POLA - NeurIPS 2022

Proximal Learning with Opponent-Learning Awareness

Stephen Zhao, Chris Lu, Roger B. Grosse, Jakob N. Foerster

October 21, 2022

Motivation

Motivation

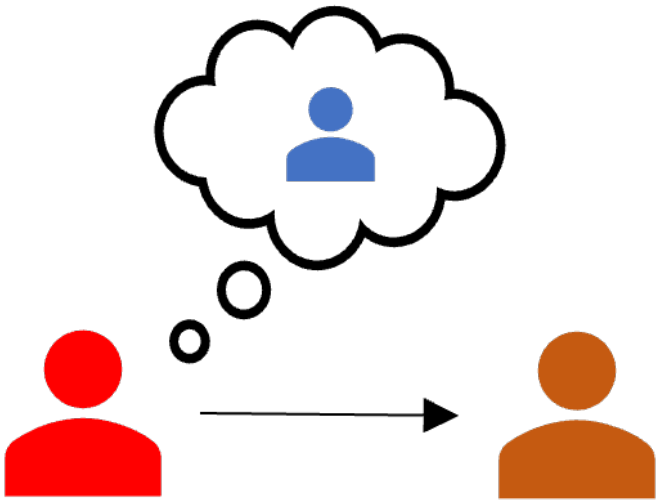
- Desired: reciprocity-based cooperation

Motivation

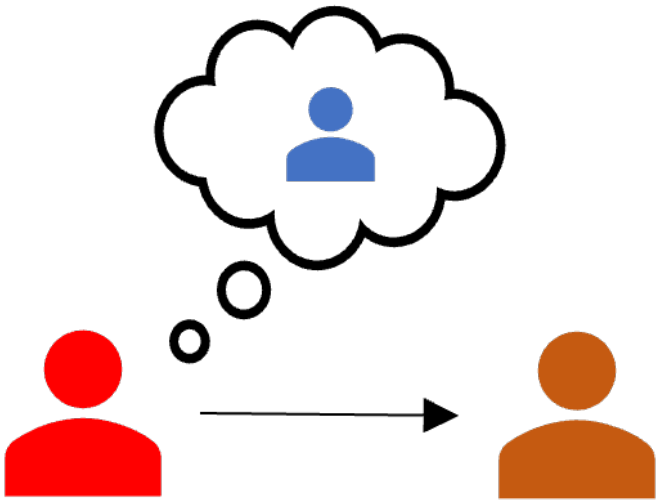
- Desired: reciprocity-based cooperation

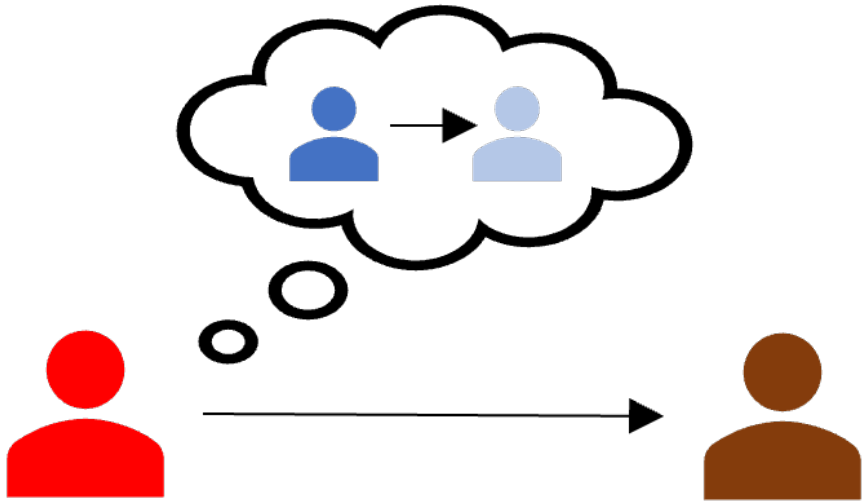
	Simple Policy Space	Complex Policy Space
LOLA (Foerster et al., 2018)	✓	×
Desired	✓	✓

Background – LOLA



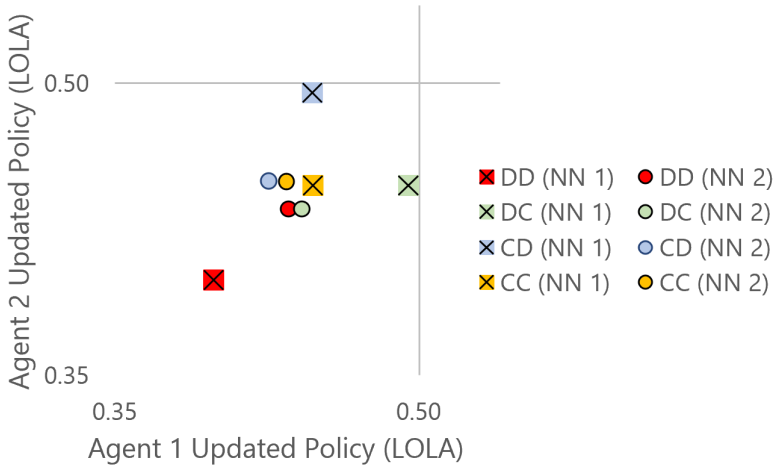
Background – LOLA

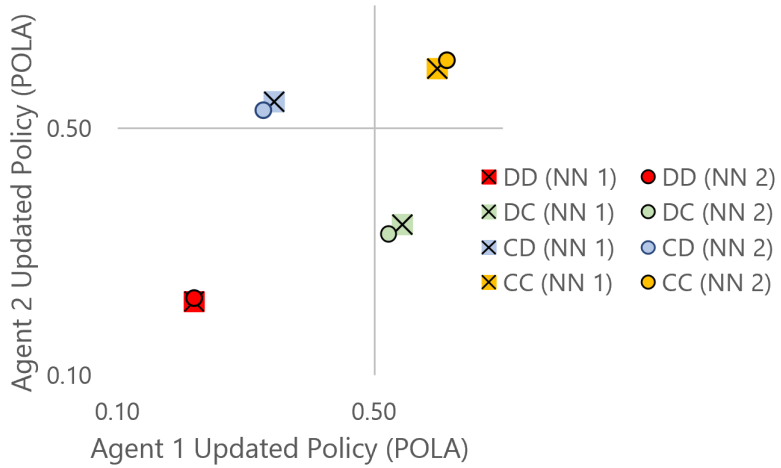




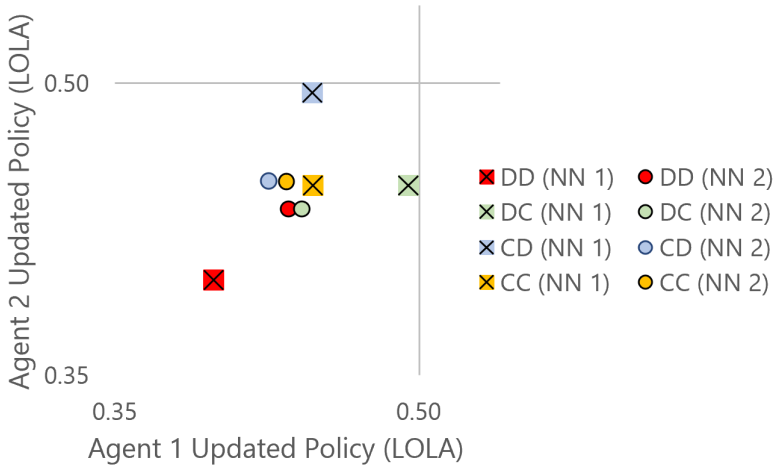
LOLA is Sensitive to Policy Parameterization

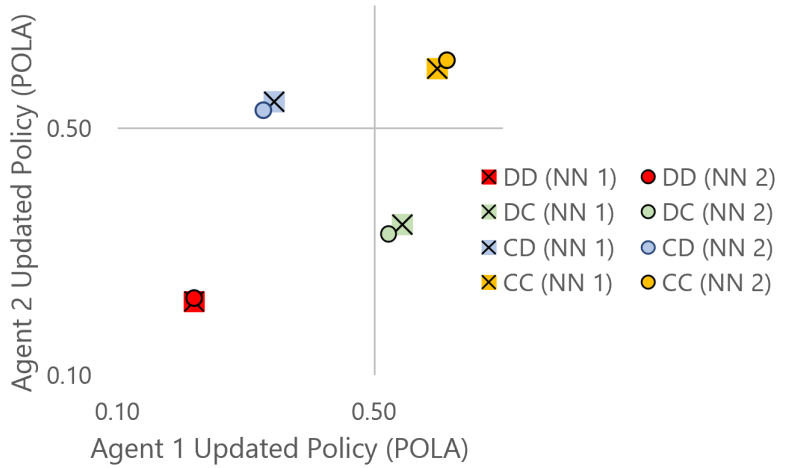
LOLA is Sensitive to Policy Parameterization



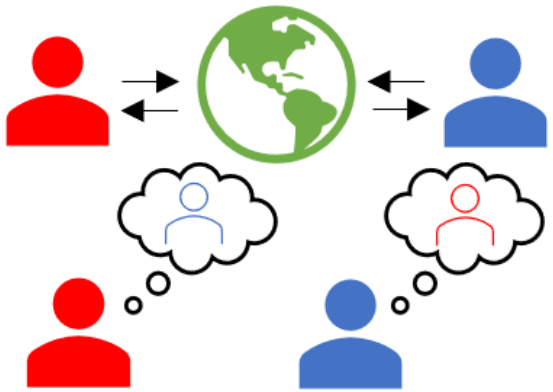


LOLA is Sensitive to Policy Parameterization





- LOLA is ill-defined under opponent modeling



POLA - Ideal Formulation

POLA - Ideal Formulation

- Agent 1's update solves for $\theta^{1\prime}$:

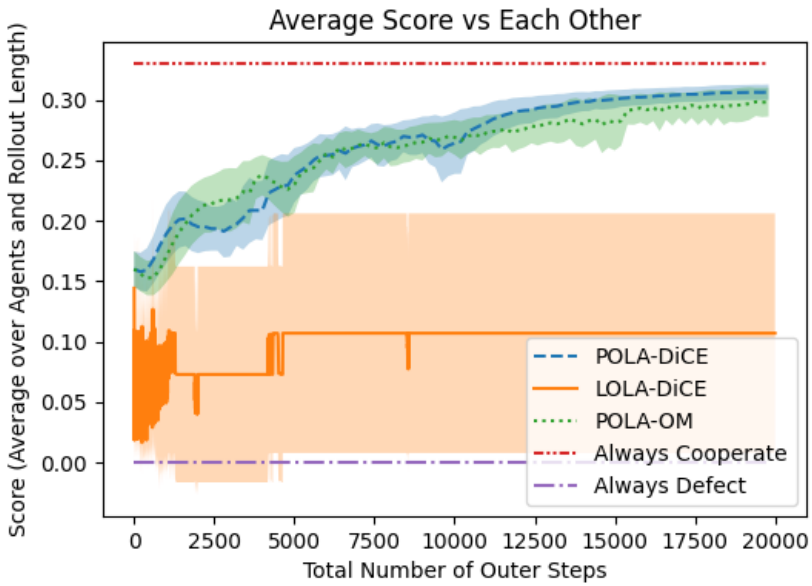
$$\theta^{1\prime}(\theta^1, \theta^2) = \arg\min_{\theta^{1\prime\prime}} \left(L^1(\pi_{\theta^{1\prime\prime}}, \pi_{\theta^{2\prime}(\theta^{1\prime\prime}, \theta^2)}) + \beta_{\text{out}} D(\pi_{\theta^1} || \pi_{\theta^{1\prime\prime}}) \right) \quad (1)$$

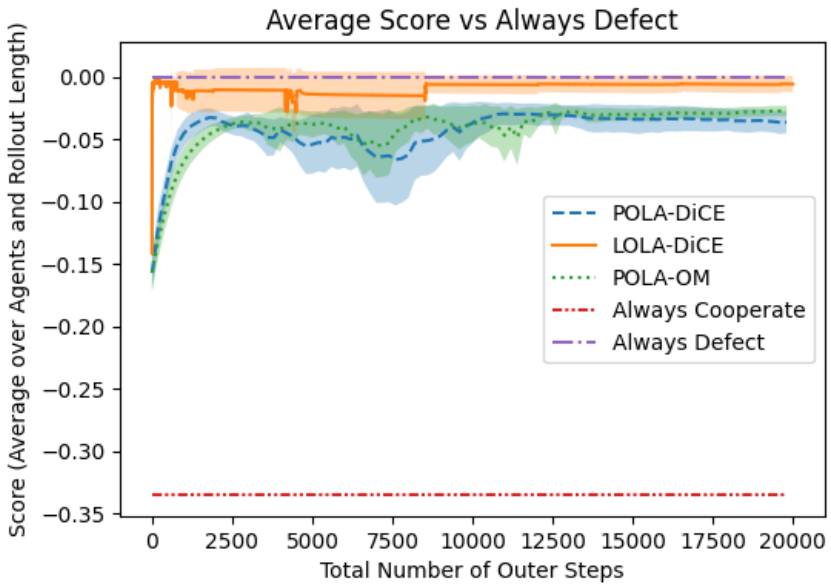
$$\theta^{2\prime}(\theta^{1\prime\prime}, \theta^2) = \arg\min_{\theta^{2\prime\prime}} \left(L^2(\pi_{\theta^{1\prime\prime}}, \pi_{\theta^{2\prime\prime}}) + \beta_{\text{in}} D(\pi_{\theta^2} || \pi_{\theta^{2\prime\prime}}) \right) \quad (2)$$

POLA - POLA-DiCE Approximation

POLA - POLA-DiCE Approximation

- IPD with full history:





POLA - POLA-DiCE Approximation

- Coin game: