EvIL😈: Evolution Strategies for Generalisable Imitation Learning

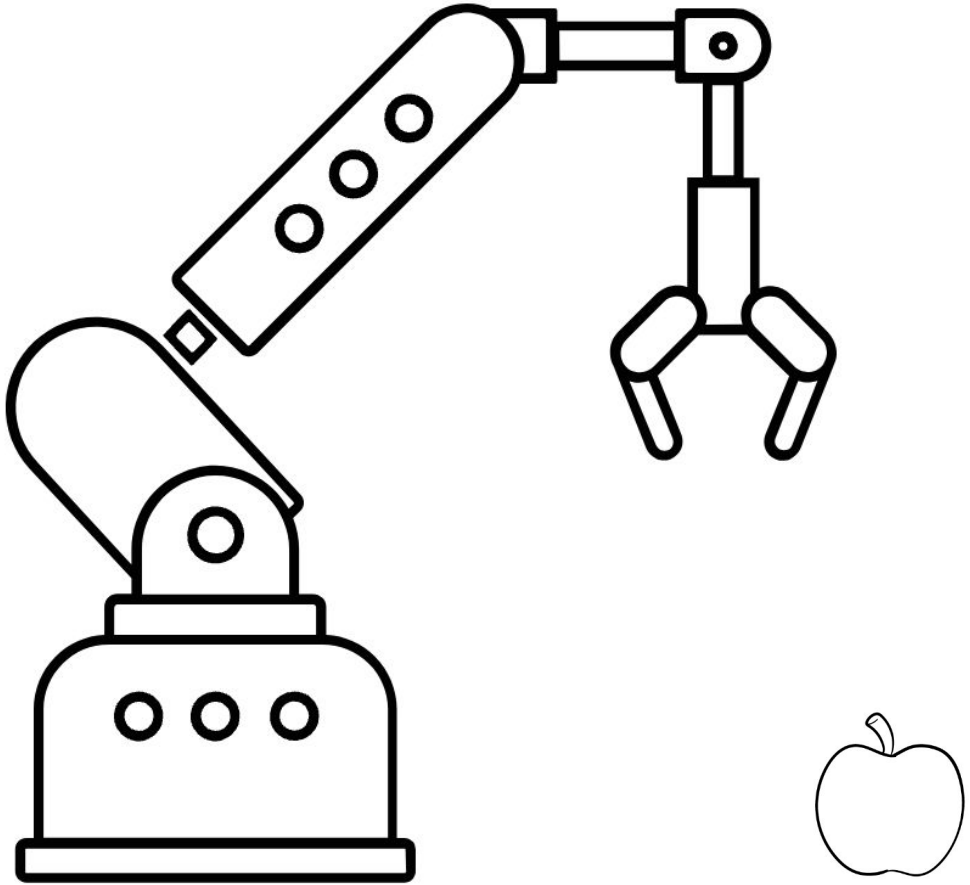
Silvia Sapora



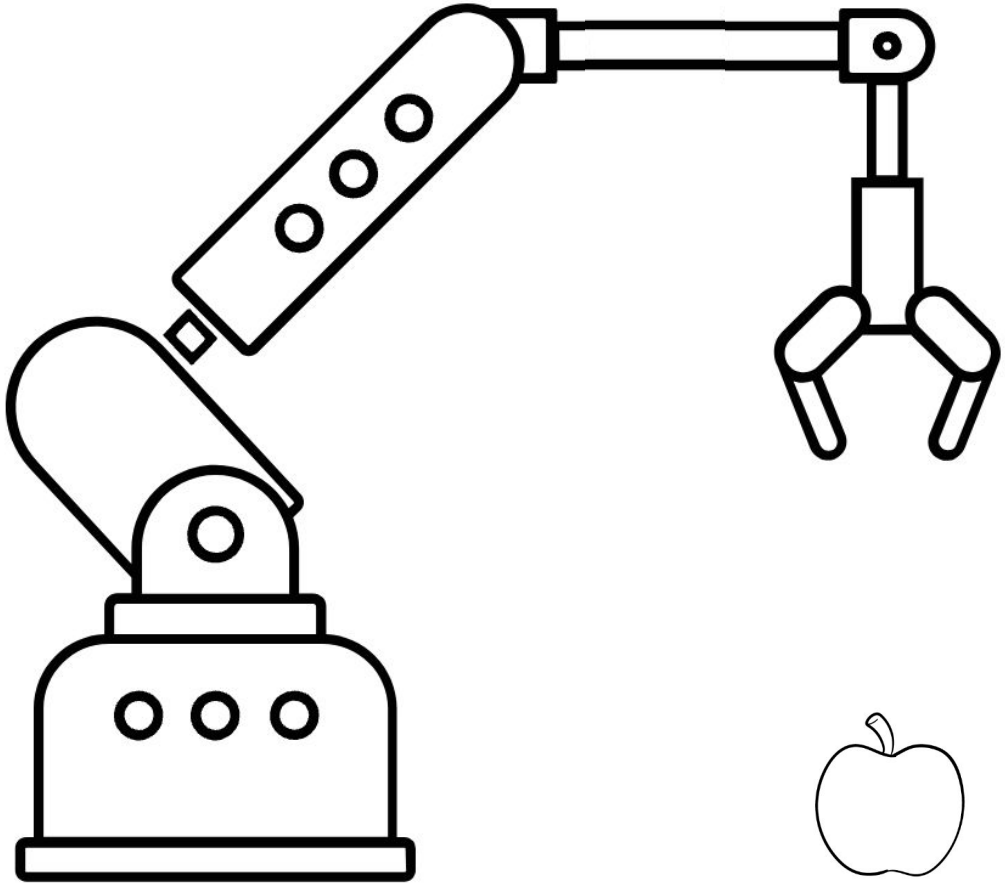


with Gokul Swamy, Chris Lu, Yee Whye Teh and Jakob Foerster

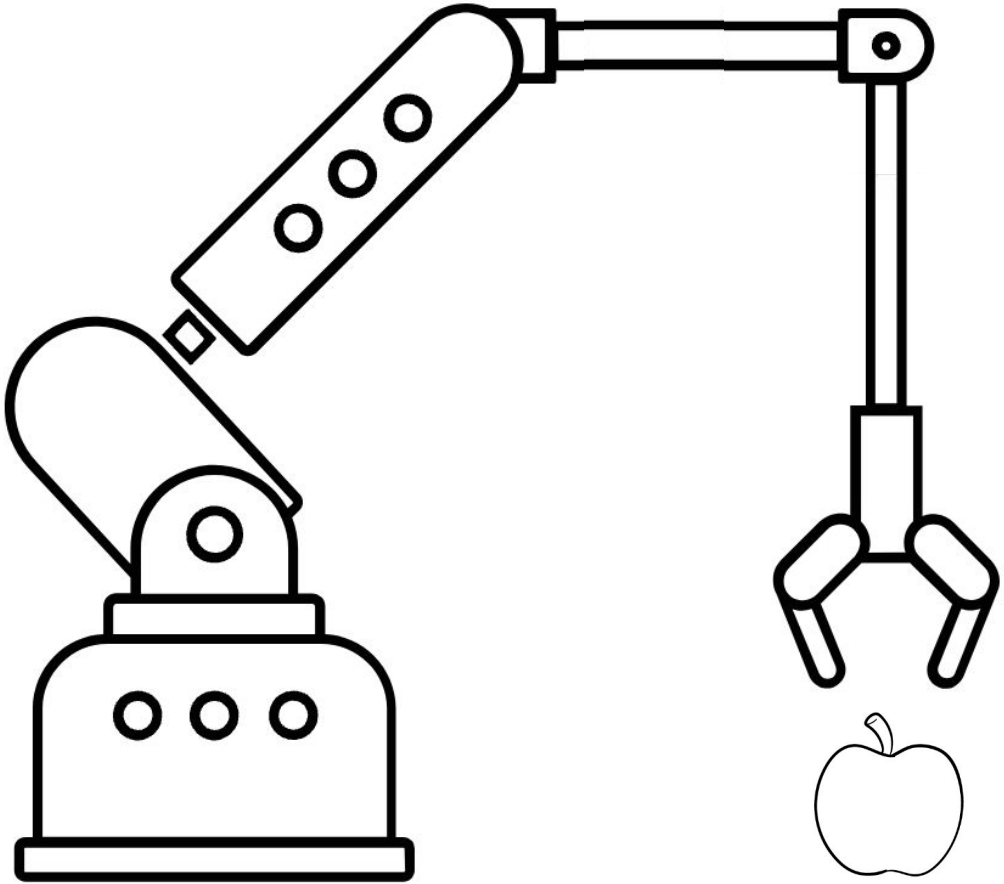
Background: Inverse Reinforcement Learning



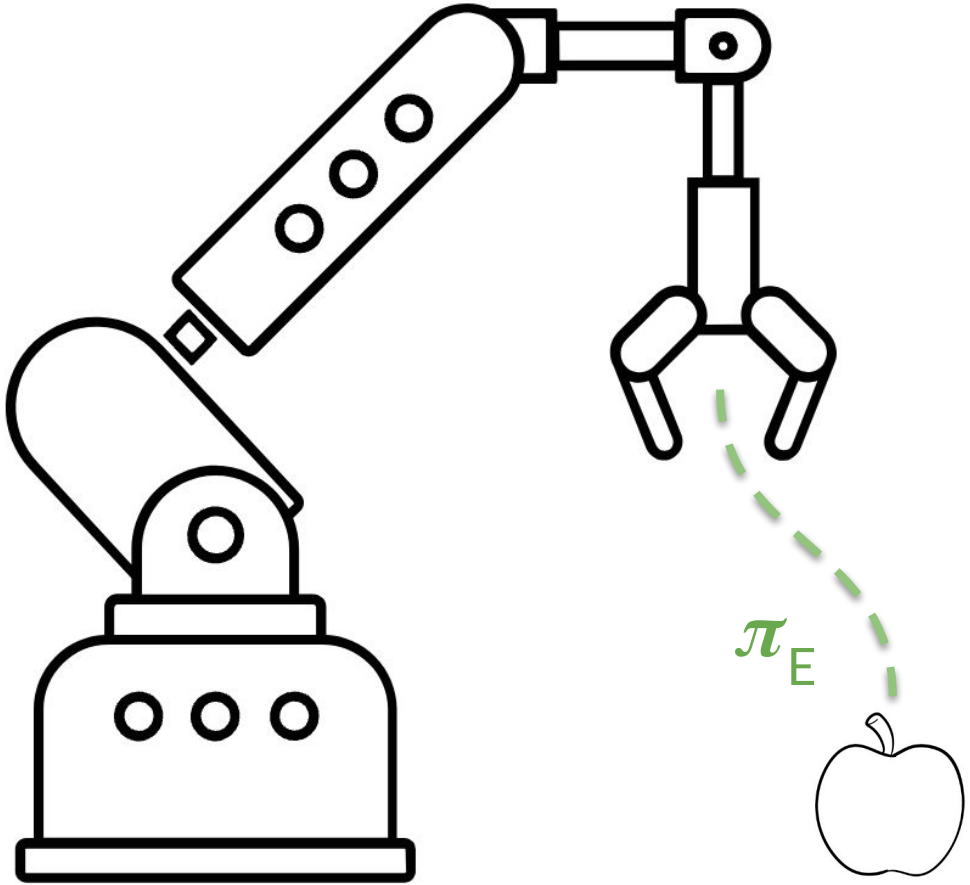
Background: Inverse Reinforcement Learning



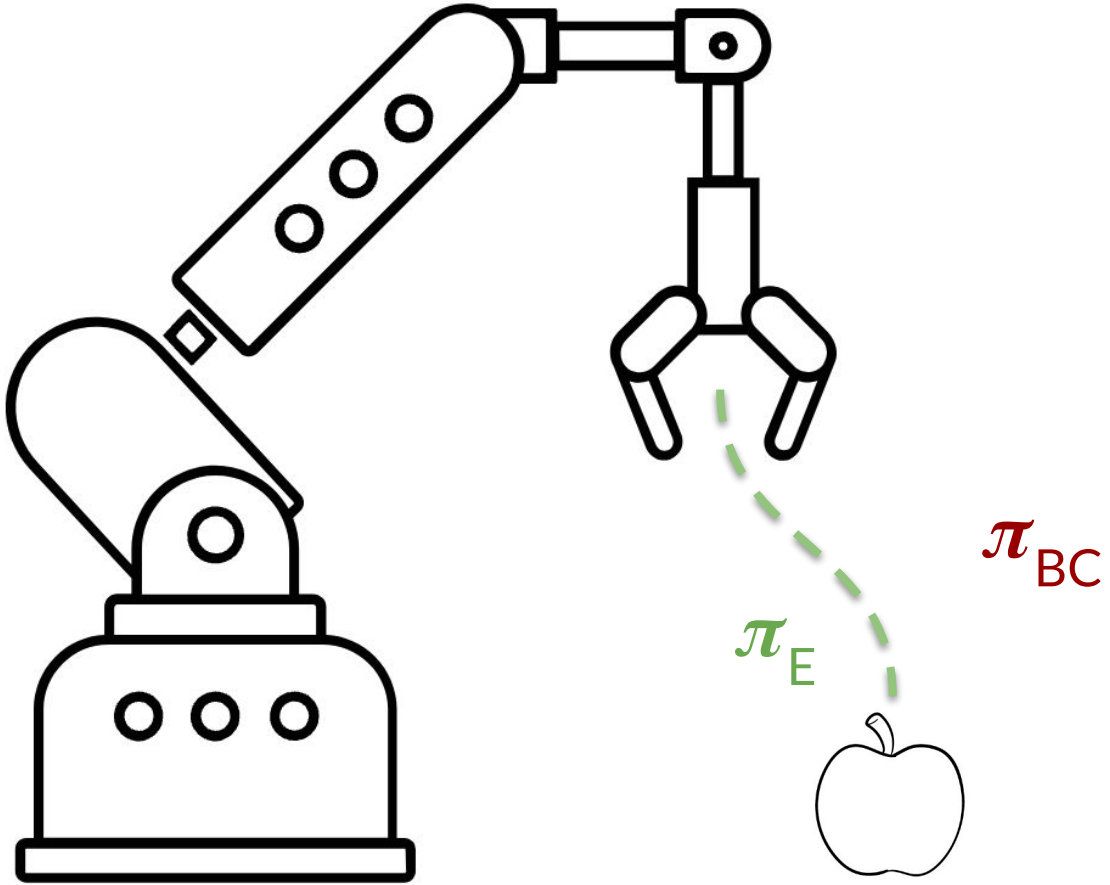
Background: Inverse Reinforcement Learning



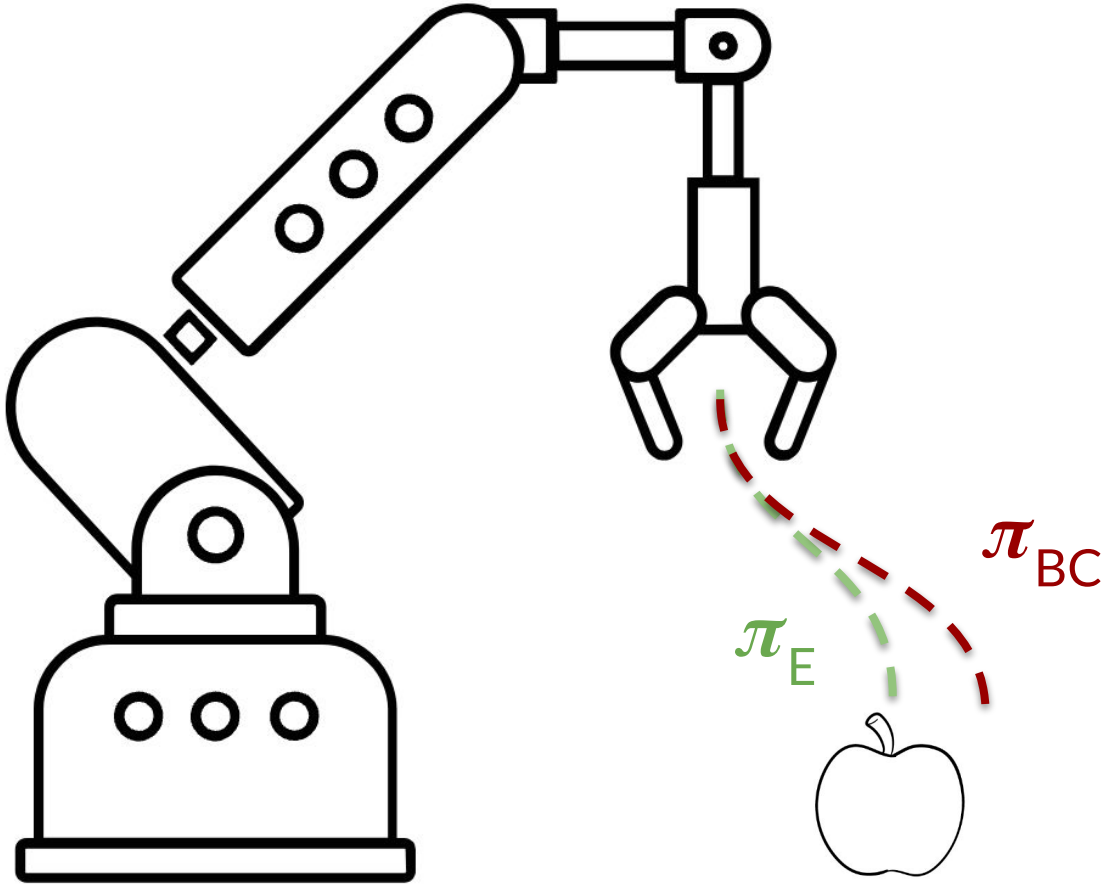
Background: Inverse Reinforcement Learning



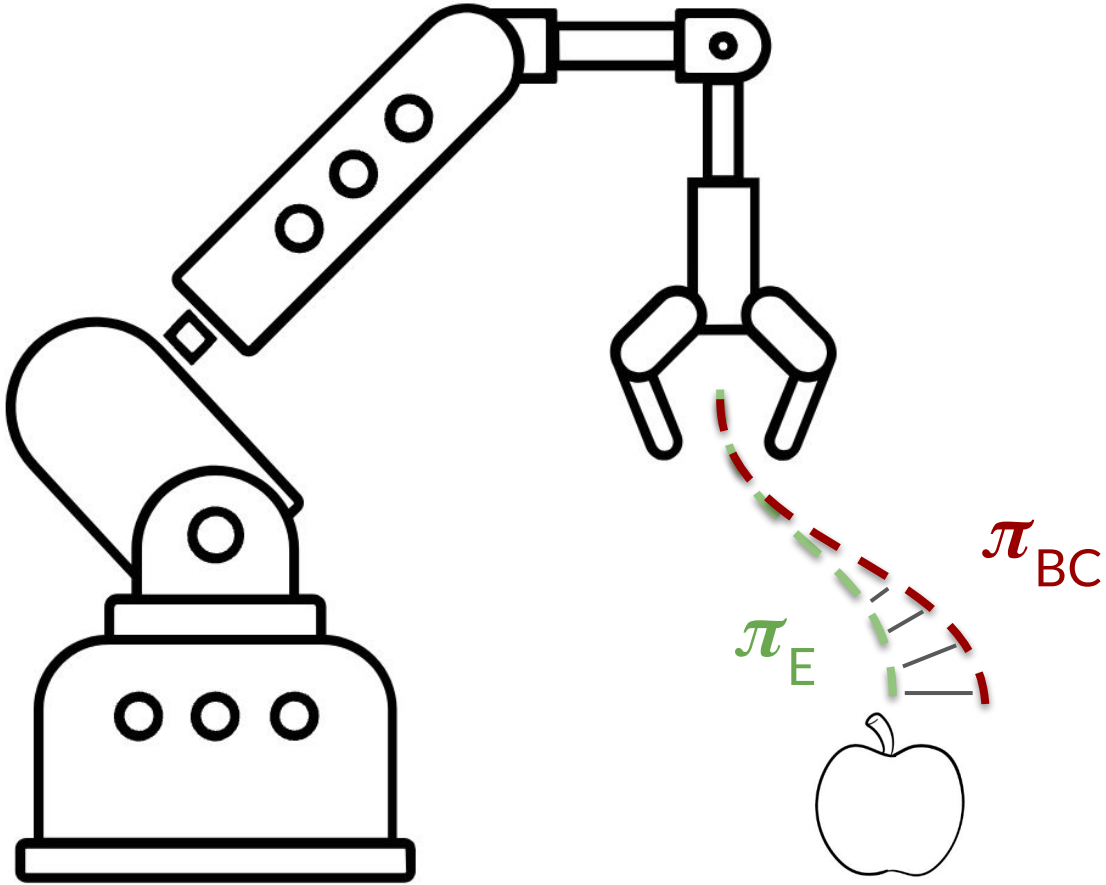
Background: Inverse Reinforcement Learning



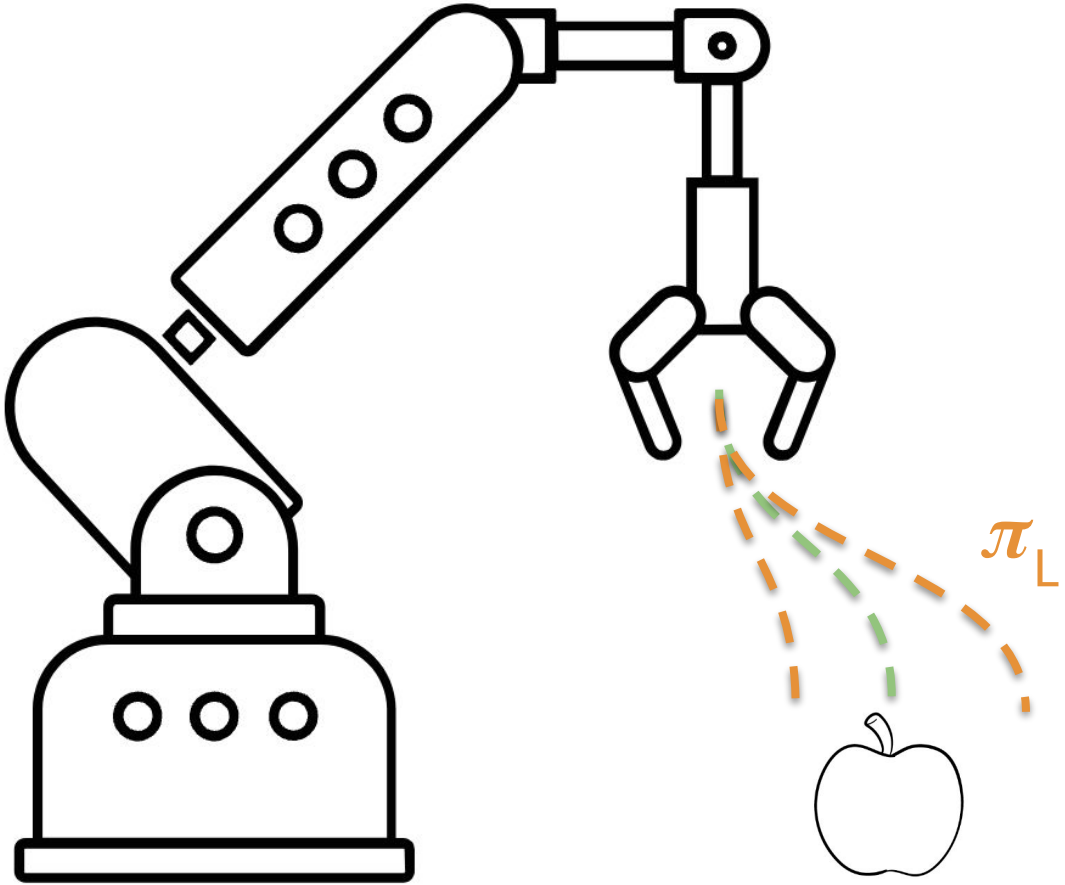
Background: Inverse Reinforcement Learning



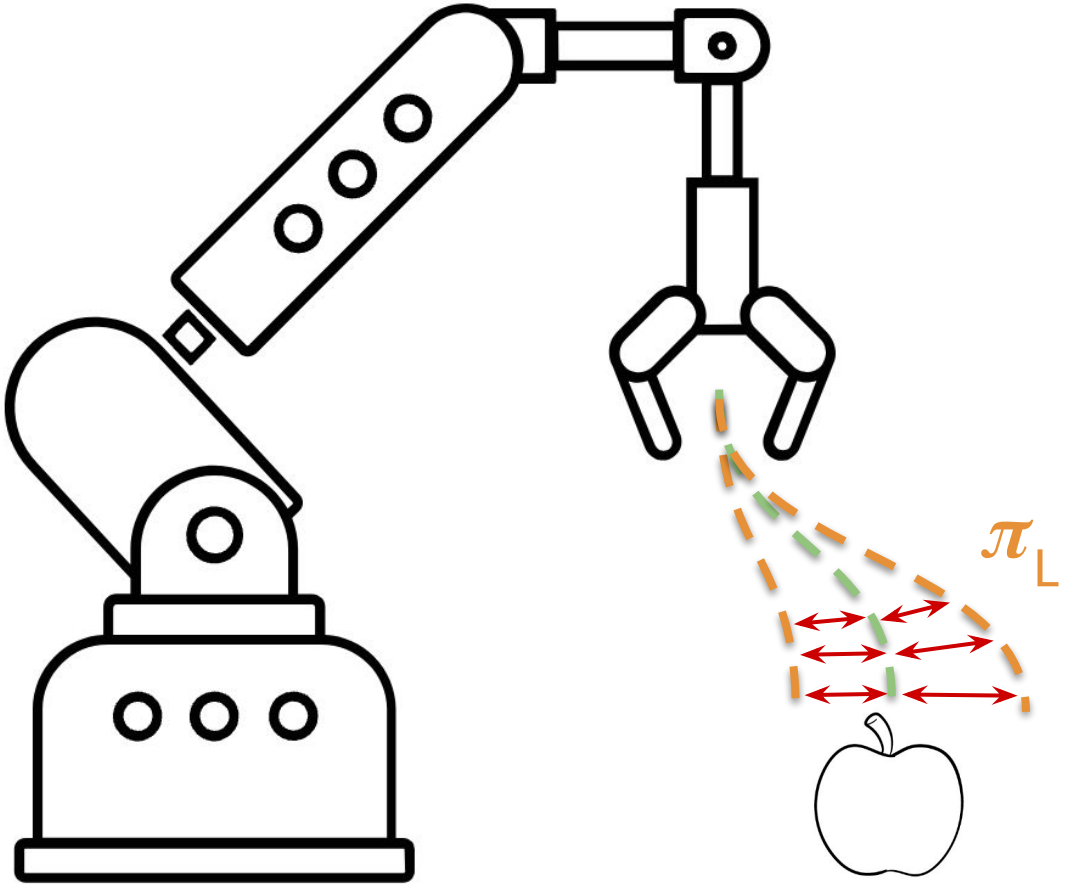
Background: Inverse Reinforcement Learning



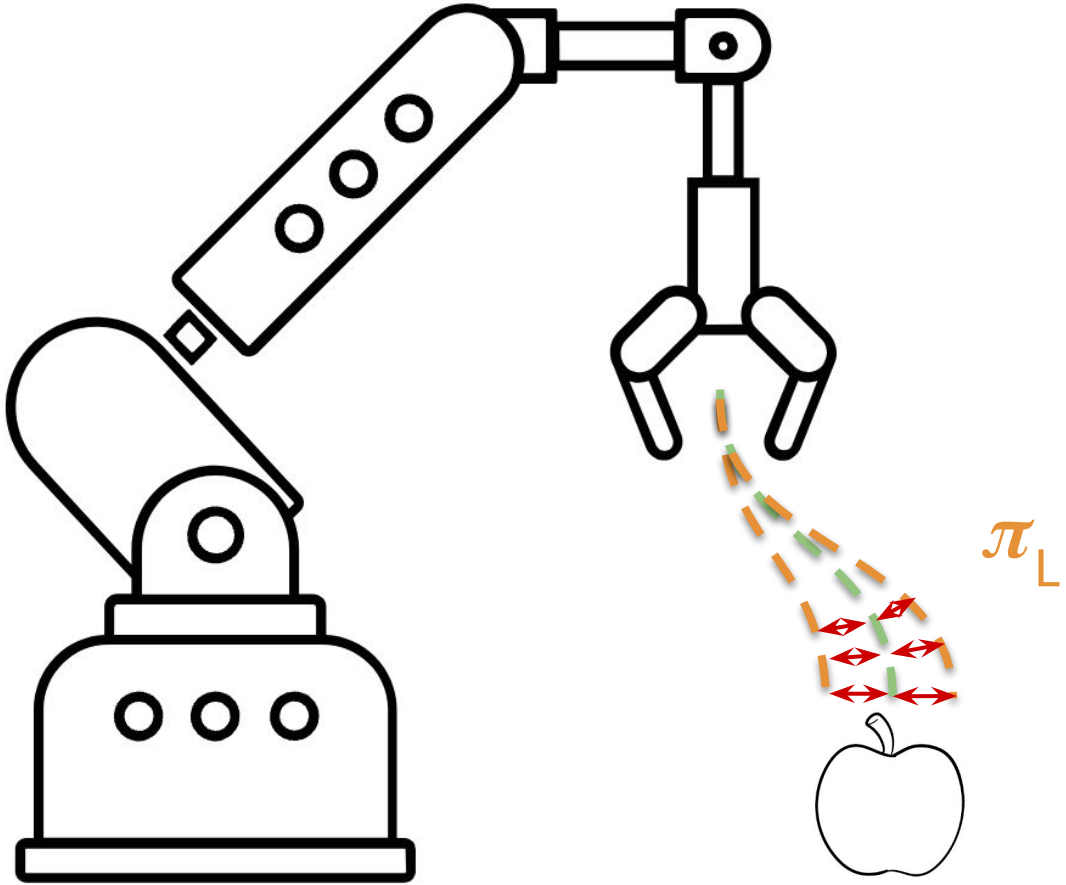
Background: Inverse Reinforcement Learning



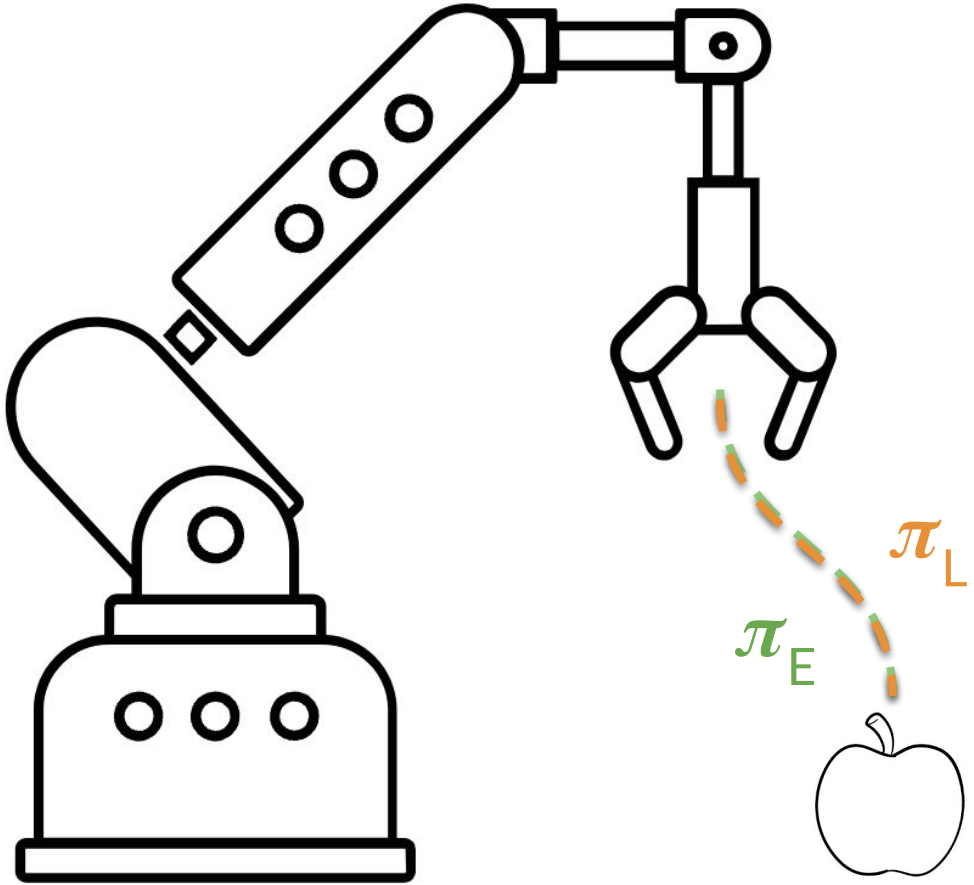
Background: Inverse Reinforcement Learning



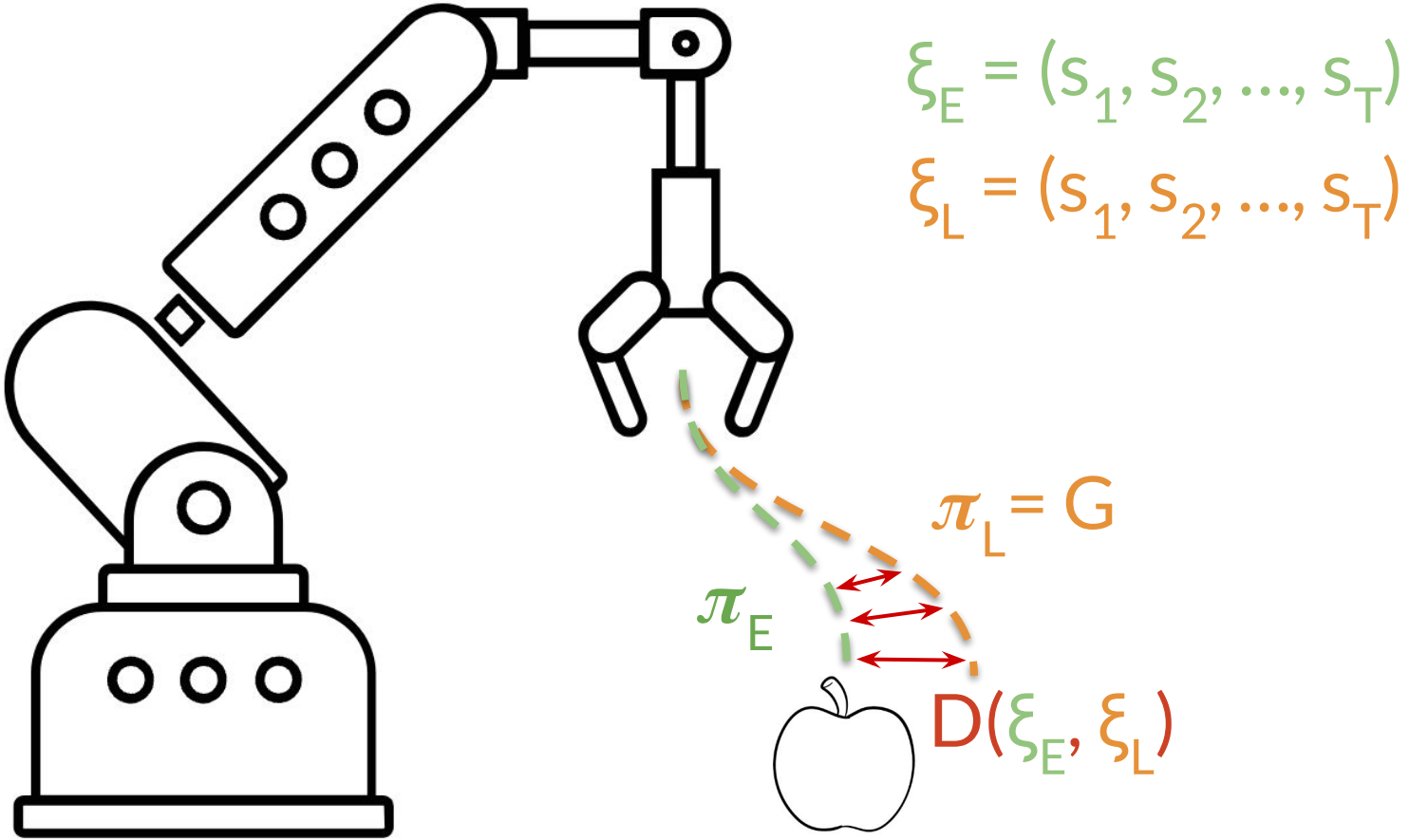
Background: Inverse Reinforcement Learning



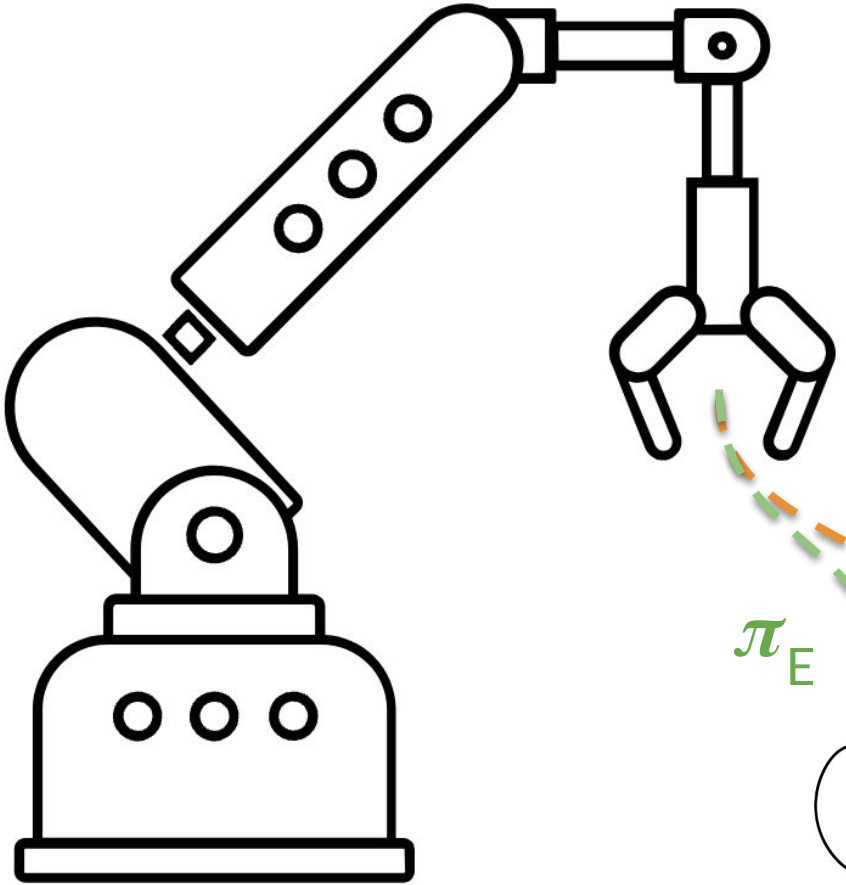
Background: Inverse Reinforcement Learning



Background: Inverse Reinforcement Learning



Background: Inverse Reinforcement Learning



$\xi_E = (s_1, s_2, \ldots, s_T)$

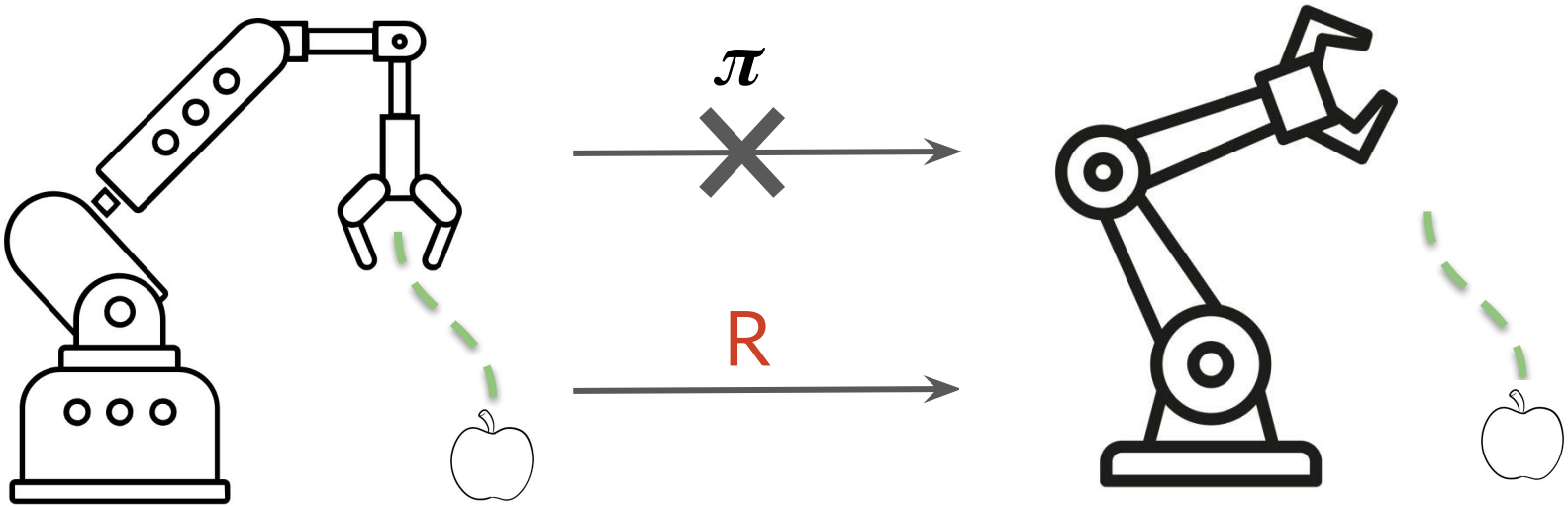
$\xi_L = (s_1, s_2, \ldots, s_T)$

$\pi_L = G$

π_E

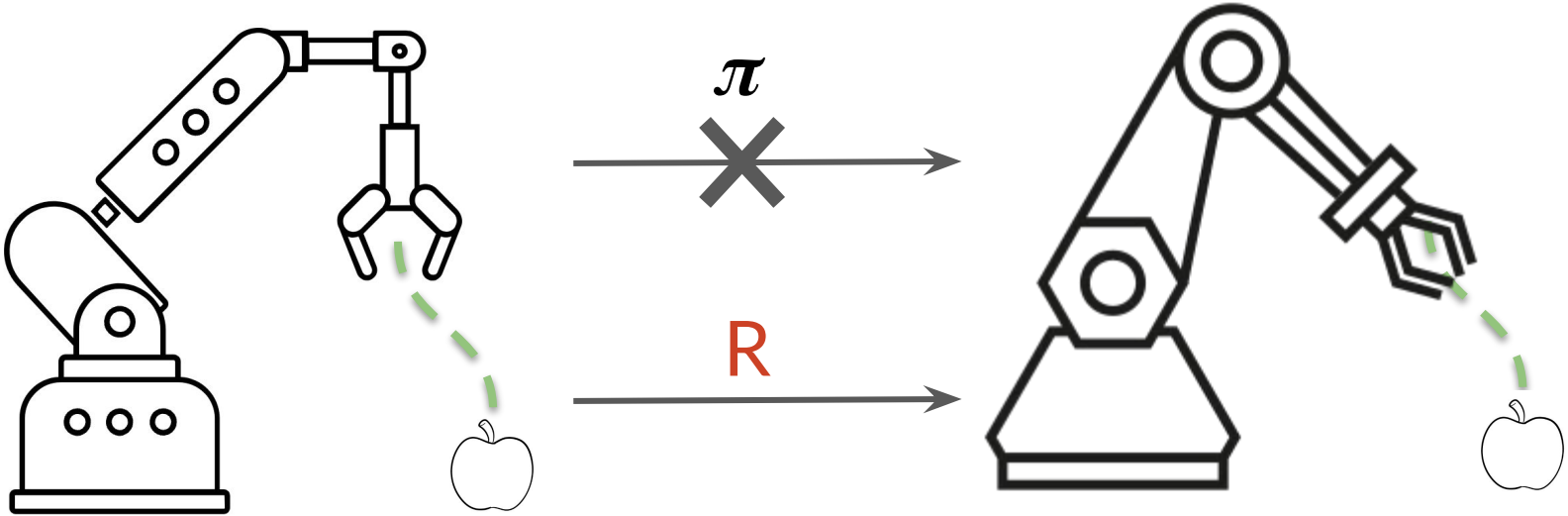
$D(\xi_E, \xi_L) = R$

Background: Inverse Reinforcement Learning

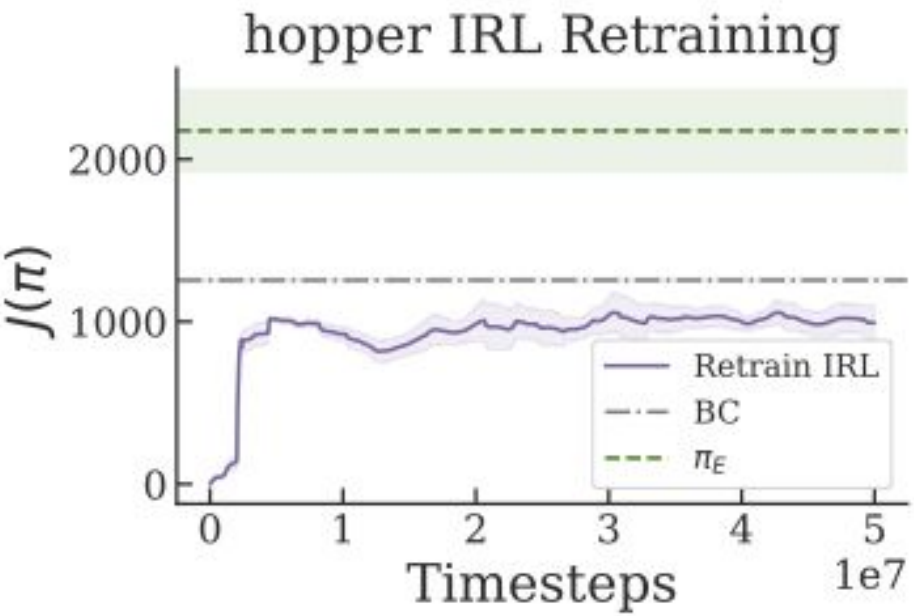


Of Moments and Matching: A Game-Theoretic Framework for Closing the Imitation Gap, Swamy et al. 2021

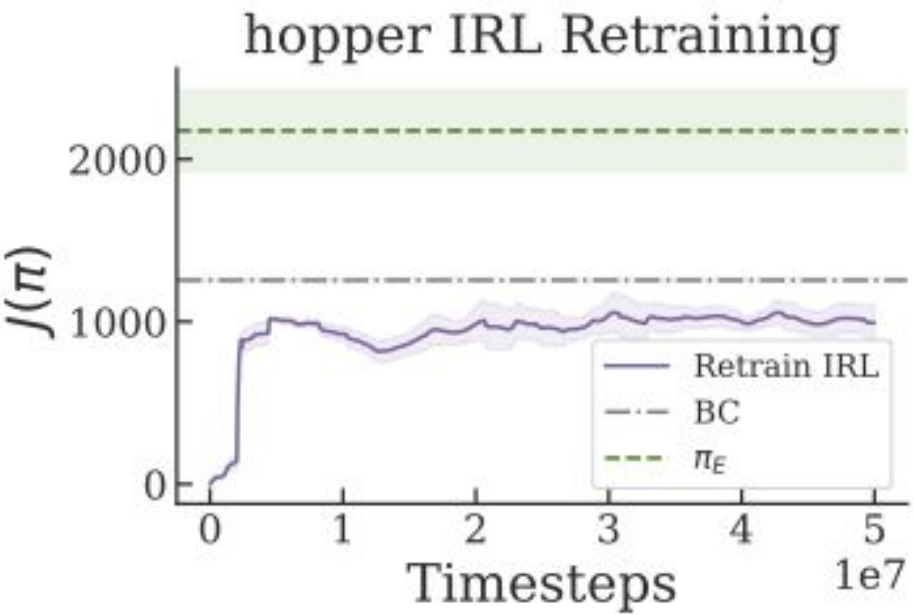
Background: Inverse Reinforcement Learning



🔑 Problems: Inverse Reinforcement Learning

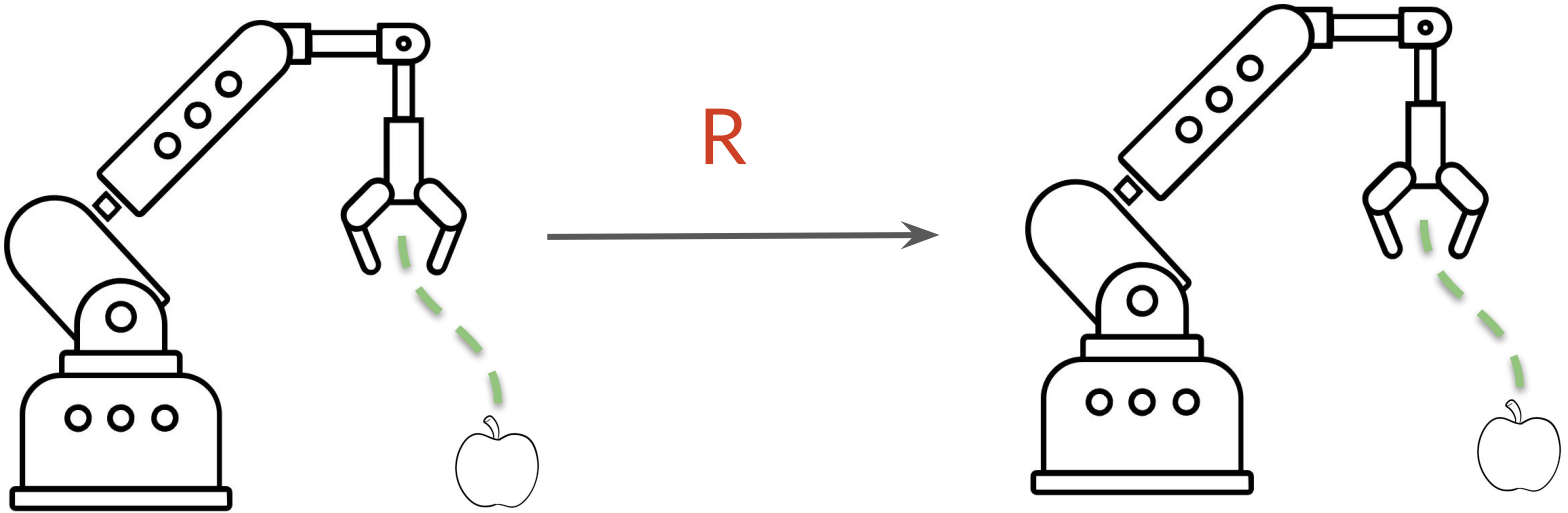


🔑 Problems: Inverse Reinforcement Learning

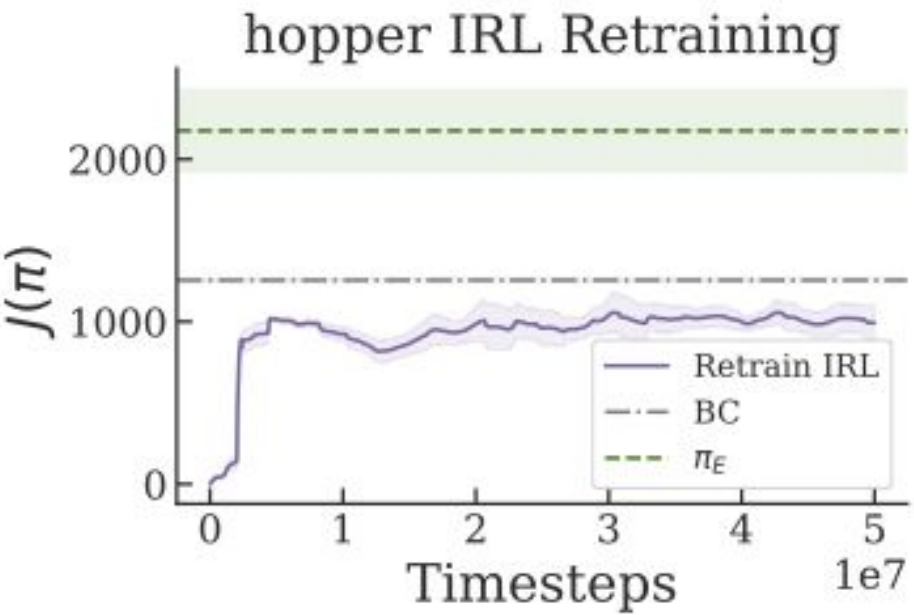


- Not **effective**

🔑 Problems: Inverse Reinforcement Learning



🔑 Problems: Inverse Reinforcement Learning



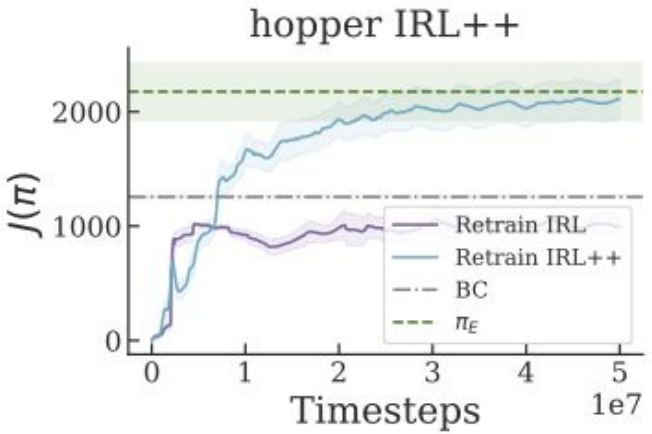
- Not **effective**
- Not **efficient**

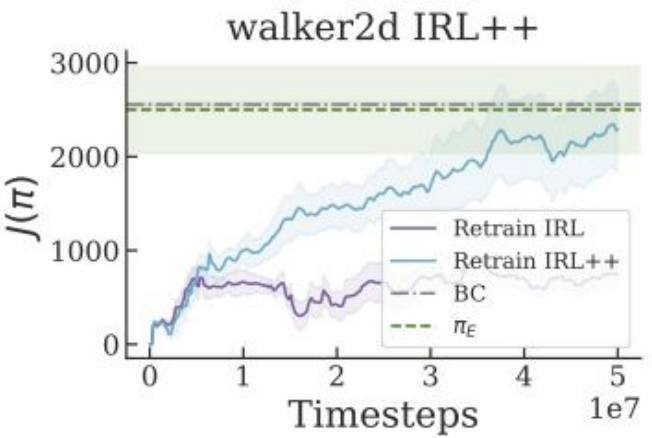
🔑 Idea: IRL++ for Effective Retraining

- **Policy Buffer**
- **Discriminator and Policy Ensembles**
- **Random Policy Resets**

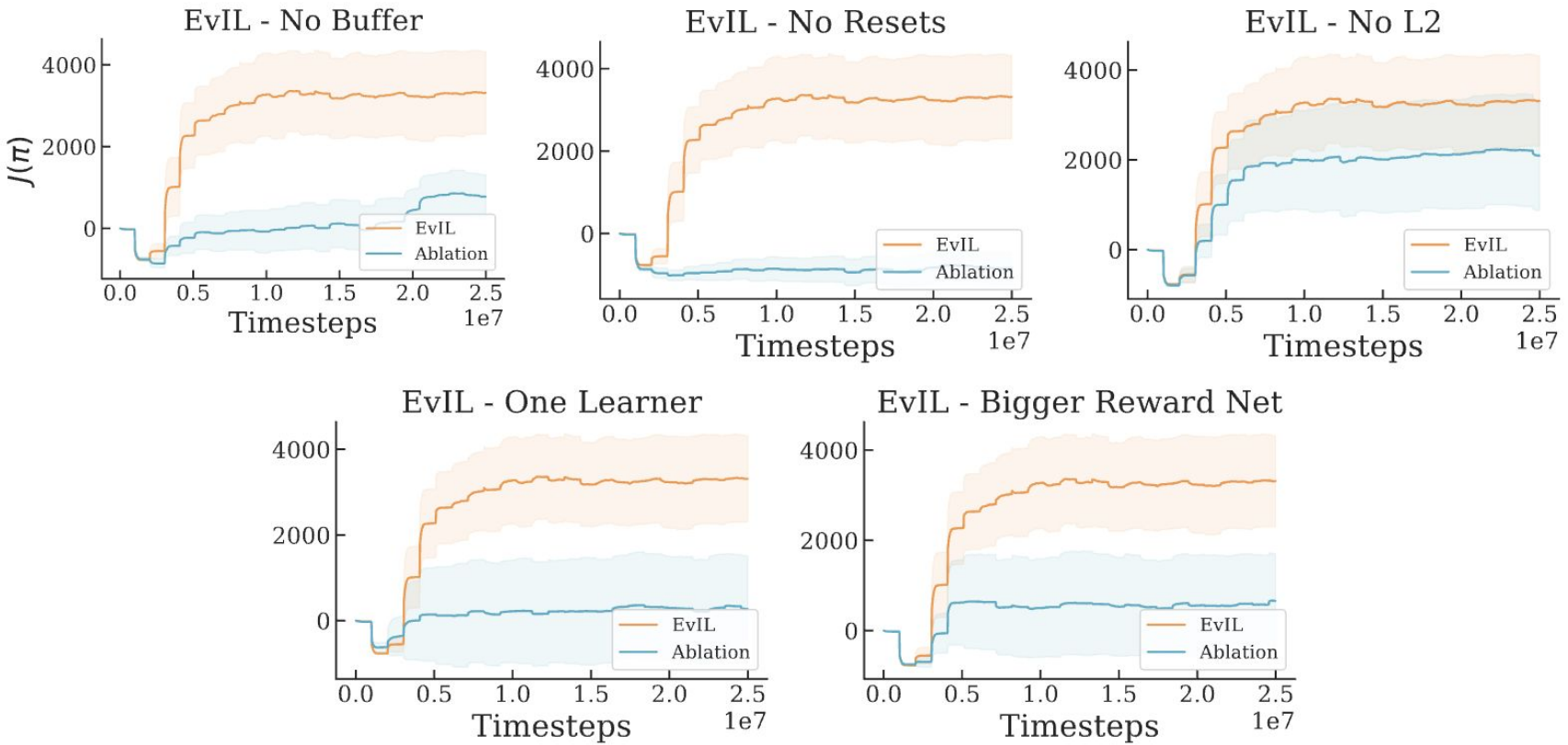
🔑 Idea: IRL++ for Effective Retraining

- **Policy Buffer**
- **Discriminator and Policy Ensembles**
- **Random Policy Resets**

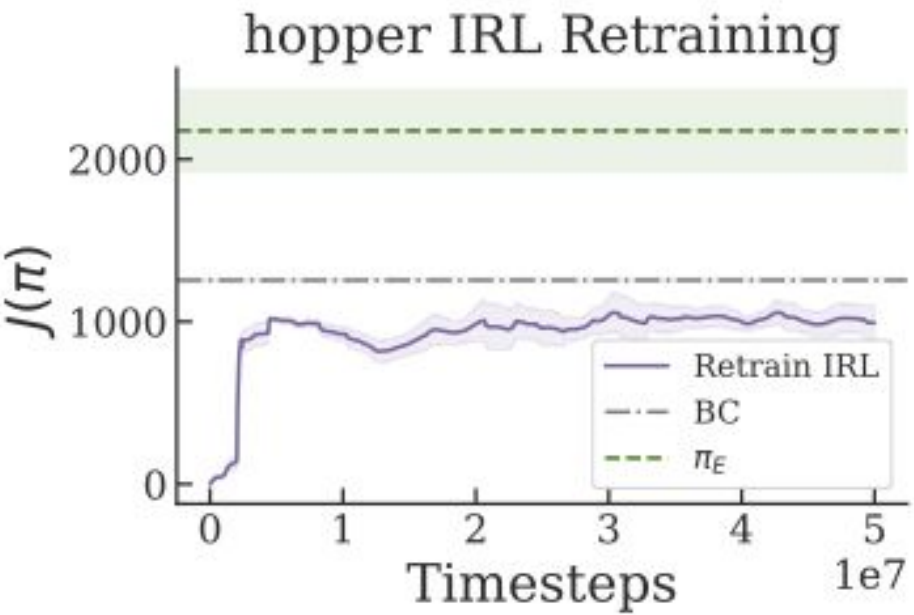




🔑 Idea: IRL++ for Effective Retraining

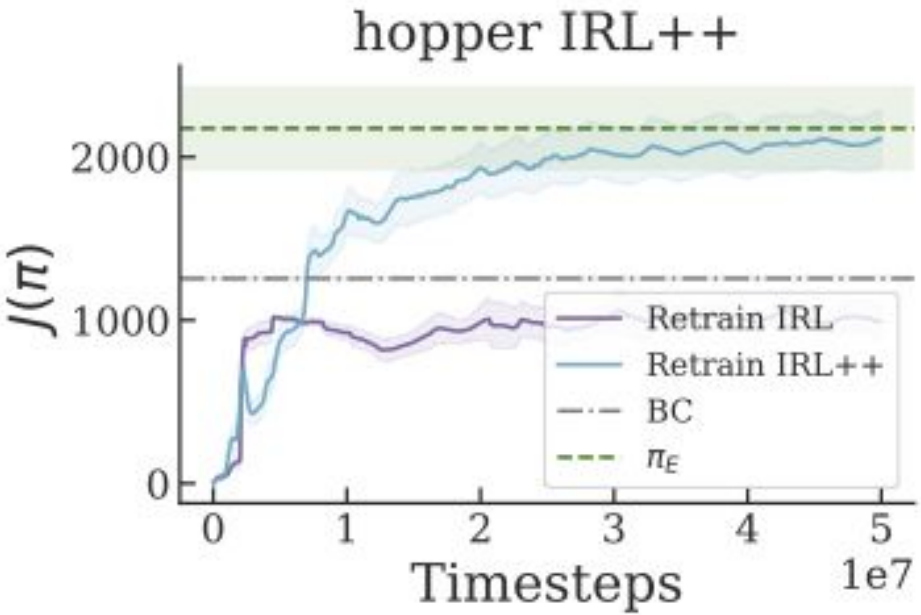


🔑 Problems: Inverse Reinforcement Learning



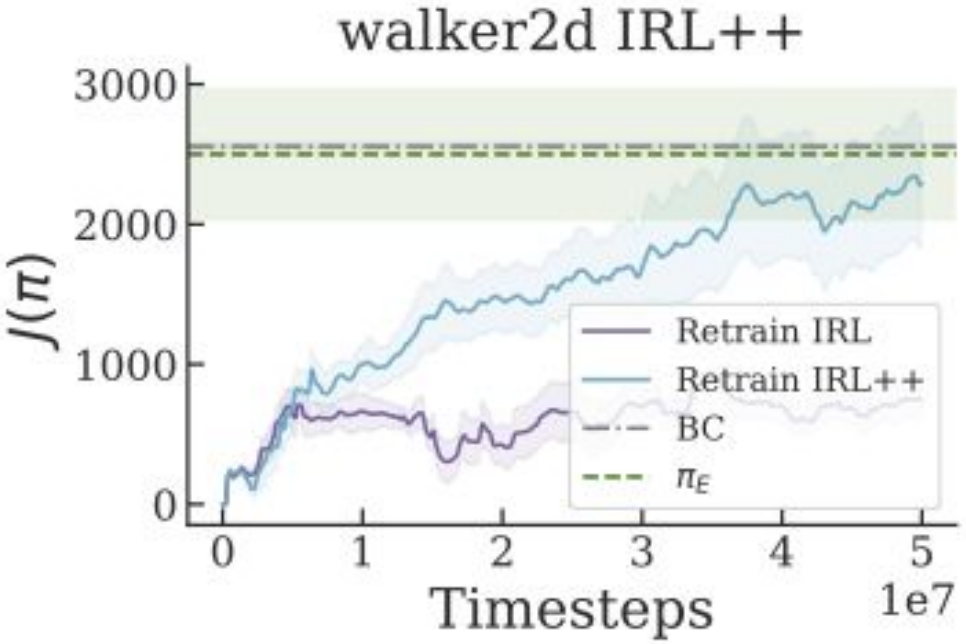
- Not **effective**
- Not **efficient**

🔑 Problems: Inverse Reinforcement Learning

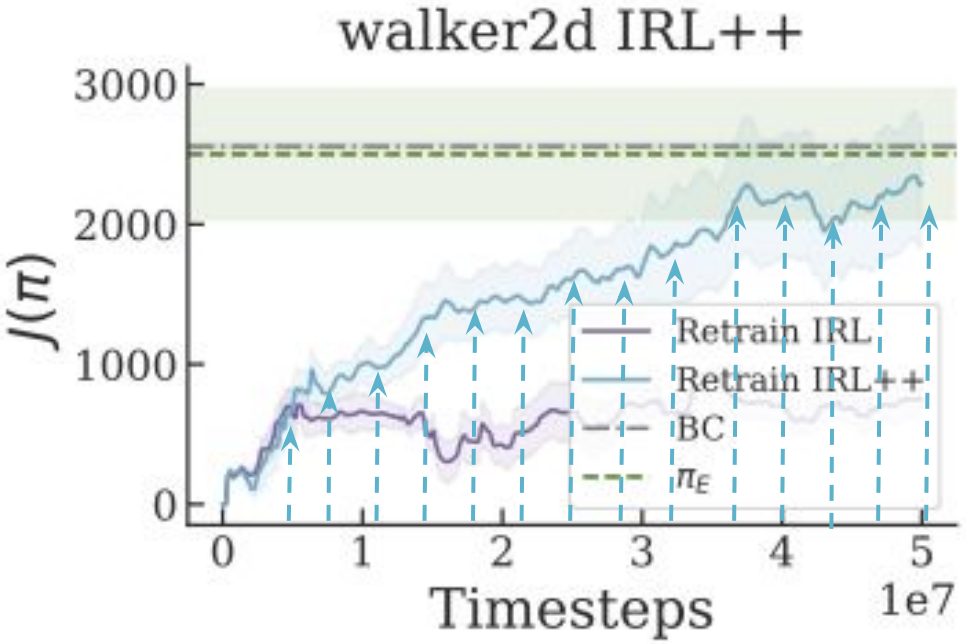


- ~~Not **effective**~~
- Not **efficient**

🔑 Idea: Reward Shaping



🔑 Idea: Reward Shaping



Background: Potential-Based Reward Shaping

$$\Phi : \mathcal{S} \rightarrow \mathbb{R} \qquad s \in \mathcal{S}$$

Policy Invariance under Reward Transformations: Theory and Applications of Reward Shaping, Ng et al. 1999

Background: Potential-Based Reward Shaping

$$\Phi : \mathcal{S} \rightarrow \mathbb{R} \qquad s \in \mathcal{S}$$

$$r'(s,a) = r(s,a) + F_{\Phi}(s,s')$$
$$F_{\Phi}(s,s') = \Phi(s') - \Phi(s)$$

Policy Invariance under Reward Transformations: Theory and Applications of Reward Shaping, Ng et al. 1999

Background: Potential-Based Reward Shaping

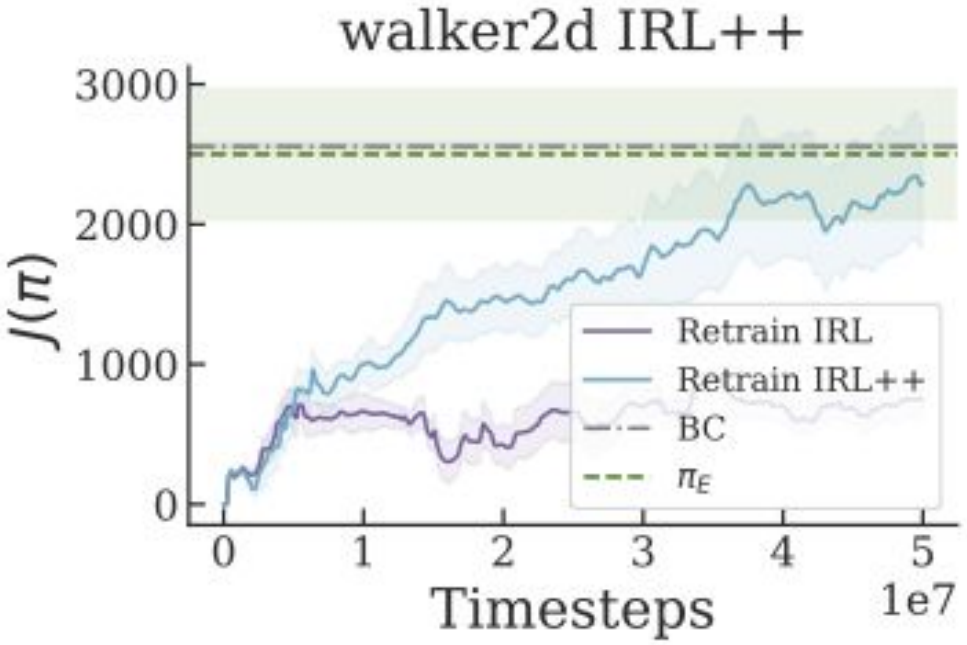
$$r'(s,a) = r(s,a) + F_\Phi(s,s')$$
$$F_\Phi(s,s') = \Phi(s') - \Phi(s)$$

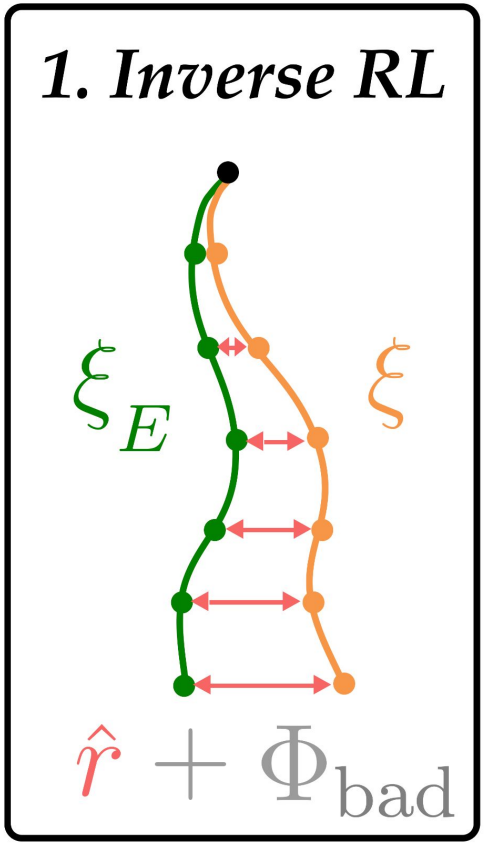
$$\Phi(s) = V^\star(s)$$

$$r'(s,a) = r(s,a) + V^\star(s') - V^\star(s) = A^\star(s,a)$$

Policy Invariance under Reward Transformations: Theory and Applications of Reward Shaping, Ng et al. 1999

Background: Potential-Based Reward Shaping





Background: Evolution Strategies

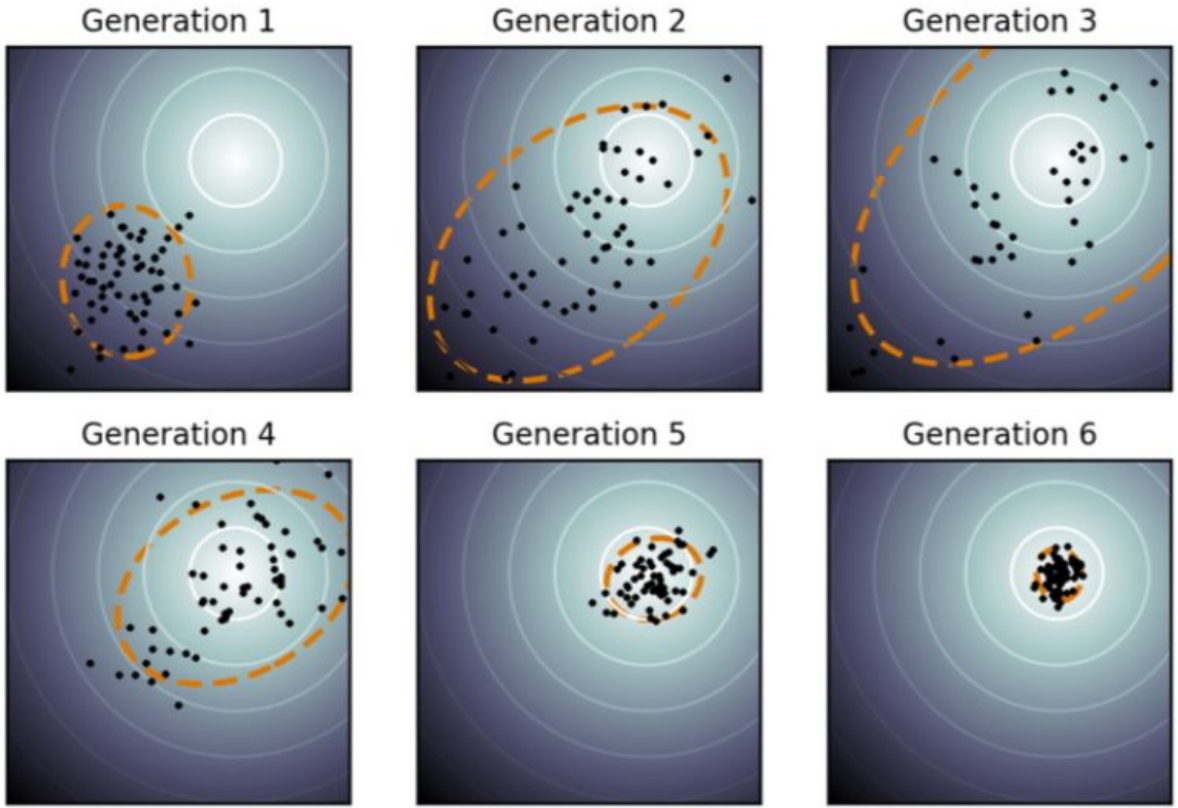
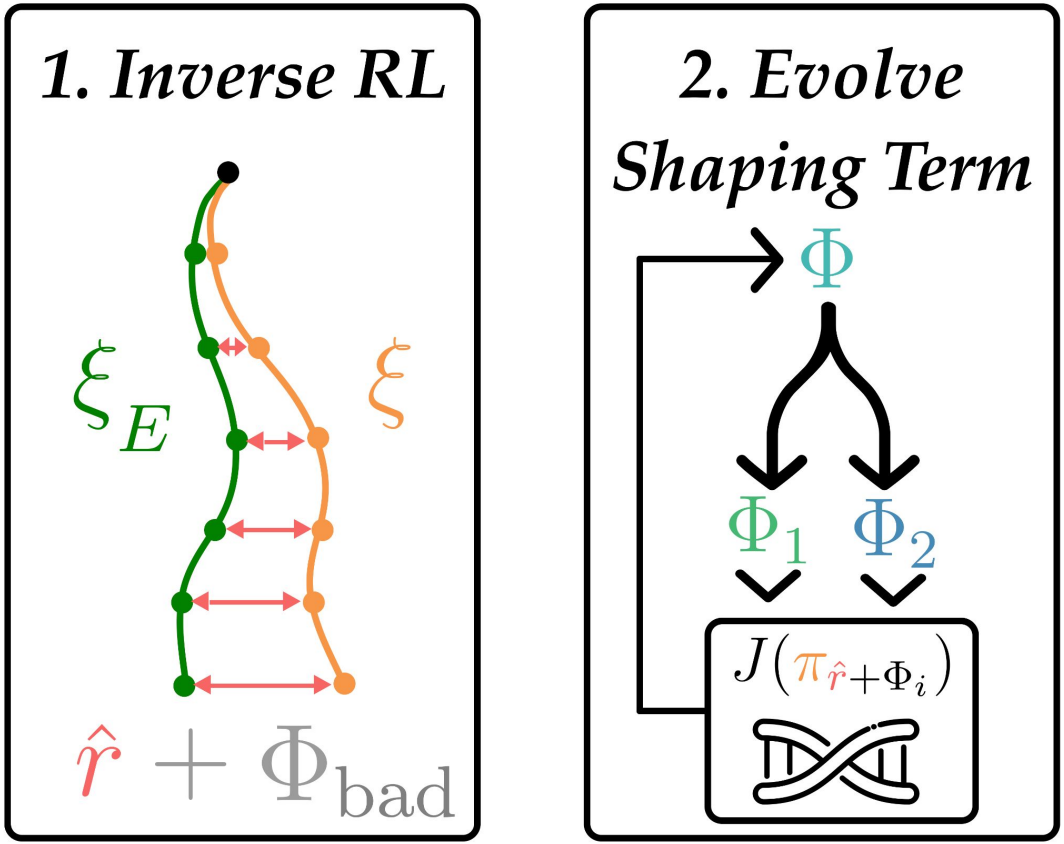
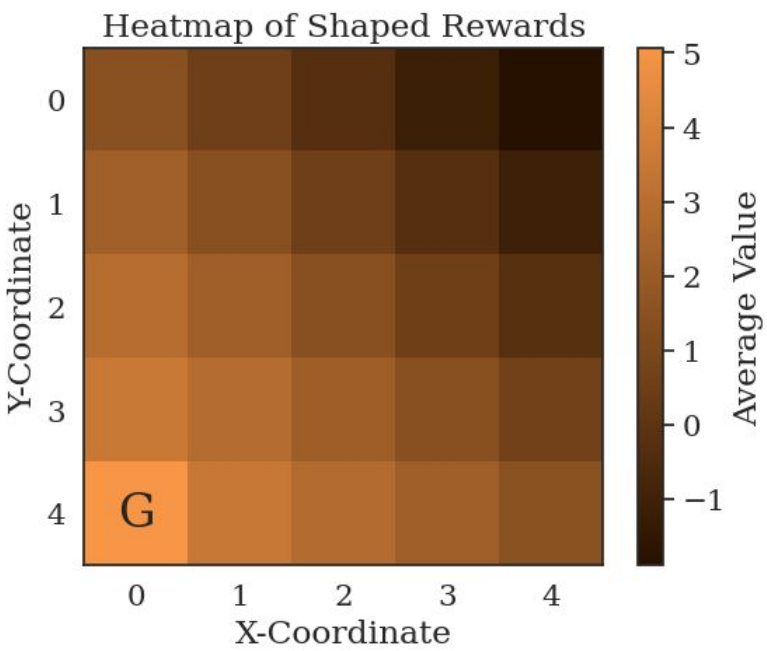


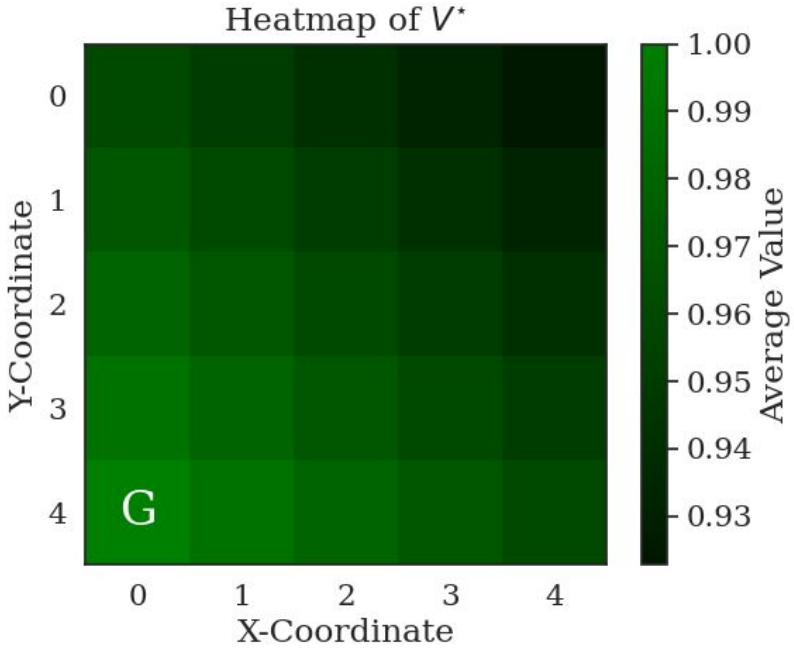
Image Source: Wikipedia CMA-ES

🔑 Idea: Reward Shaping - EvIL 😈

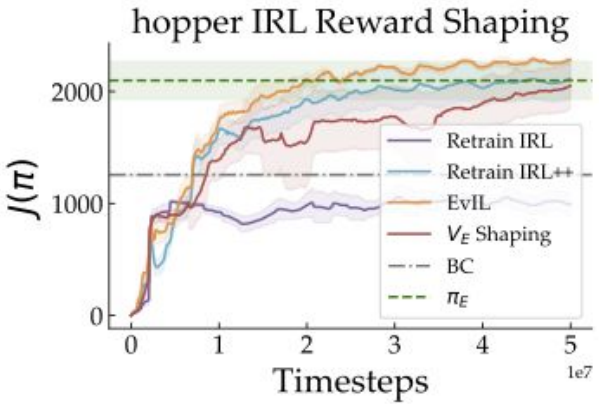


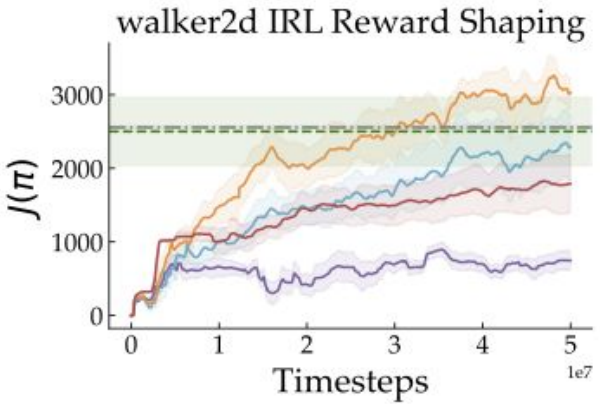
🔑 Idea: Reward Shaping - EvIL😈

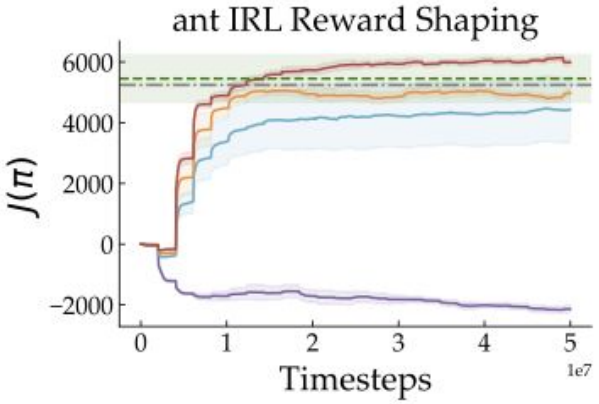


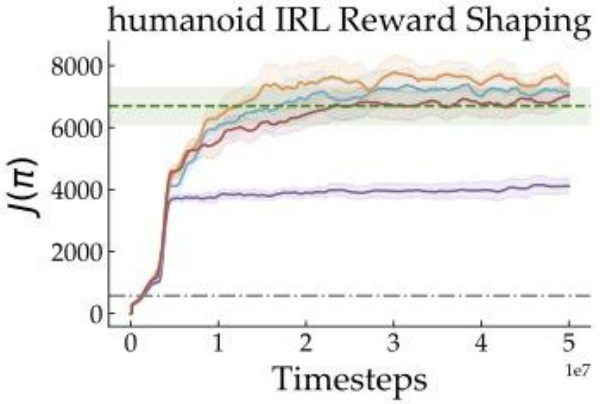


Results: Reward Shaping - EvIL😈

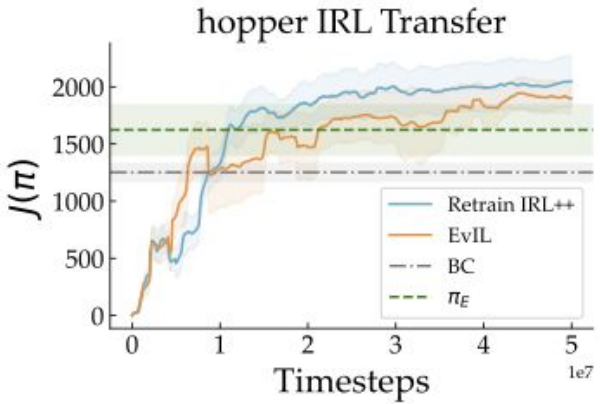


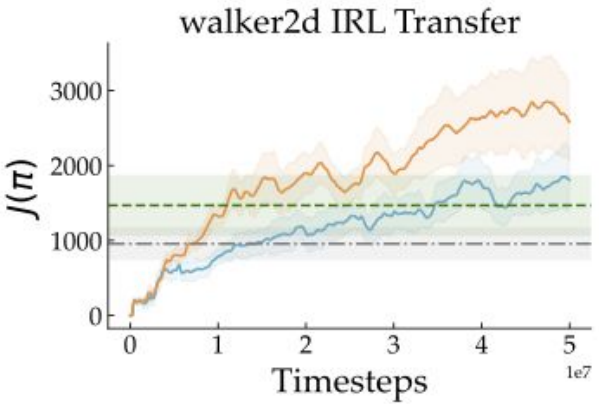


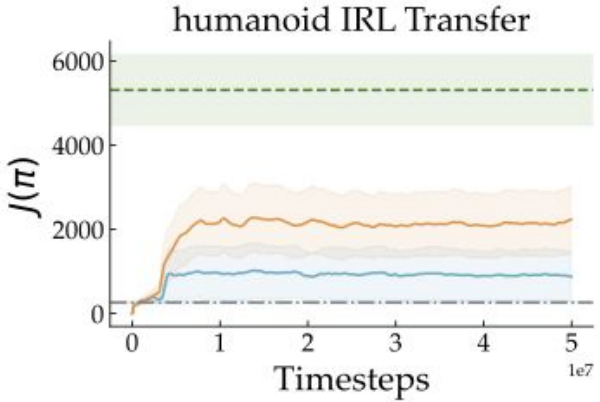


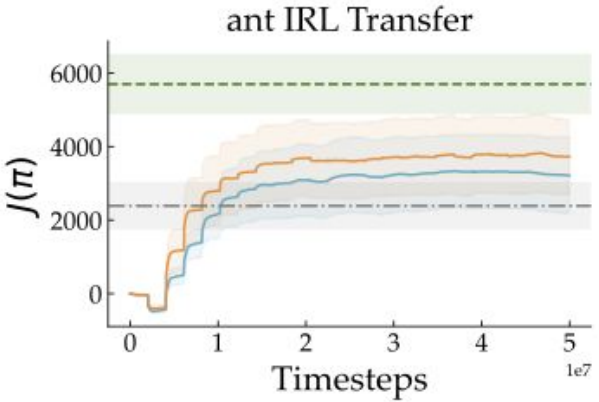


Results: Transfer - EvIL 😈









Open Source Code

github.com/SilviaSapora/evil