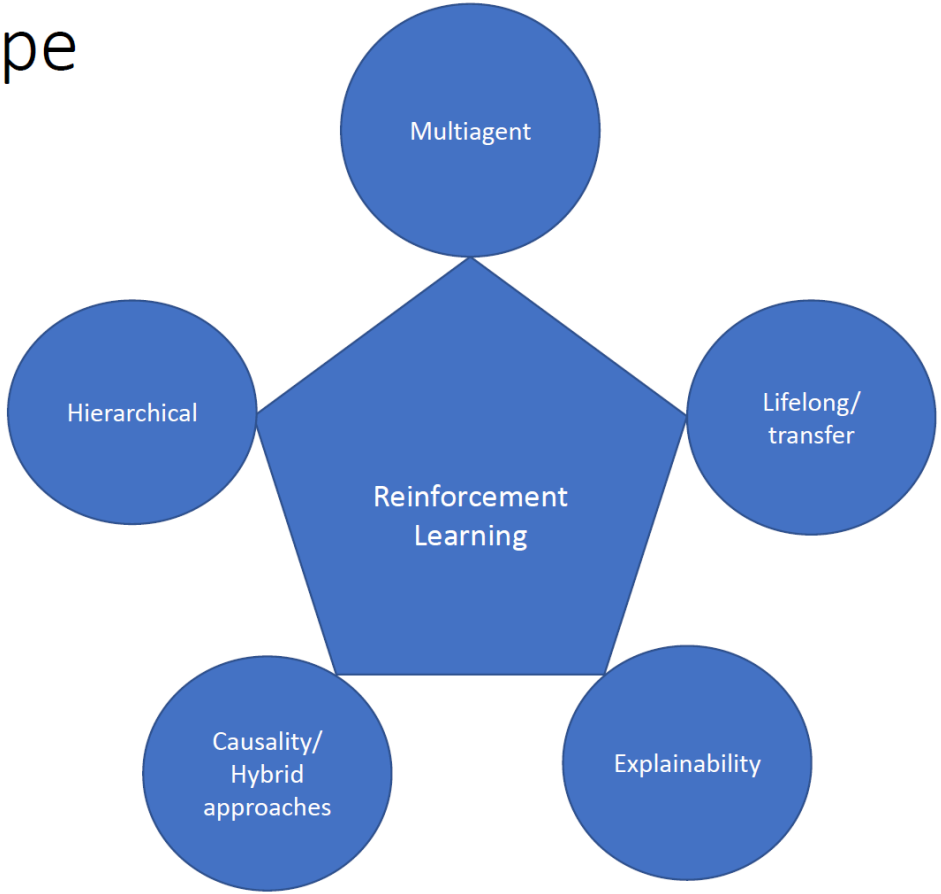
A Policy Gradient Algorithm for Learning to Learn in Multiagent Reinforcement Learning

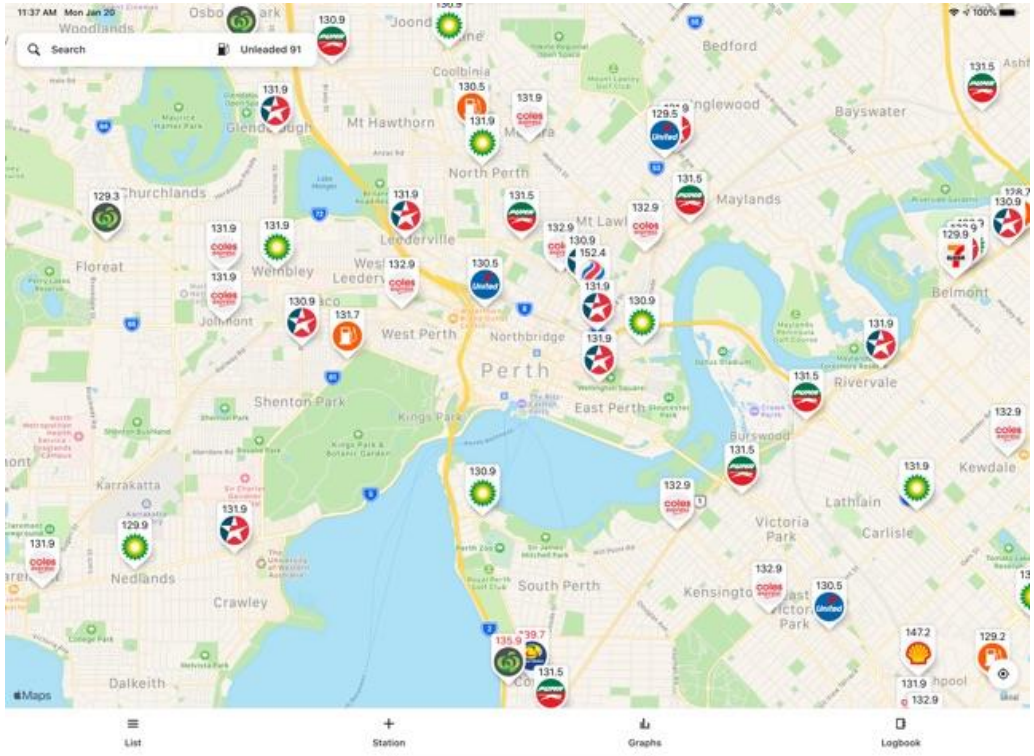
Miao Liu
IBM Research

Research Scope



Pricing in Agent Economies

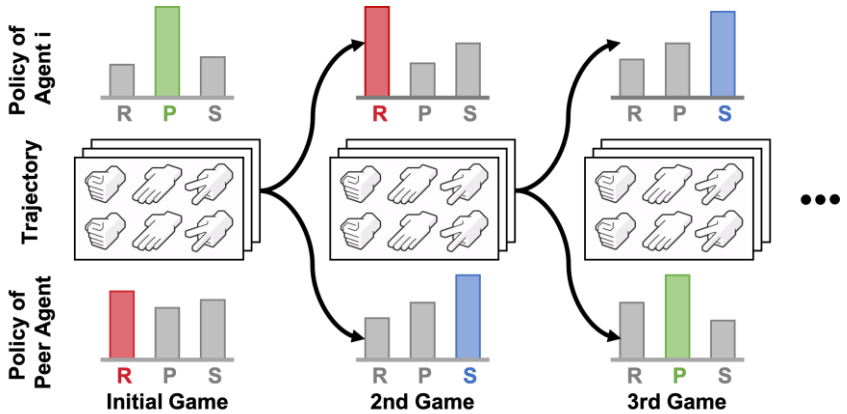
- For a population of agent each trying to adapt in the presence of other adaptive agents
- Information about the other agents might be imperfect



Pricing in Agent Economies Using Multiagent Q-learning. G. Tesauro and J. Kephart, Autonomous Agents and Multiagent Systems 2002
Shopbots and Pricebots. A. Greenwald and J. Kephart, IJCAI99

Motivation: Non-Stationary in MARL

- **Fundamental challenge in multiagent RL (MARL) :** presence of non-stationary policies due to simultaneously learning agents
  - Motivating example: changing strategies between games in play rock-paper-scissors
- **Key observation:** future policies of fellow agents are dependent on current performance and vice-versa
- Is it possible to use this sequential dependency and improve adaptation performance?



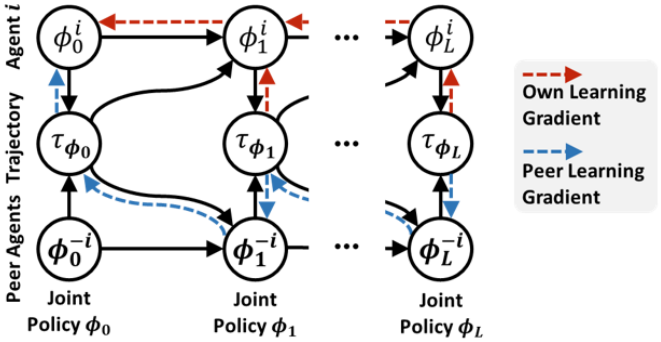
Overview: Our Contribution

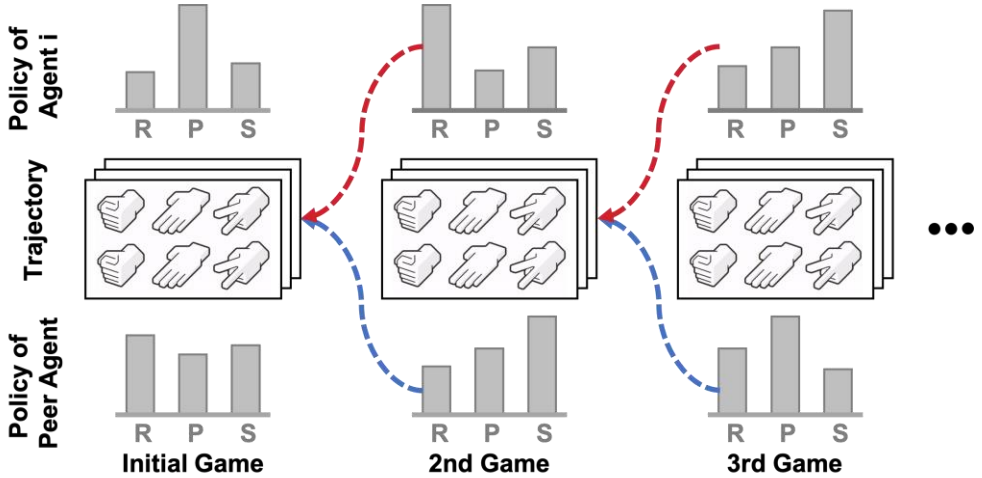
- **Objective**:
  - Adapt efficiently by considering sequential dependencies in non-stationary policy dynamics inherent to multiagent learning settings
- **Challenge**:
  - How to adapt w.r.t. potentially unknown changes in policies of fellow agents?
  - How to adapt in sample efficient manner as other agents' policies can update again after small number of interactions?
- **Key idea:**
  - To fully address non-stationarity, it is important to model both agent's own non-stationary policies dynamics and non-stationary policies of other agents

A Policy Gradient Algorithm for Learning to Learn in Multiagent Reinforcement Learning. D K Kim, M Liu, M Riemer, C Sun, M Abdulhai, G Habibi, S Lopez-Cot, G Tesauro and J P How., ICML2021

Meta-MAPG: Policy Gradient for Learning to Learn in MARL

- **Meta-multiagent policy gradient (Meta-MAPG):** Optimize agent i's initial policy with meta-learning so that it maximizes expected adaptation performance over games
  - **Own learning: Considers impact of initial policy on its own future policies**
  - **Peer learning: Actively learns to influence peers' future policies**

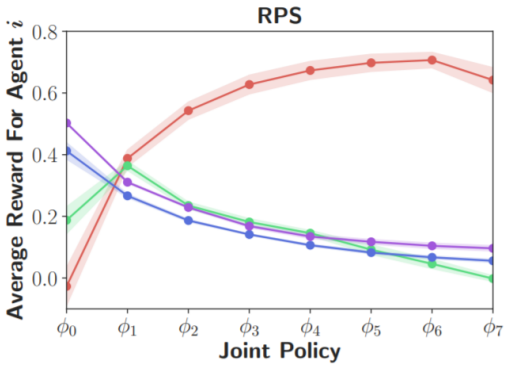


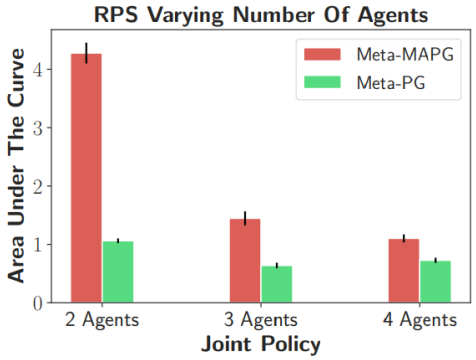


A Policy Gradient Algorithm for Learning to Learn in Multiagent Reinforcement Learning. D K Kim, M Liu, M Riemer, C Sun, M Abdulhai, G Habibi, S Lopez-Cot, G Tesauro and J P How., ICML2021

Result: Adaptation Performance in Rock-Paper-Scissors (RPS)

- **Baselines:**
  - **Meta-PG (AI-Shevidat et al., 2018): without peer learning**
  - **LOLA-DICE (Foerster et al., 2018): without own learning**
  - **REINFORCE (Williams, 1992): Without own and peer learning**



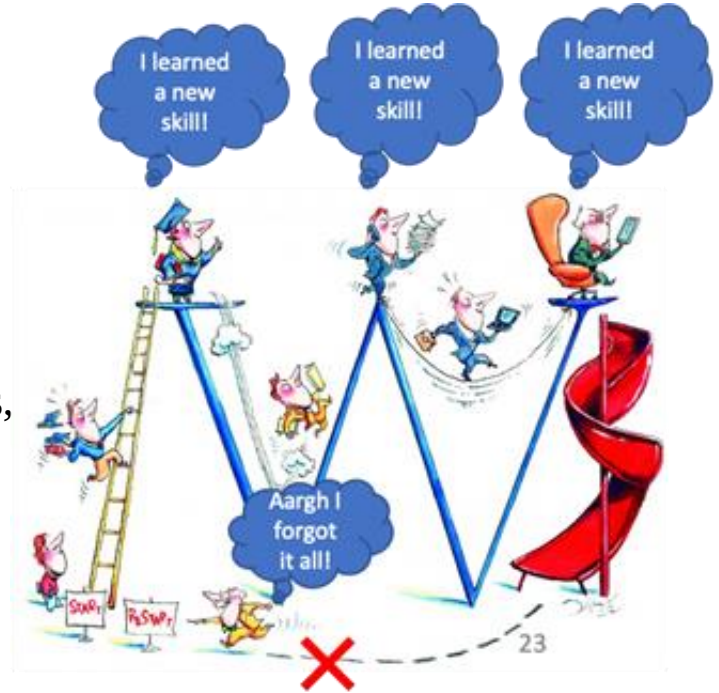


Matthew Riemer: Intro to My Work

Researcher at IBM and PhD Student at Mila

Research Topic: Continual Reinforcement Learning

- **Main Interest:** understanding how to perform machine learning on non-stationary distributions
- **Problem Setting:** large state spaces, large mixing times, and highly non-Markovian dynamics
- **Main Approaches**: learning to learn (non-stationary policies), temporal abstraction, and state abstraction



IBM

Mila

Influencing Long-Term Behavior in Multiagent Reinforcement Learning

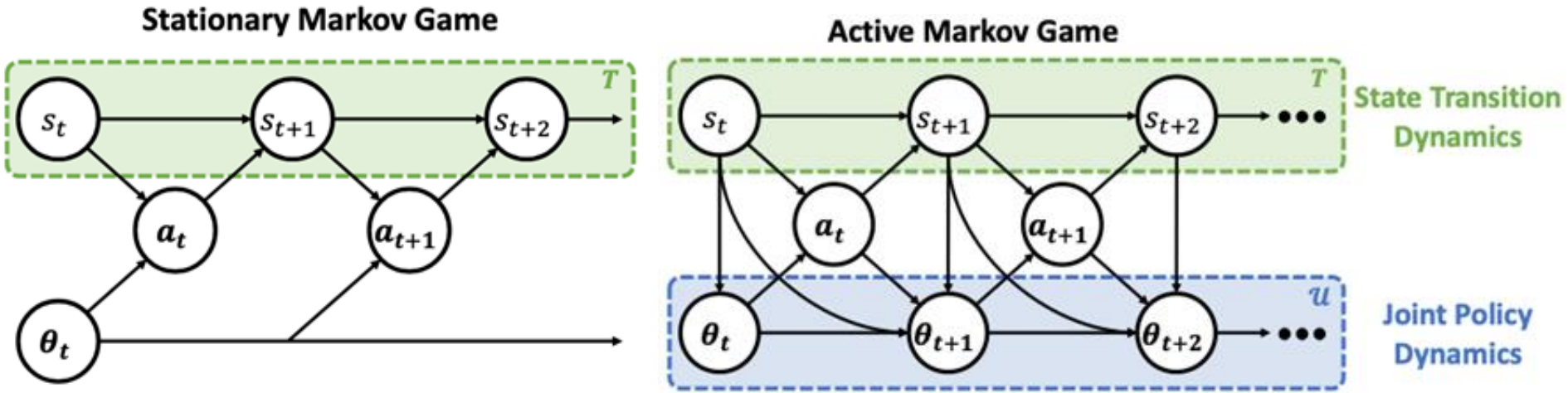
To Be Presented at NeurIPS 2022

By Dong-Ki Kim, Matthew Riemer, Miao Liu, Jakob Foerster, Michael Everett, Chuangchuang Sun, Gerald Tesauro, and Jonathan How

IBM

Mila

Modeling “Active” Markov Games



IBM

Mila

Active Markov Games: Equilibria

Definition (Active Equilibrium)

An active equilibrium is a set of joint policy parameters $\boldsymbol{\theta}^* = \{\theta^{i*}, \boldsymbol{\theta}^{-i*}\}$ *with associated joint update functions* $\boldsymbol{\mathcal{U}}^* = \{\mathcal{U}^{i*}, \boldsymbol{\mathcal{U}}^{-i*}\}$ *such that:*

$$\rho^i(s, \theta^{i*}, \boldsymbol{\theta}^{-i*}, \mathcal{U}^{i*}, \boldsymbol{\mathcal{U}}^{-i*}) \geq \rho^i(s, \theta^i, \boldsymbol{\theta}^{-i*}, \mathcal{U}^i, \boldsymbol{\mathcal{U}}^{-i*})$$
$$\forall i \in \mathcal{I}, s \in \mathcal{S}, \theta^i \in \Theta^i, \mathcal{U}^i \in \mathbb{U}^i$$

IBM

Mila