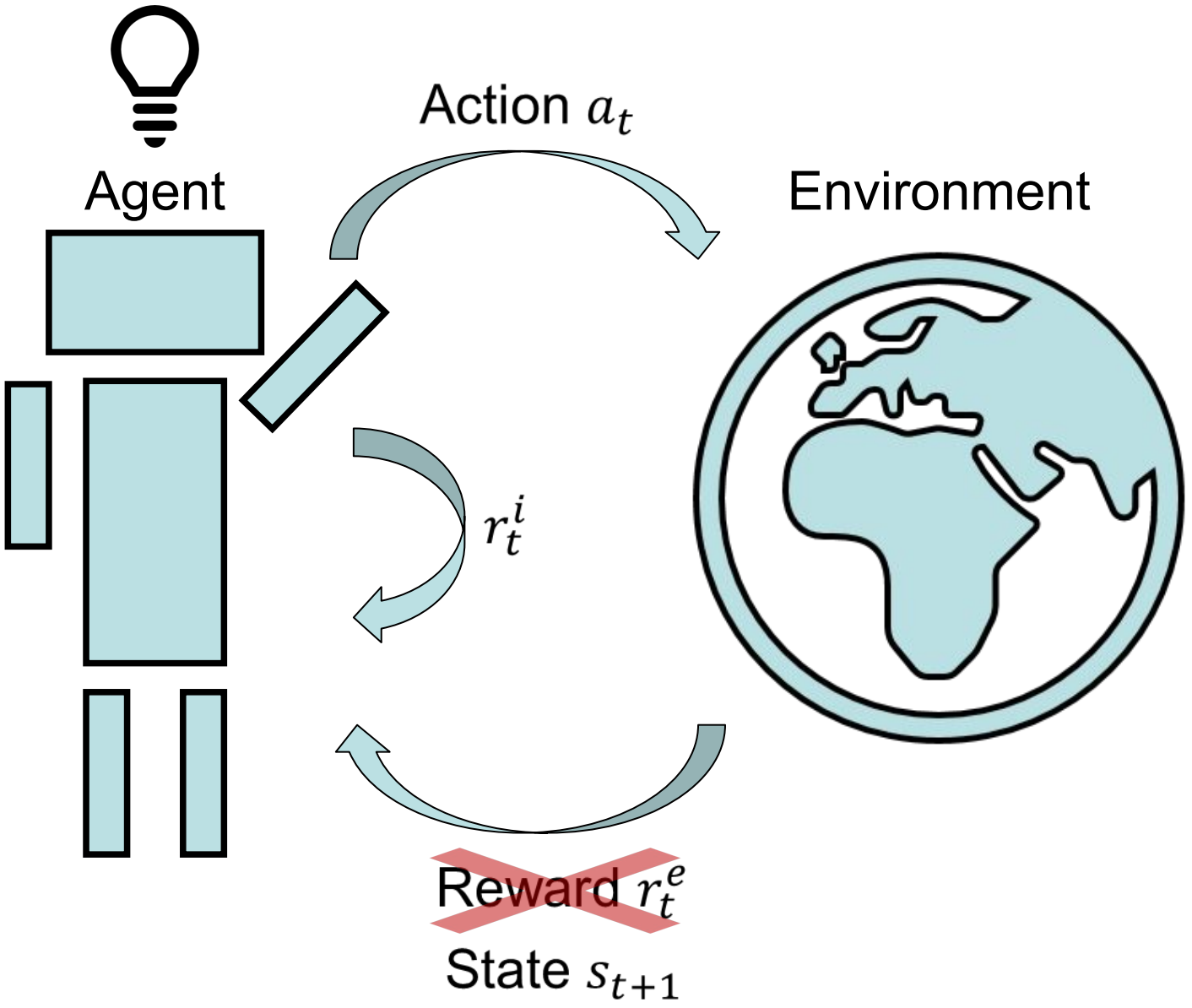
Reinforcement learning



Sparse reward settings



Use intrinsic rewards r_t^i

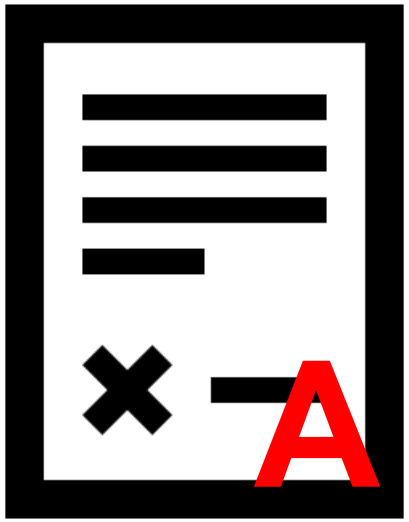


Reward the exploration of novel states

Reward the exploration of novel states

How to find novel states?

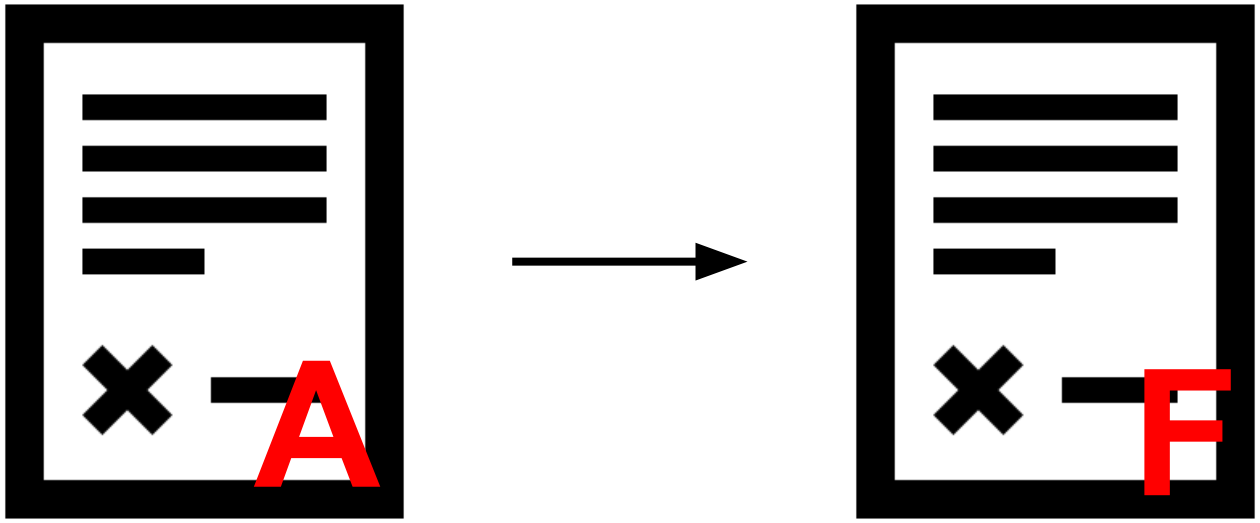
make predictions



How to find novel states?

make predictions

get surprised



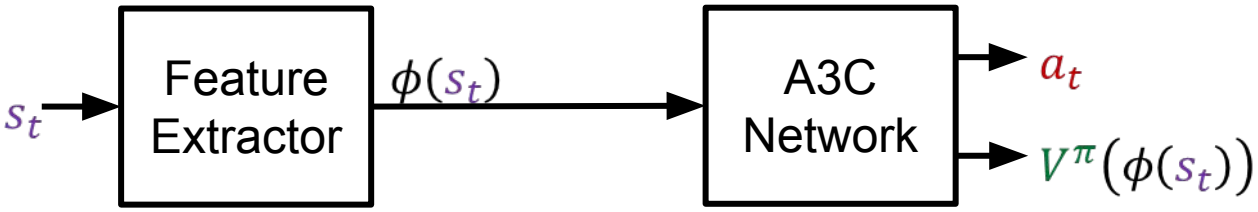
Curiosity

$$r_t^i = \beta \left\| \hat{\phi}(s_{t+1}) - \phi(s_{t+1}) \right\|^2$$

prediction

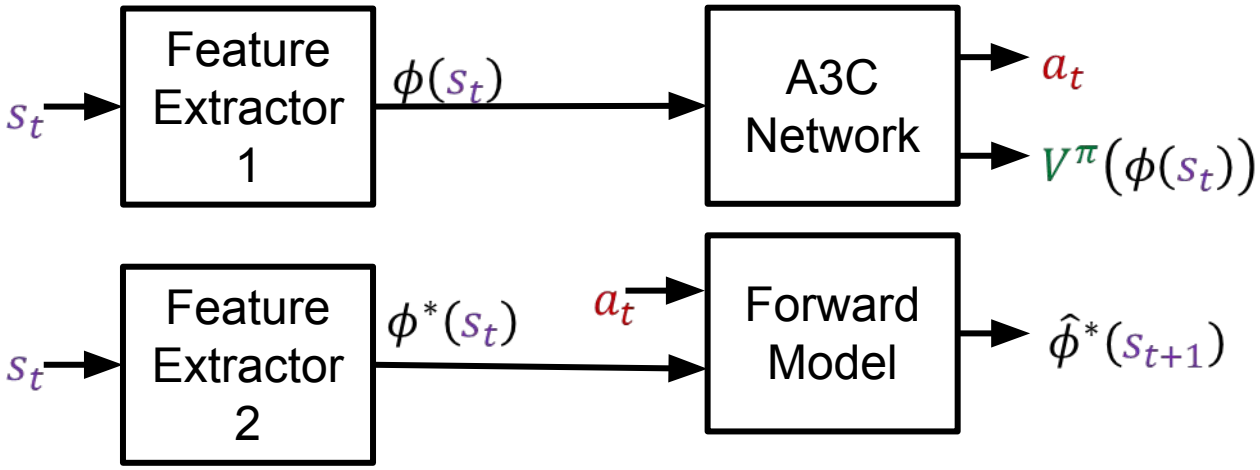
reality

Asynchronous Advantage Actor-Critic architecture (A3C)

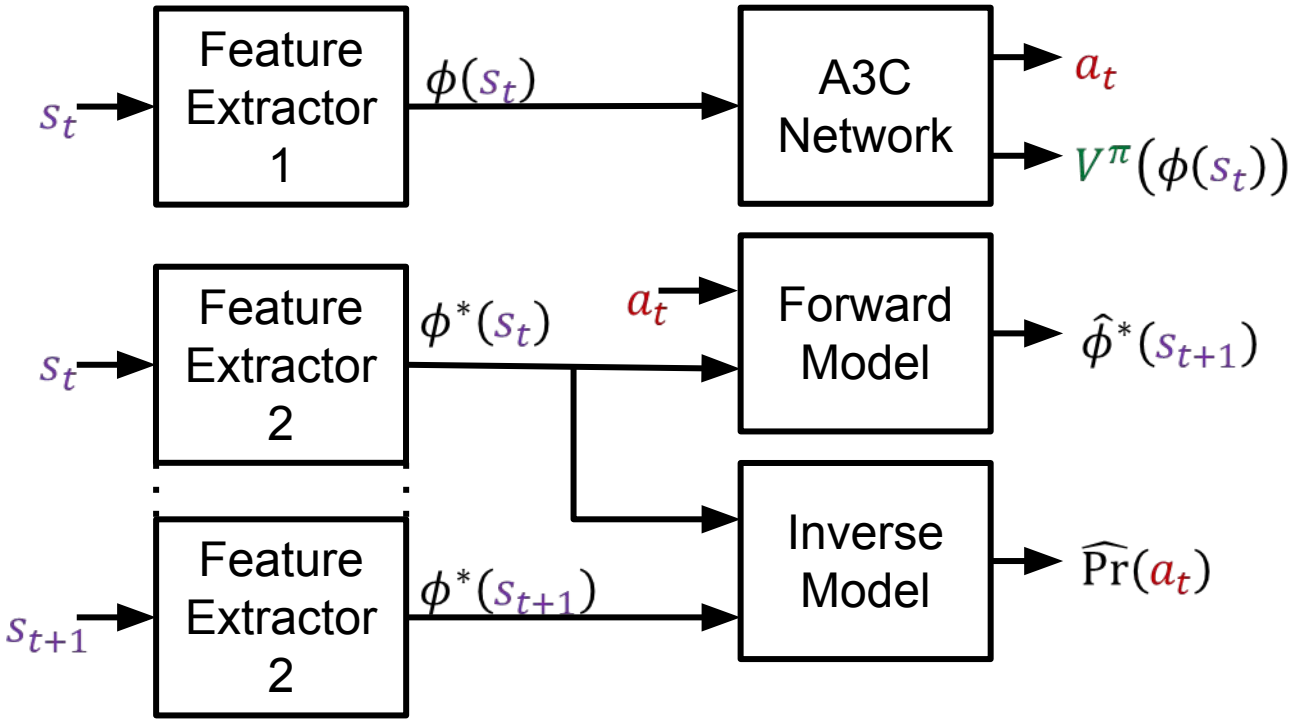


A3C

Adding curiosity

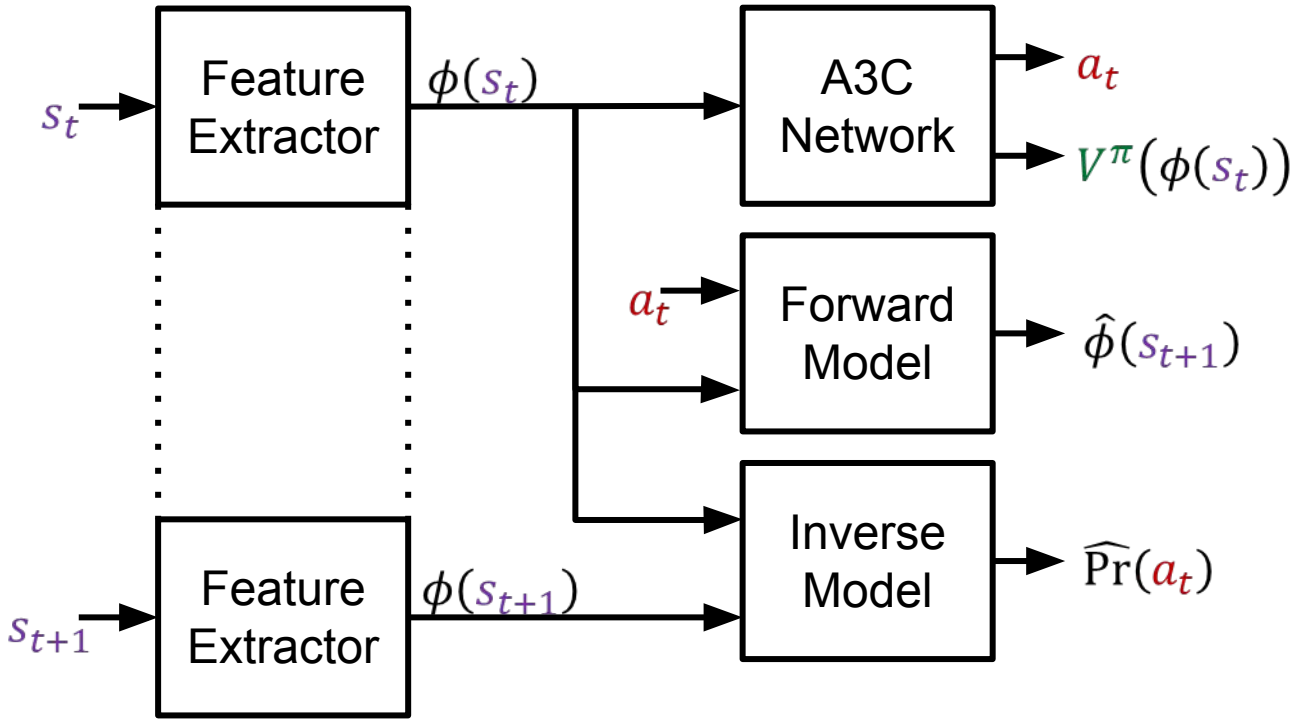


Learning good features



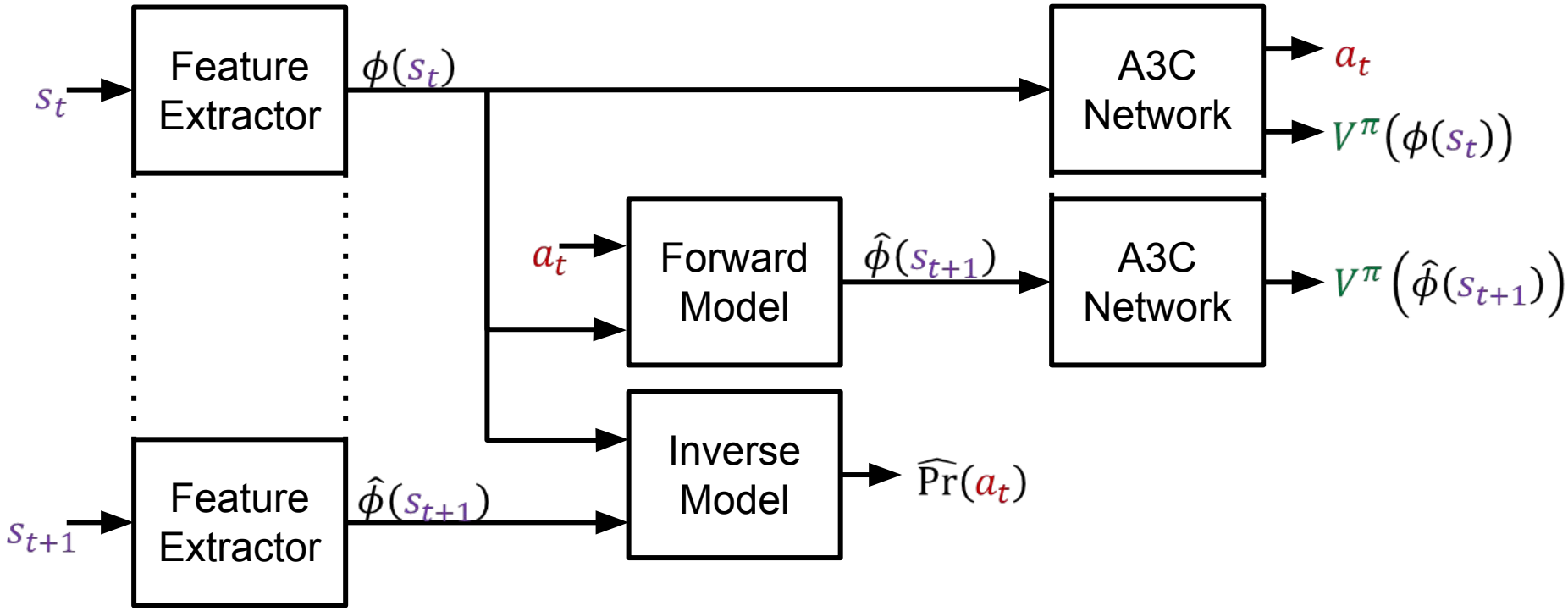
Pathak et. al, ICML 2017, **A3C + ICM**

Good features for all



A3C + Pred

Adding Value Prediction



A3C + Pred + VPC

Value Prediction Consistency

$$V^{\pi}(s_t) = \mathbb{E}_{\pi}[r_t] + \gamma V^{\pi}(s_{t+1})$$

Value Prediction Consistency

$$V^{\pi}(s_t) = \mathbb{E}_{\pi}[r_t] + \gamma V^{\pi}(s_{t+1})$$

$$V^{\pi}(s_{t+1}) = \frac{V^{\pi}(s_t) - \mathbb{E}_{\pi}[r_t]}{\gamma}$$

Value Prediction Consistency

$$V^{\pi}(s_t) = \mathbb{E}_{\pi}[r_t] + \gamma V^{\pi}(s_{t+1})$$

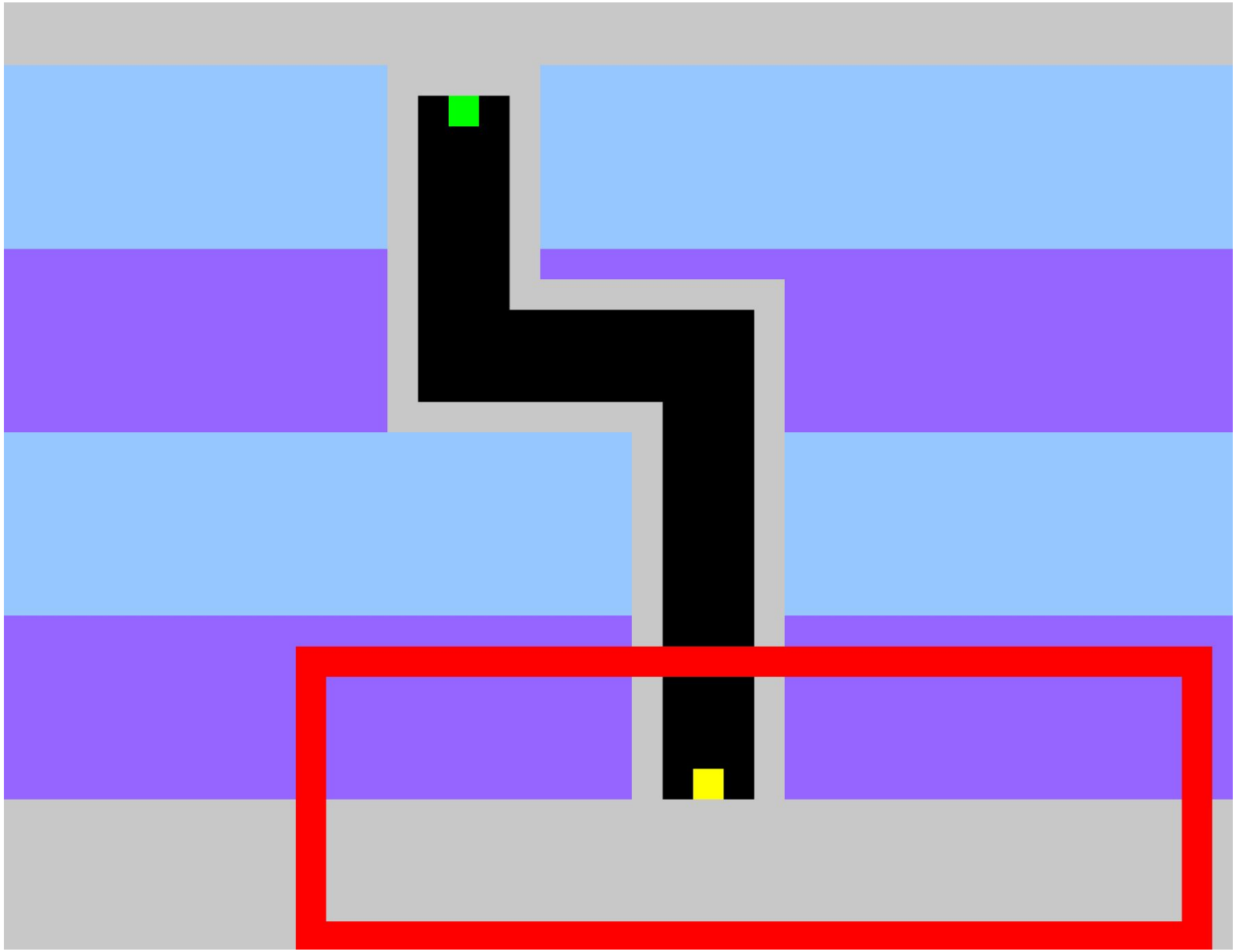
$$\downarrow$$

$$V^{\pi}(s_{t+1}) = \frac{V^{\pi}(s_t) - \mathbb{E}_{\pi}[r_t]}{\gamma}$$

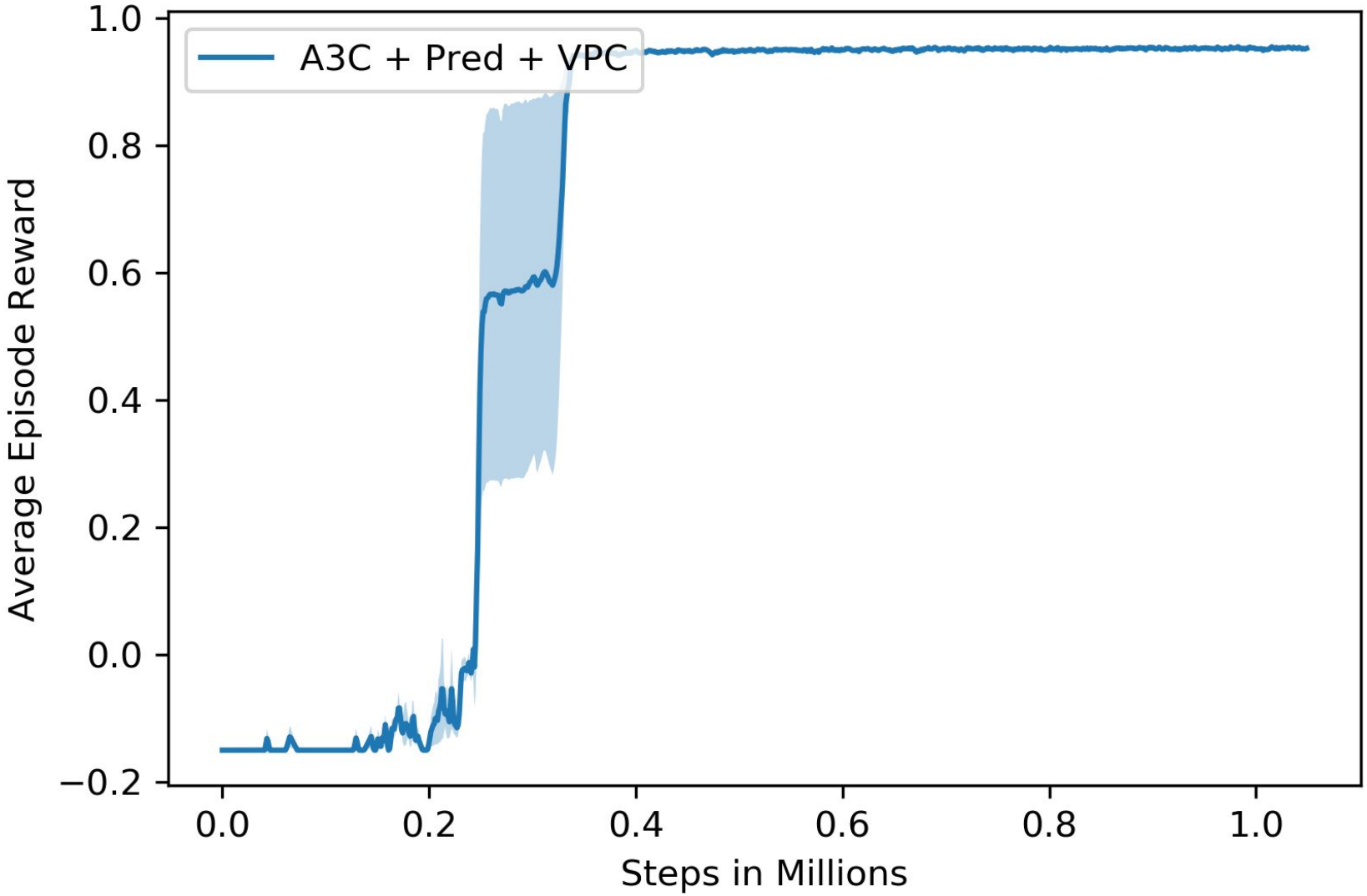
$$\downarrow$$

$$e_{\text{VPC}} = \left\| V^{\pi}\left(\hat{\phi}(s_{t+1})\right) - \frac{V^{\pi}(s_t) - \mathbb{E}_{\pi}[r_t]}{\gamma} \right\|^2$$

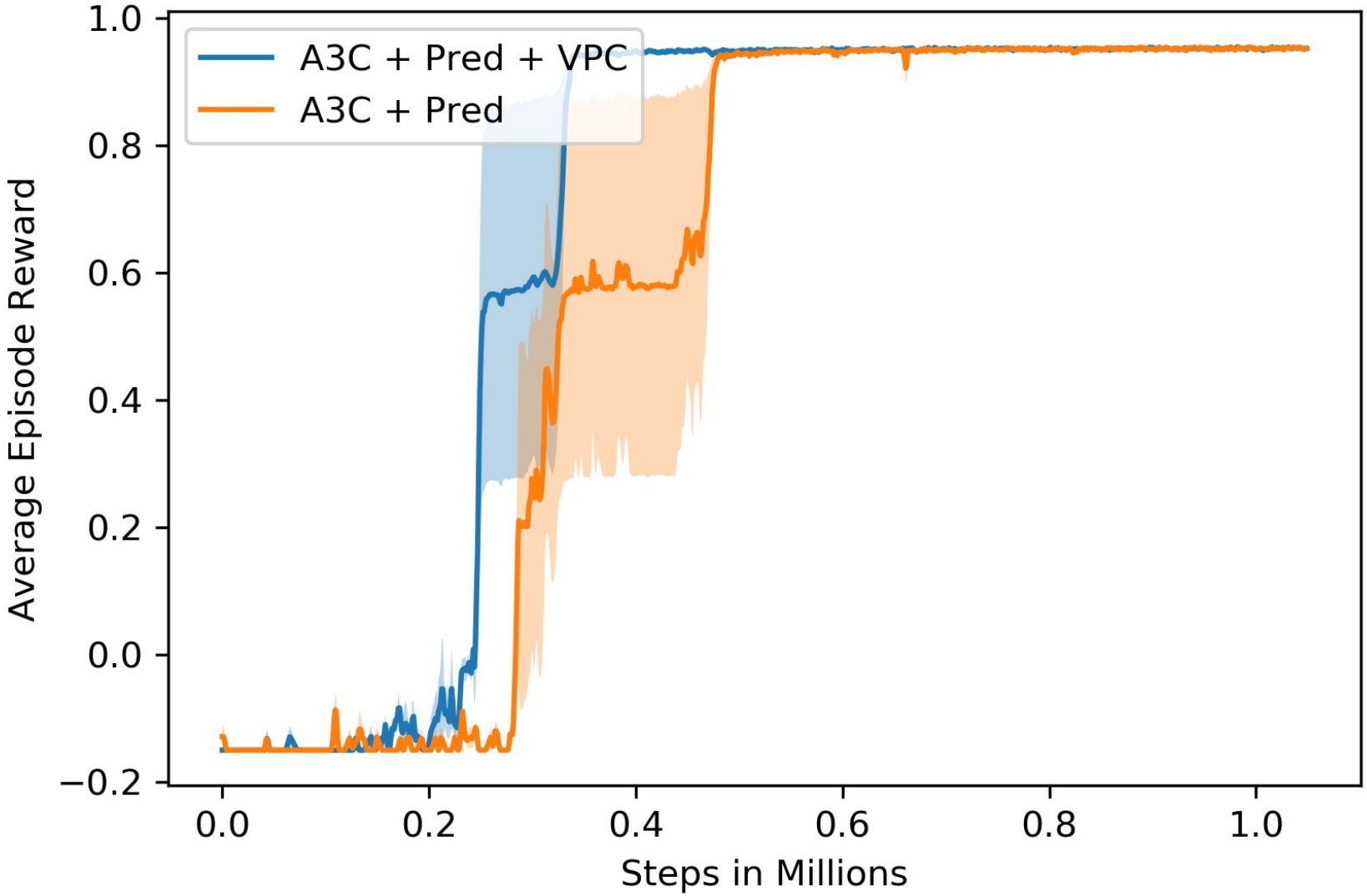
Let’s see how it works in practice



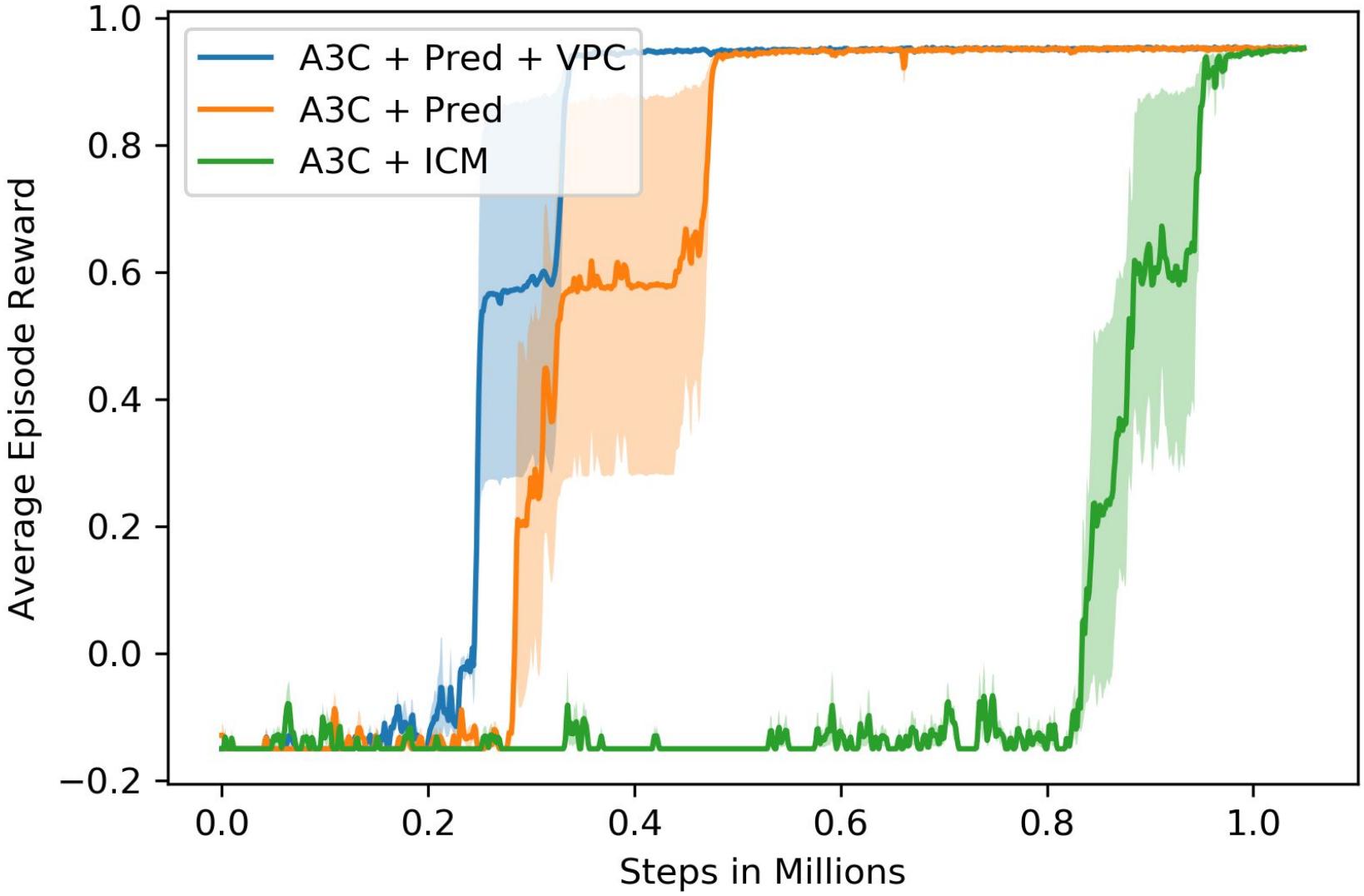
Rewards per episode



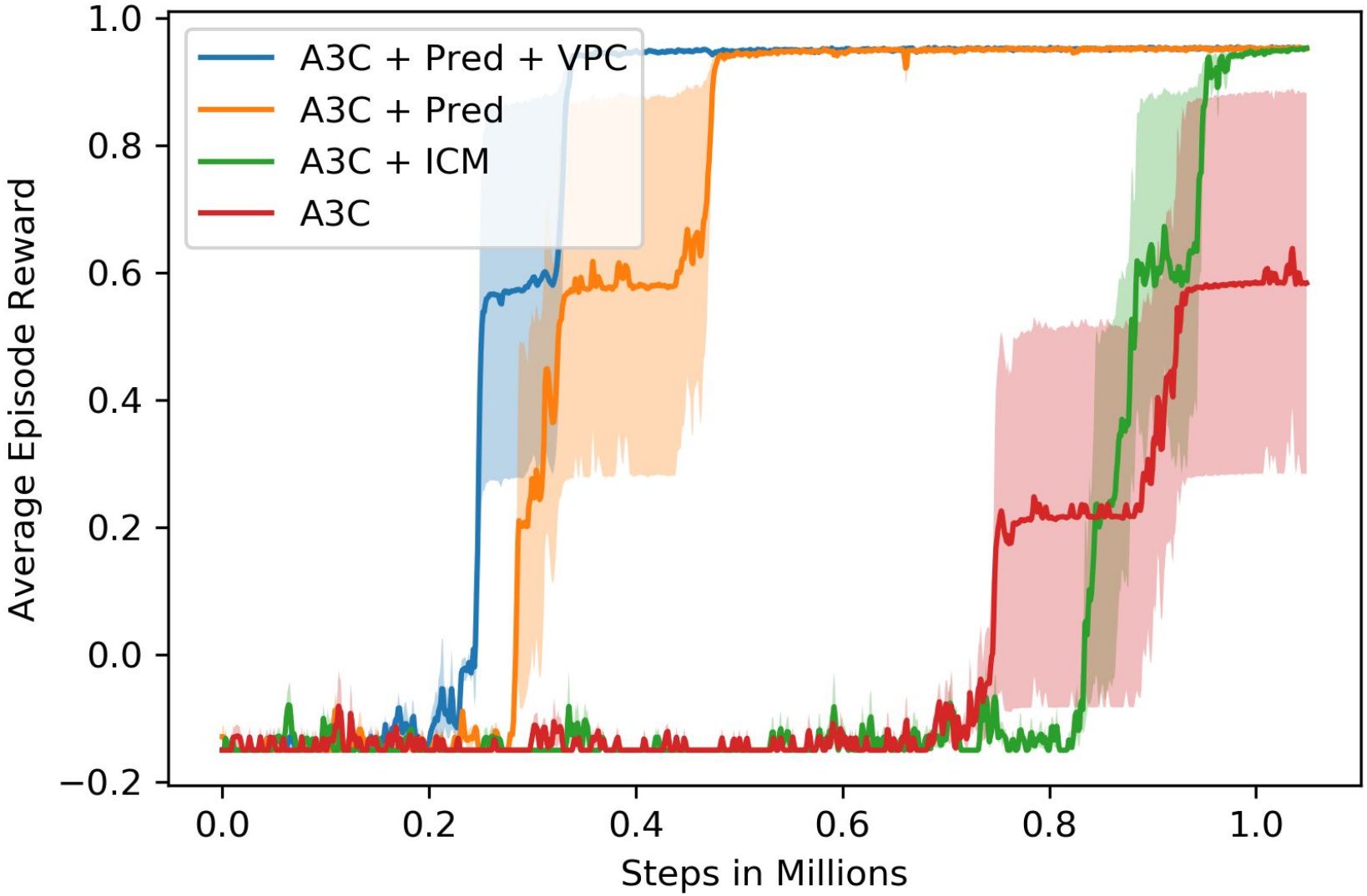
Rewards per episode



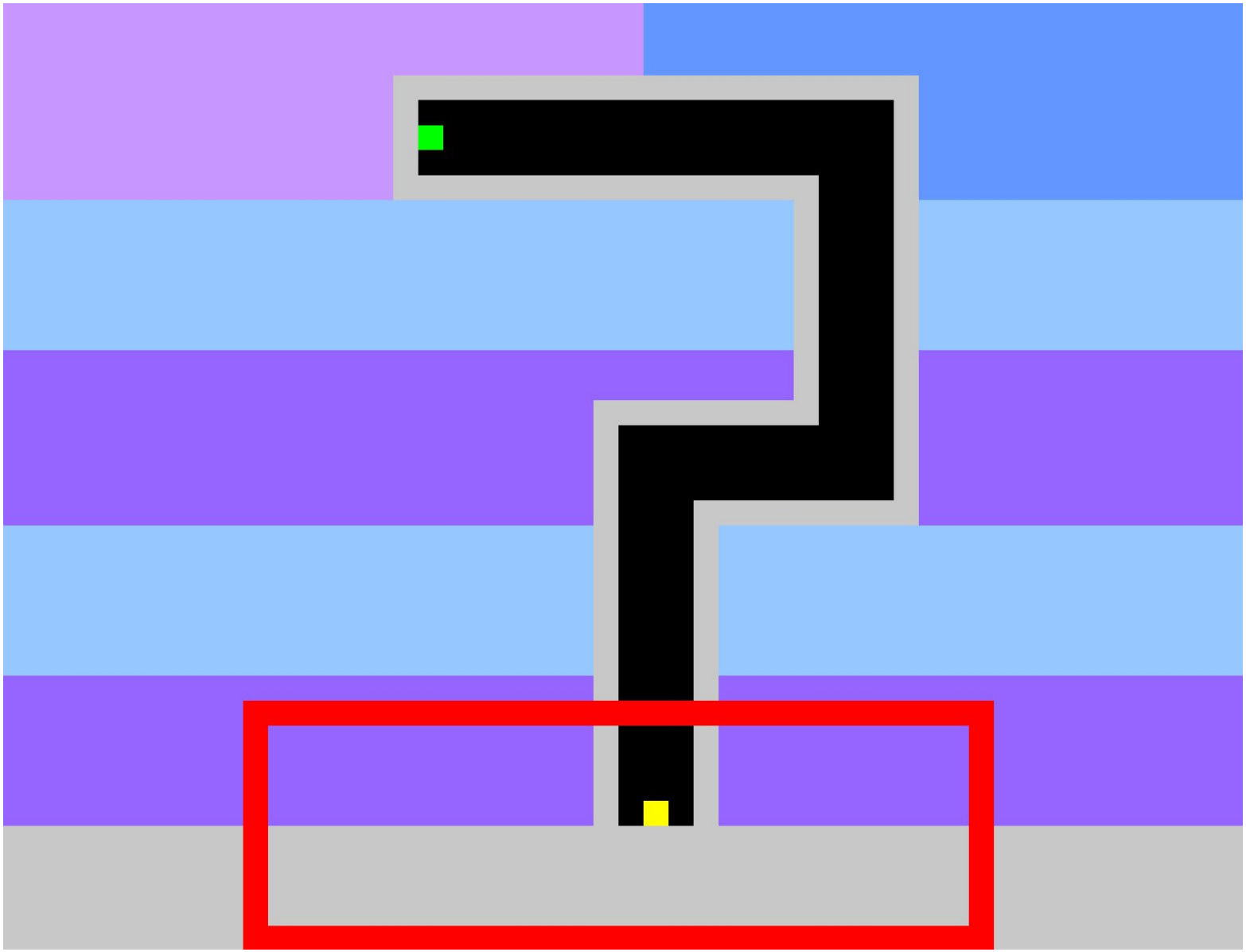
Rewards per episode



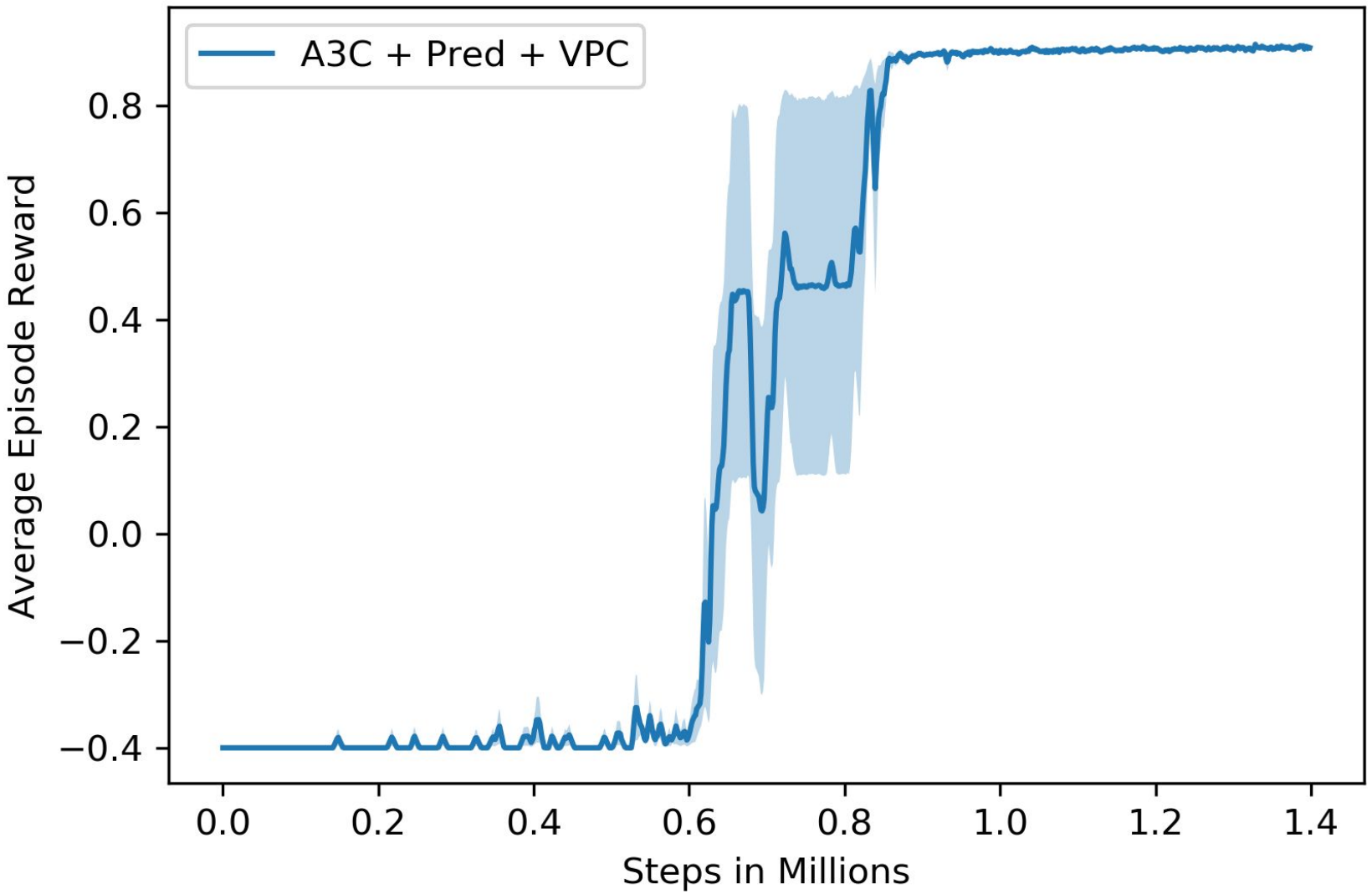
Rewards per episode



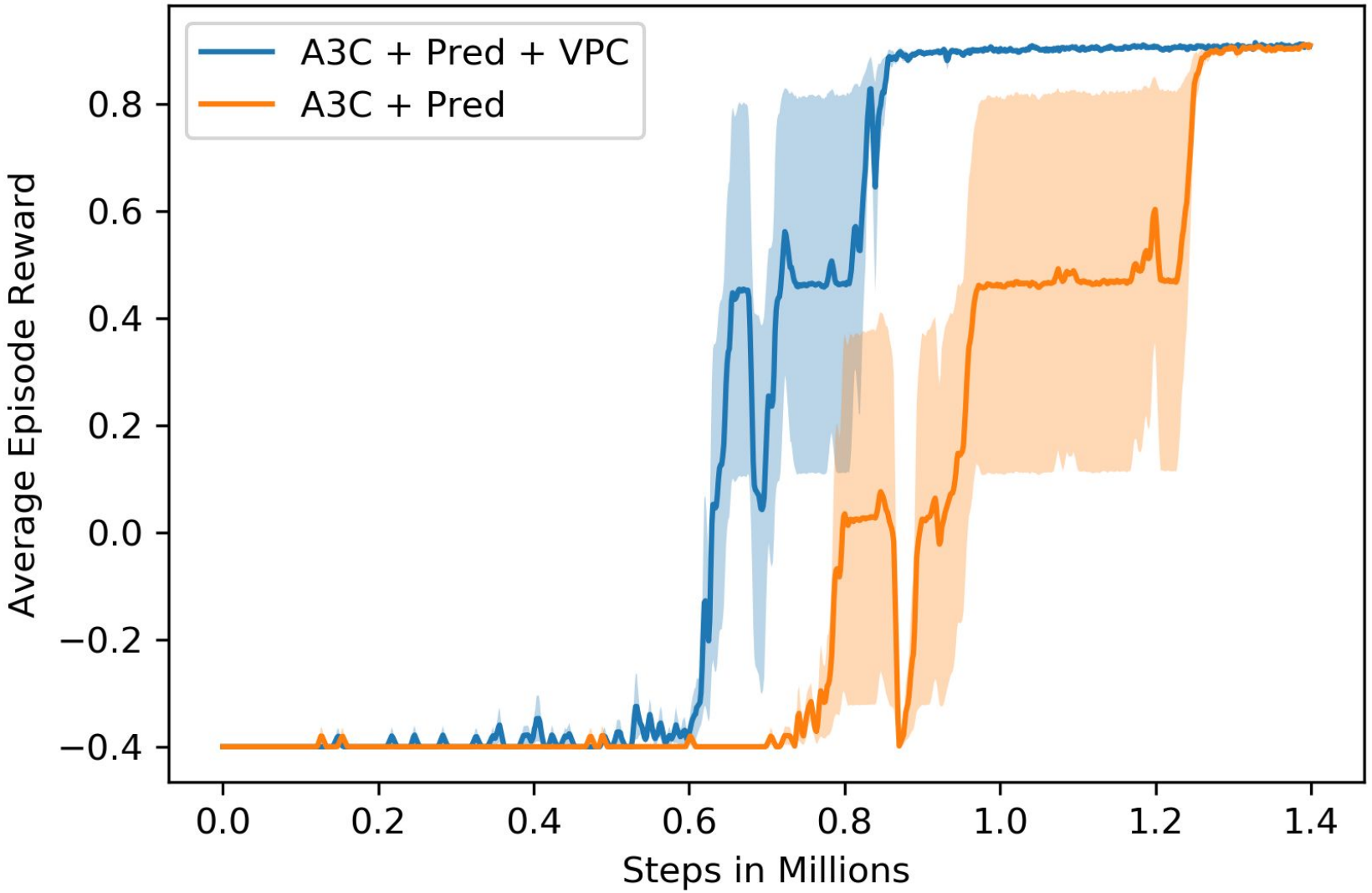
Thinking bigger



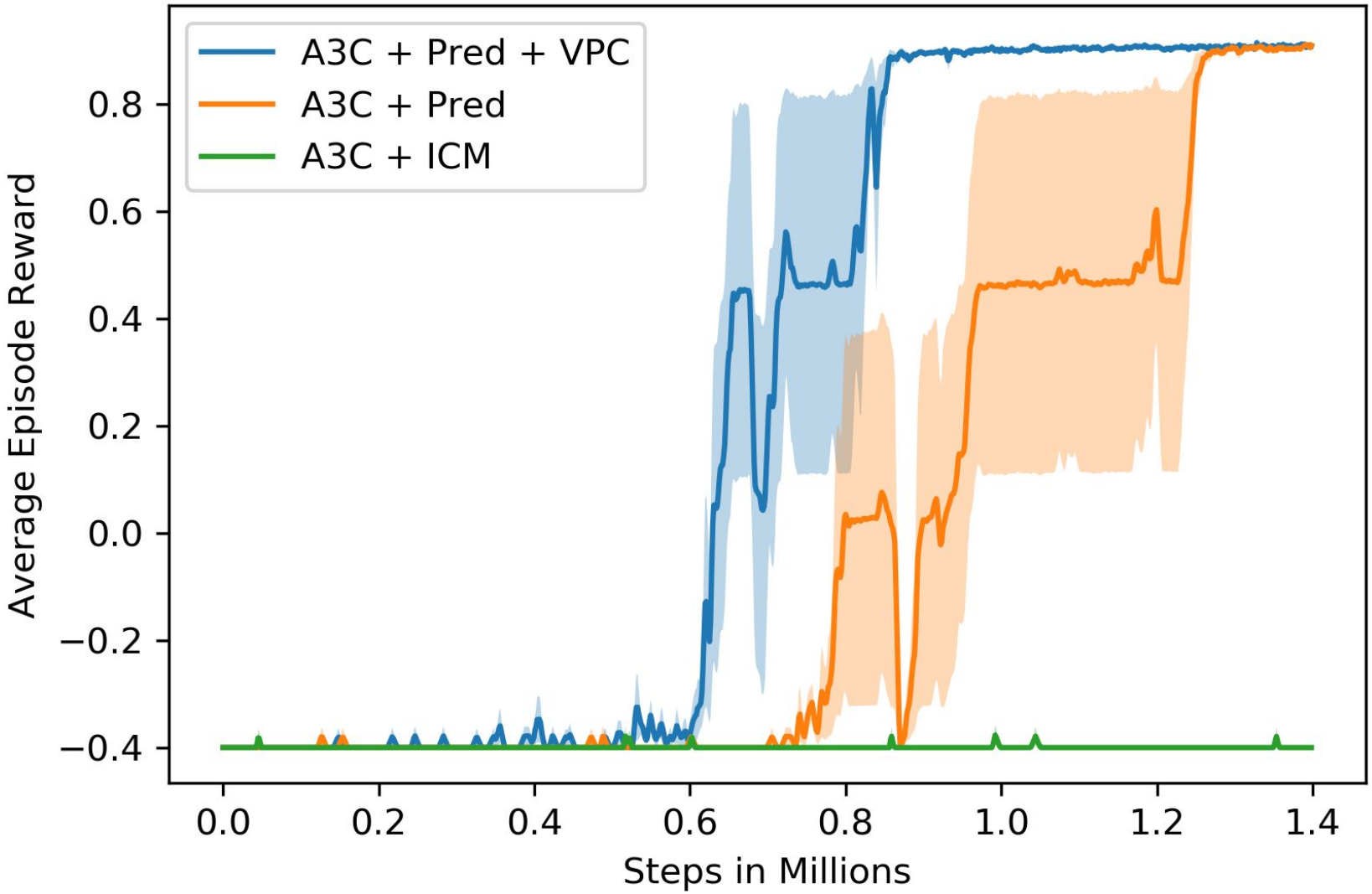
Rewards per episode



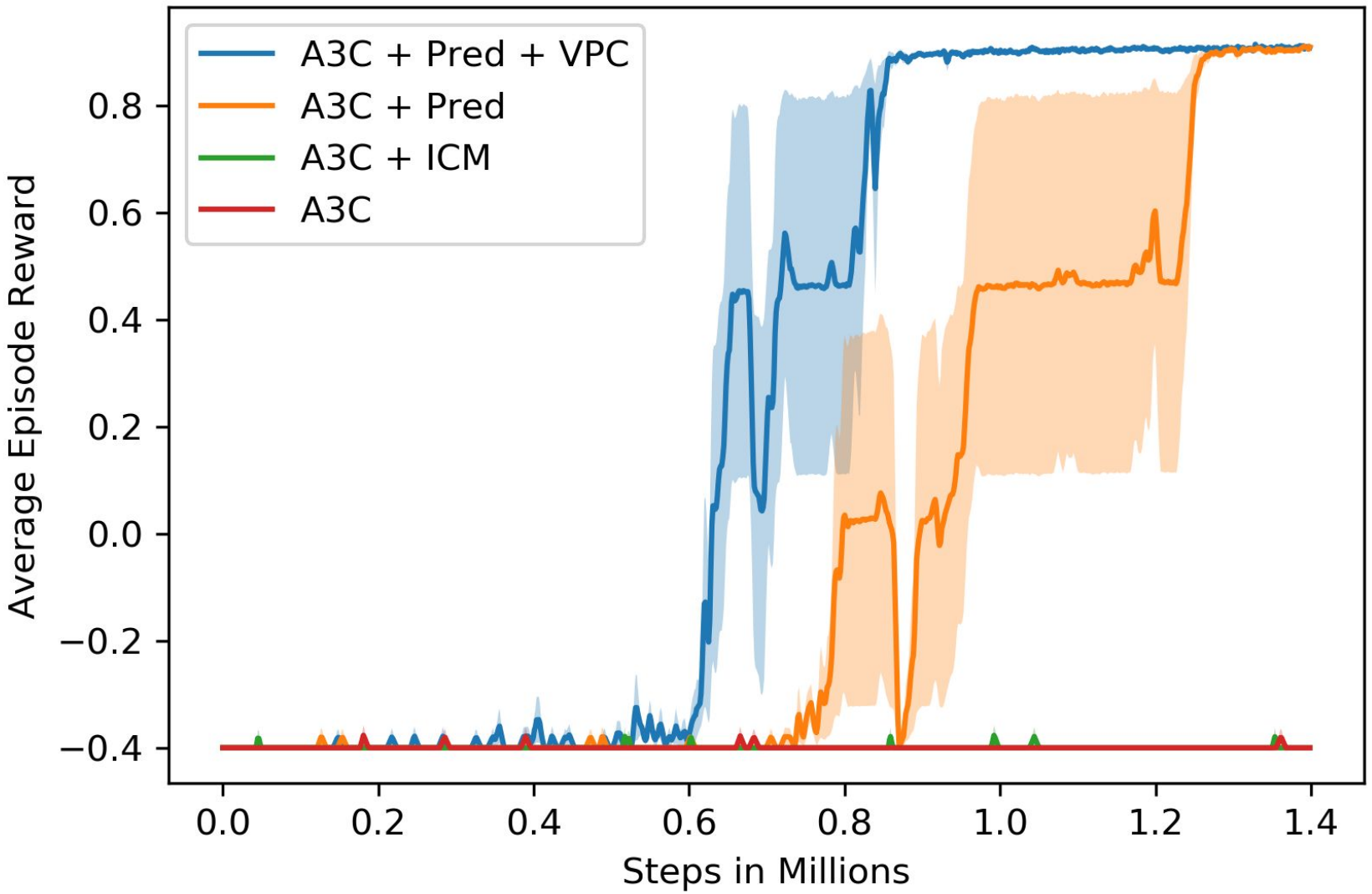
Rewards per episode



Rewards per episode

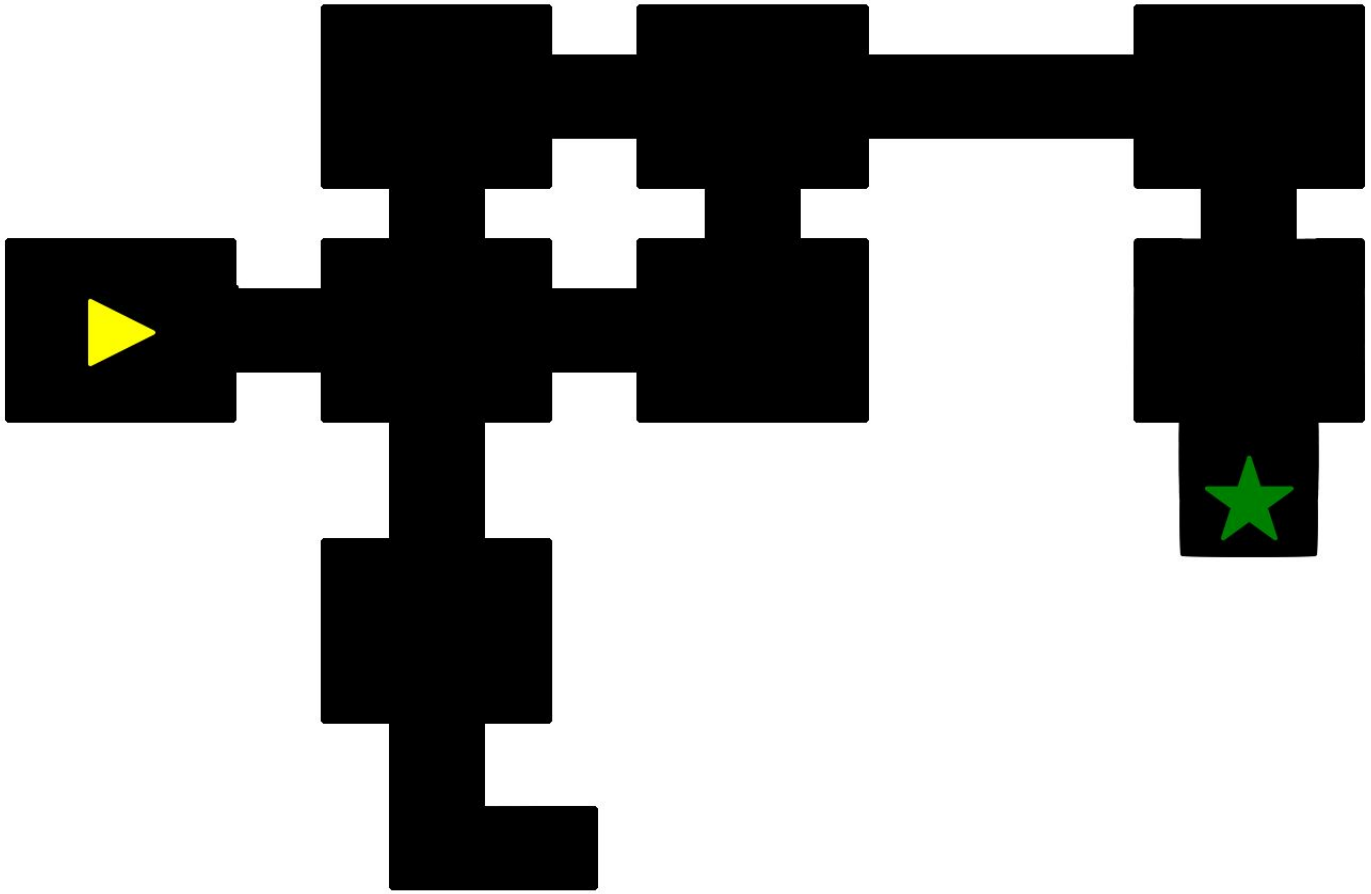


Rewards per episode

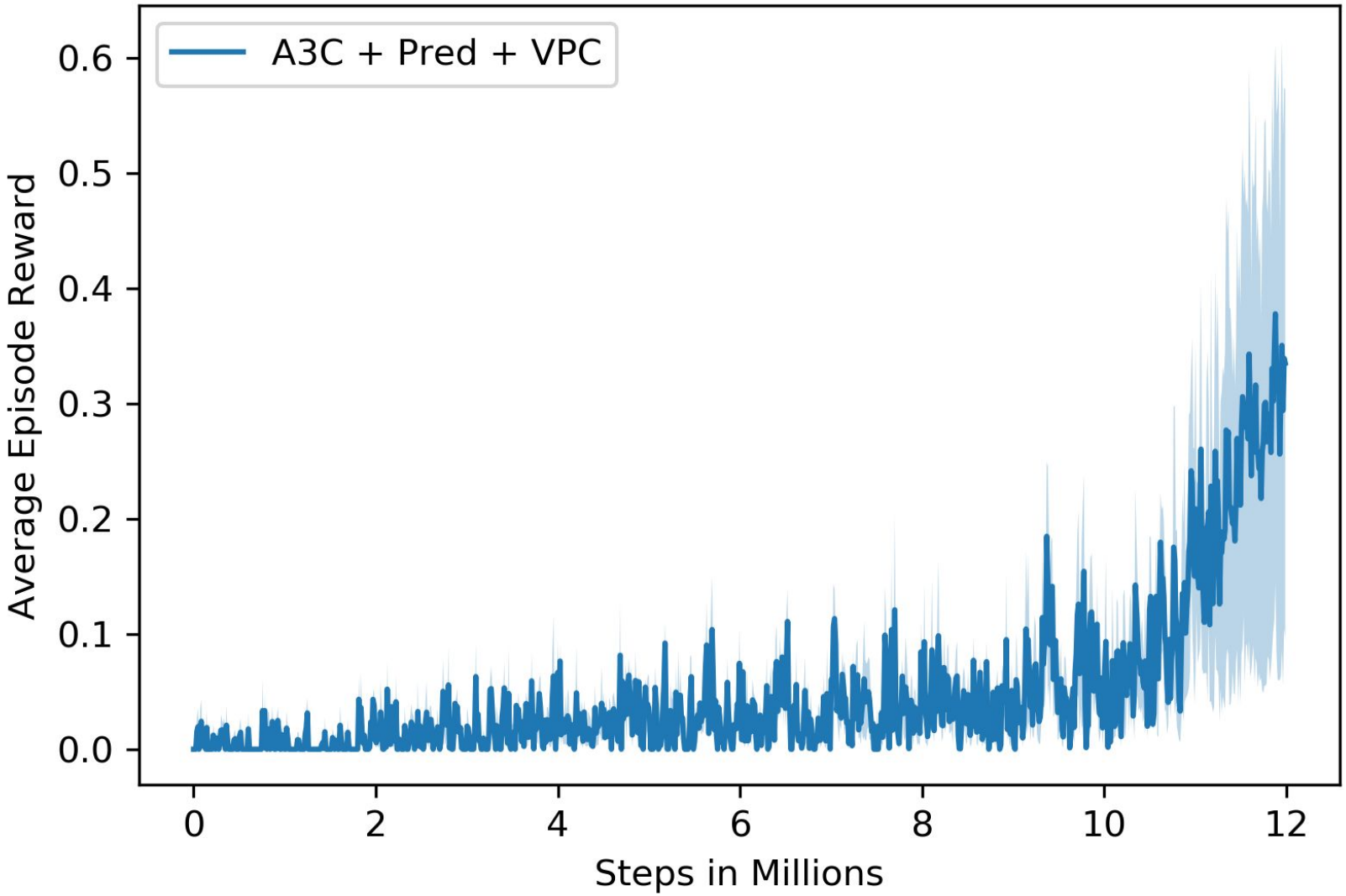


Doom environment

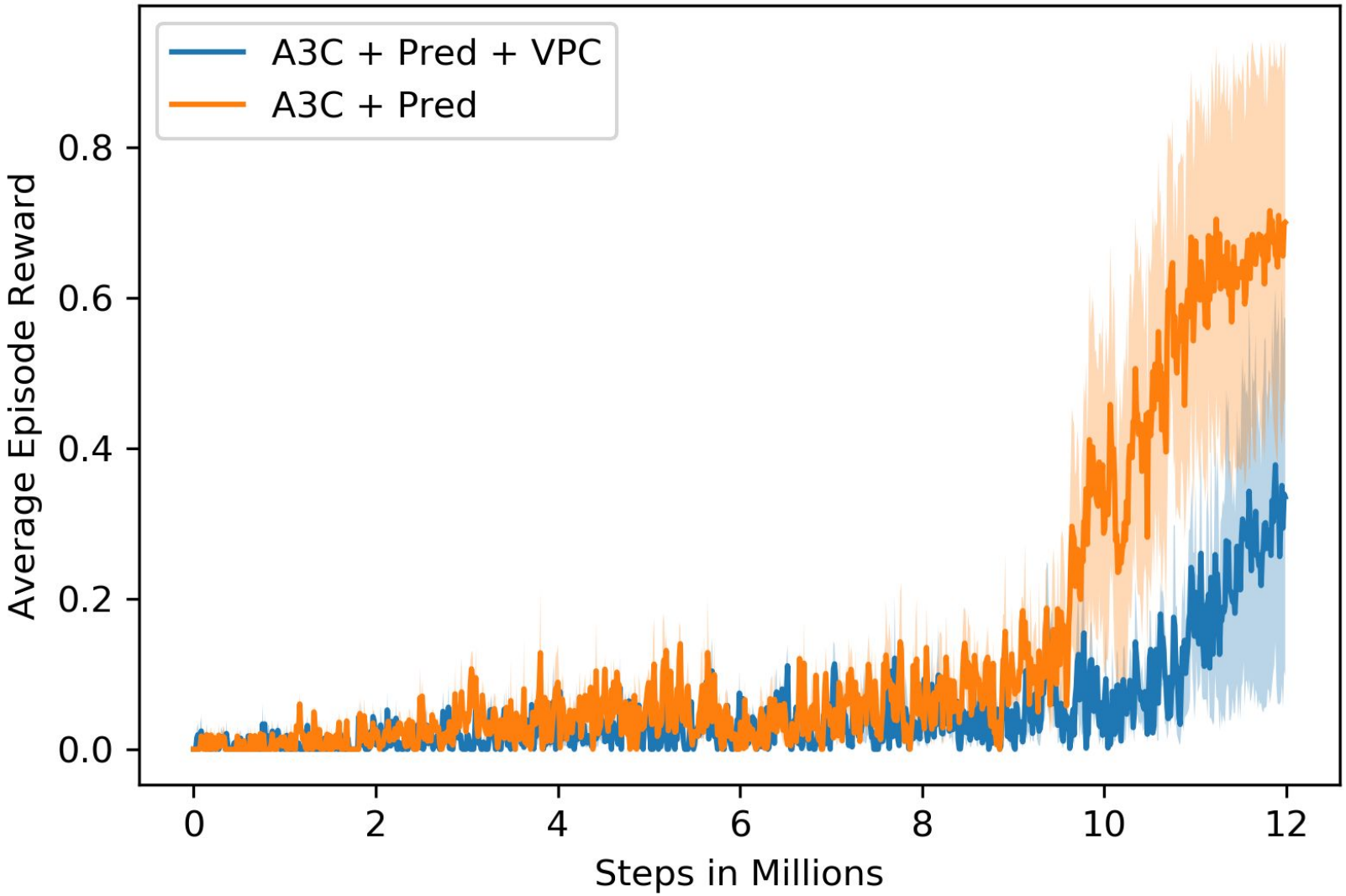
Doom Setup



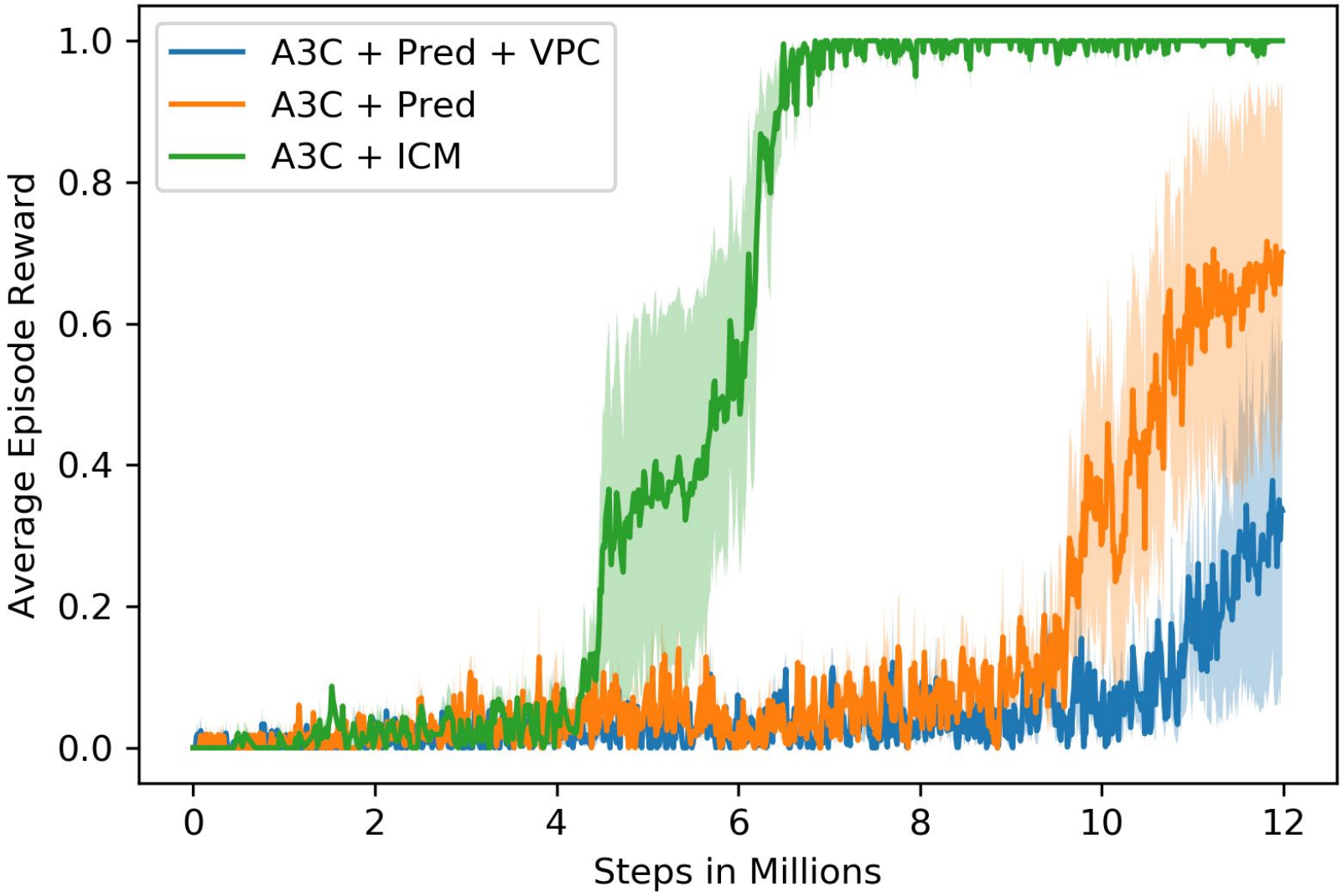
Rewards per episode



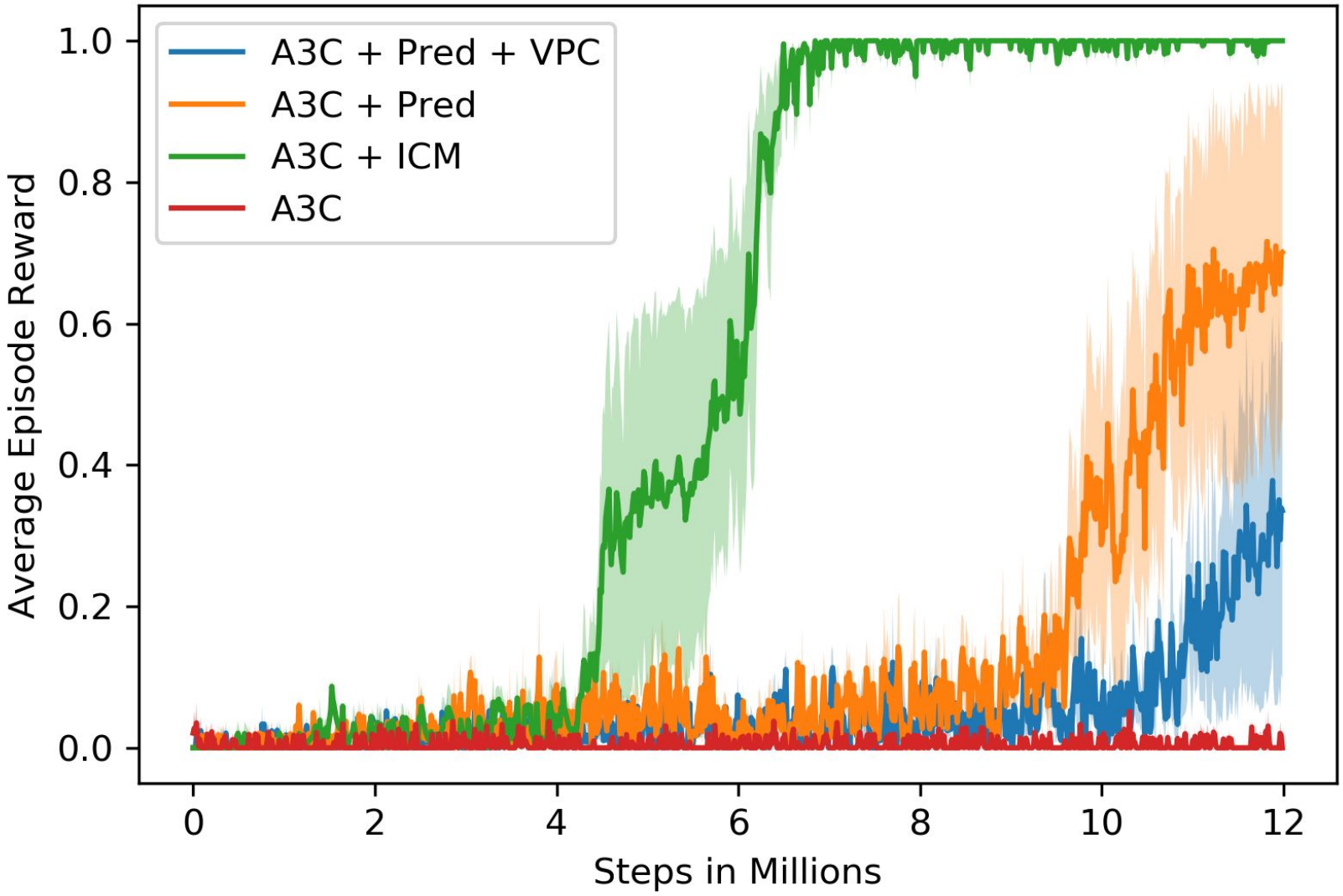
Rewards per episode



Rewards per episode



Rewards per episode



Question & Answers