Breaking the Curse of Multiagency in Robust Multi-Agent Reinforcement Learning

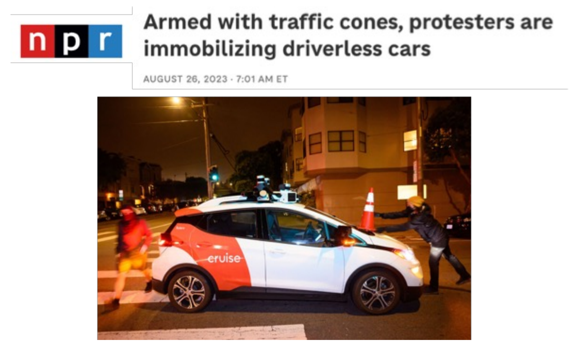
Laixi Shi* Jingchu Gai* Eric Mazumdar Yuejie Chi Adam Wierman

JHU CMU Caltech CMU Caltech

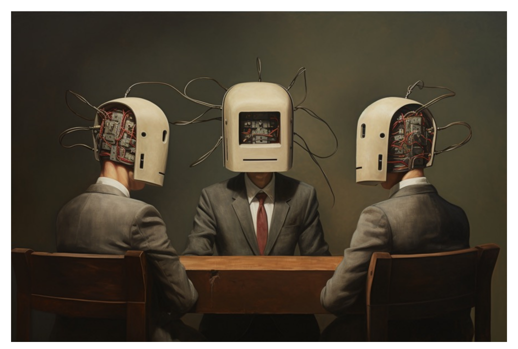
ICML International Conference On Machine Learning

Background: Multi-Agent Reinforcement Learning (MARL)

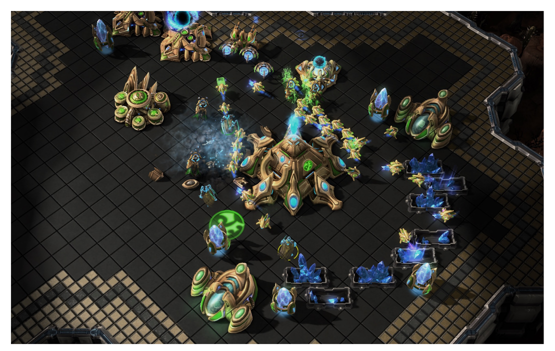
- **MARL:** AI agents are increasingly enmeshed in strategic human agent society



Autonomous driving

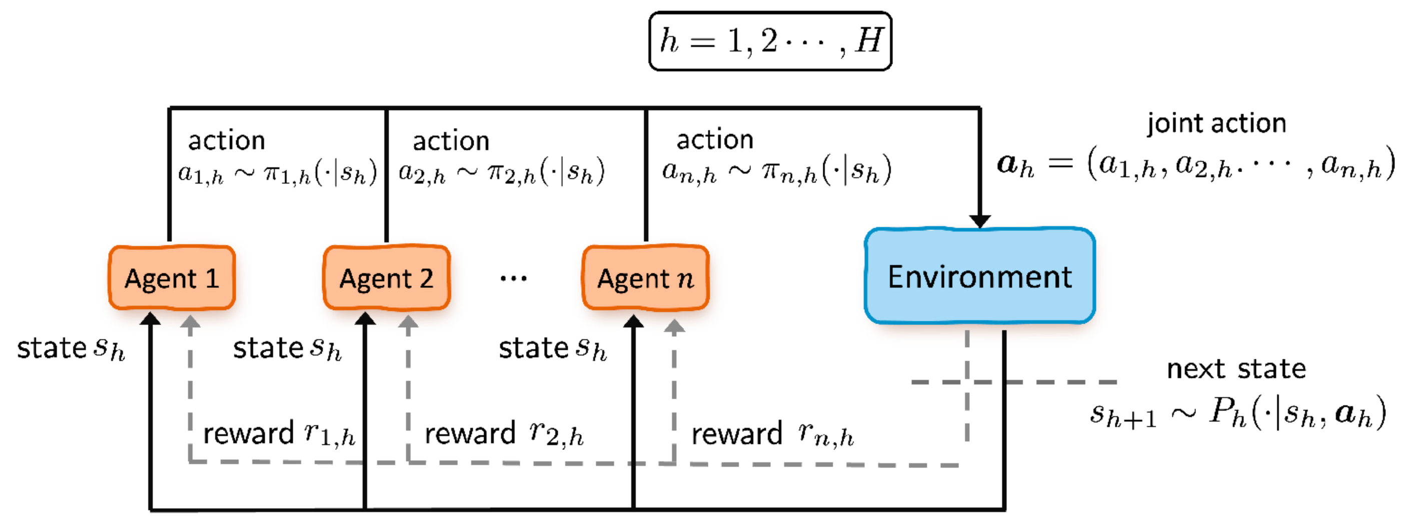


Human-AI collaboration



Multi-player games

- **Formulation: multi-player general-sum Markov games (MGs)**
  - n-player, finite state space $\mathcal{S}$, finite action spaces $\mathcal{A}_i$ for the i-th agent.

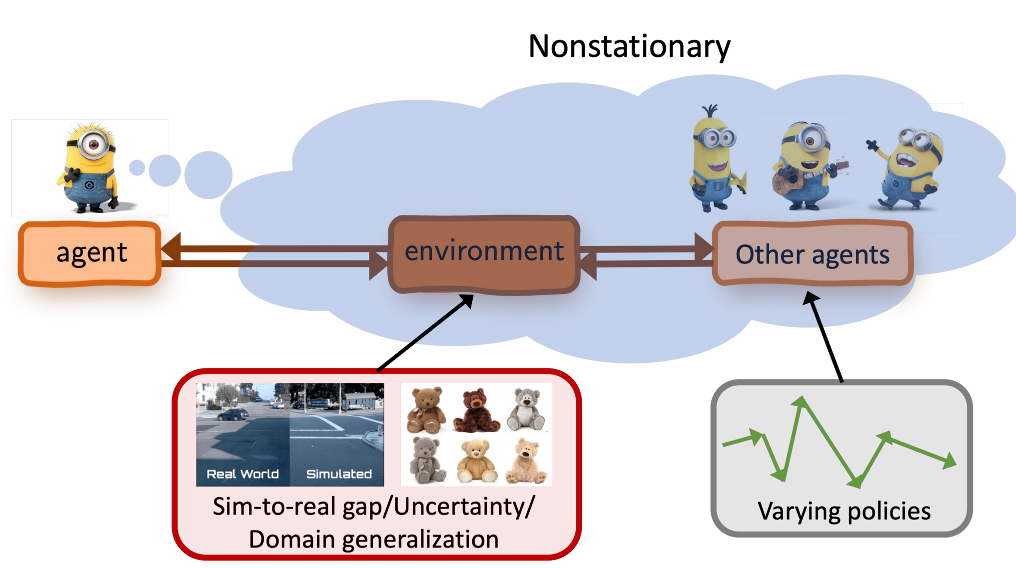


- **Value functions:** for any joint policy π, the cumulative reward is

$$\forall (i,s) \in [n] \times \mathcal{S}: \quad V_{i,h}^{\pi,P}(s) := \mathbb{E}_{\pi,P}\left[\sum_{t=h}^{H} r_{i,t}(s_t, \boldsymbol{a}_t) \mid s_h = s\right]$$

Robust MARL

- **Robust MARL:** promote robustness to environment shift and nonstationary of agents



- **Formulation: robust MGs (RMGs) with uncertainty set $\mathcal{U}_\rho^{\sigma_i}(P^0, \cdot)$**
  - ➢ ρ: divergence function; σ_i: uncertainty set radius
  - ➢ the transition kernel P is not fixed; vary within a prescribed uncertainty set determined by (possibly the current policy and) a nominal kernel P^0 (e.g., the training environment)
  - ➢ Robust value functions: $V_{i,h}^{\pi,\sigma_i}(s) := \inf_{P \in \mathcal{U}_\rho^{\sigma_i}(P^0,\pi)} V_{i,h}^{\pi,P}(s)$

- **Goal:** find some game-theoretical equilibrium strategies:
  - robust NE: a product policy $\pi: \mathcal{S} \times [H] \mapsto \prod_{1 \le i \le n} \Delta(\mathcal{A}_i)$ s.t. $V_{i,1}^{\pi,\sigma_i}(s) = \max_{\pi_i'} V_{i,1}^{\pi_i' \times \pi_{-i},\sigma_i}(s), \forall i, s$
  - robust CCE: a joint policy $\pi: \mathcal{S} \times [H] \mapsto \Delta(\prod_{1 \le i \le n} \mathcal{A}_i)$ s.t. $V_{i,1}^{\pi,\sigma_i}(s) \ge \max_{\pi_i'} V_{i,1}^{\pi_i' \times \pi_{-i},\sigma_i}(s), \forall i, s$

Challenges of Robust MARL

- **1. Construction of realistic uncertainty sets:** enabled by richness of robust MGs
  - ➢ Existing (s,a)-rectangular uncertainty set consider each agent's objective function using independent risk-aware outcome on each joint action

$$\mathbb{E}_{a_{-i} \in \pi_{-i}}[\mathbf{Risk}(V_{i,h}^{\pi,P}(a_i, \boldsymbol{a}_{-i}))]$$

  - ➢ **Observations from behavioral economics [Goeree et al., 2005]:** people often use a risk-aware metric outside of the expected outcome of other players' joint policy

$$\mathbf{Risk}(\mathbb{E}_{a_{-i} \in \pi_{-i}}[V_{i,h}^{\pi,P}(a_i, \boldsymbol{a}_{-i})])$$ ✅

- **2. Data efficiency --- The curse of multiagency:**
  - P^0 is unknown, need data to query samples from P^0 for uncertainty set estimation
  - The existing sample complexity requirement scales exponentially with the number of agents (using (s,a)-rectangular uncertainty set)

$$\widetilde{O}\left(\frac{H^3 S \prod_{i=1}^n A_i}{\varepsilon^2} \min\left\{H, \frac{1}{\min_{1 \le i \le n} \sigma_i}\right\}\right)$$

Robust MGs with Fictitious Uncertainty Sets

- **Expected nominal transition kernel:** for any joint policy $\pi: \mathcal{S} \times [H] \mapsto \Delta(\mathcal{A})$
  - conditioned on: the i-th agent plays action a_i and others play $\boldsymbol{a}_{-i} \sim \pi_h(\cdot \mid s, a_i)$

$$\forall (h,s,a_i) \in [H] \times \mathcal{S} \times \mathcal{A}_i: \quad P_{h,s,a_i}^{\pi_{-i}} = \mathbb{E}_{\boldsymbol{a} \sim \pi_h(\cdot \mid s, a_i)}\left[P_{h,s,\boldsymbol{a}}^0\right] = \sum_{\boldsymbol{a}_{-i} \in \mathcal{A}_{-i}} \frac{\pi_h(a_i, \boldsymbol{a}_{-i} \mid s)}{\pi_{i,h}(a_i \mid s)}\left[P_{h,s,\boldsymbol{a}}^0\right]$$

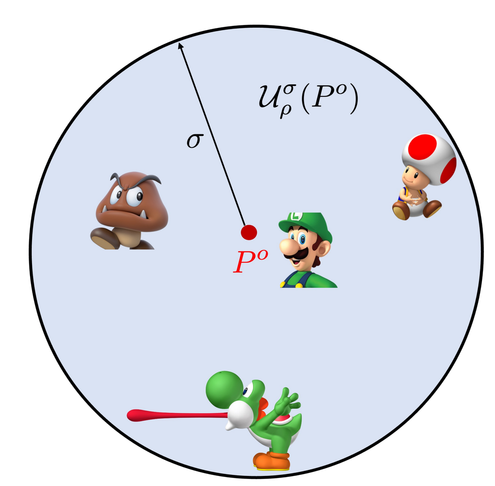
- **Fictitious uncertainty set:**

Others-integrated (s, a_i)-rectangular set

$$\forall i \in [n]: \ \mathcal{U}_\rho^{\sigma_i}(P^0, \pi) := \otimes\, \mathcal{U}_\rho^{\sigma_i}\left(P_{h,s,a_i}^{\pi_{-i}}\right),$$

$$\mathcal{U}_\rho^{\sigma_i}\left(P_{h,s,a_i}^{\pi_{-i}}\right) := \left\{P \in \Delta(\mathcal{S}): \rho\left(P, P_{h,s,a_i}^{\pi_{-i}}\right) \le \sigma_i\right\}$$

  - ➢ For i-th agent and each (s, a_i), the uncertainty set $\mathcal{U}_\rho^{\sigma_i}\left(P_{h,s,a_i}^{\pi_{-i}}\right)$ is a ball around the expected nominal transition kernel $P_{h,s,a_i}^{\pi_{-i}}$



- **Why fictitious uncertainty set**
  - ➢ **Realistic and predictive of human decisions** in comparisons to prior works using $(s, \boldsymbol{a})$-rectangular set (others-separated uncertainty set)

$$\mathcal{U}_\rho^{\sigma_i}(P^0) := \otimes \mathcal{U}^{\sigma_i}(P_{h,s,\boldsymbol{a}}^0), \quad \text{where} \quad \mathcal{U}_\rho^{\sigma_i}(P_{h,s,\boldsymbol{a}}^0) = \left\{P_{h,s,\boldsymbol{a}} \in \Delta(\mathcal{S}): \rho(P_{h,s,\boldsymbol{a}}, P_{h,s,\boldsymbol{a}}^0) \le \sigma_i\right\}$$

  - ➢ **A natural adaptation from single-agent robust RL**: Fixing other agents' policy π_i, from the viewpoint of the individual I, RMGs with fictitious uncertainty set degrades to a single-agent robust RL problem

- **Properties of RMGs with fictitious uncertainty set**

Theorem 1: Existence of robust NE, CCE, and CE

This Work: design Robust MGs with realistic uncertainty sets and sample complexity guarantees breaking the curse of multiagency

Breaking Curse of Multiagency of Sample Complexity

- **Setting:**
  - Using total variation (TV) as ρ: $\forall P, P' \in \Delta(\mathcal{S})$: $\quad \rho_{\mathrm{TV}}(P, P') := \frac{1}{2}\|P - P'\|_1$
  - Data collection mechanism: a generative model for the true nominal kernel P^0

$$s_{h,s,\boldsymbol{a}}^i \overset{i.i.d}{\sim} P_h^0(\cdot \mid s, \boldsymbol{a}), \qquad i = 1, 2, \ldots$$

  - Goal: find an ε-approximate robust-CCE ξ, i.e.,

$$\mathsf{gap}_{\mathrm{CCE}}(\xi) := \max_{s \in \mathcal{S}, 1 \le i \le n}\left\{\mathbb{E}_{\pi \sim \xi}\left[V_{i,1}^{\star,\pi_{-i},\sigma_i}(s)\right] - \mathbb{E}_{\pi \sim \xi}\left[V_{i,1}^{\pi,\sigma_i}(s)\right]\right\} \le \varepsilon$$

- **Algorithm design: Robust-Q-FTRL**
  - ➢ Using tailored **online adversarial learning** algorithm: tailored FTRL
  - ➢ **Using N-sample estimation** for empirical kernel and robust Q-function:
    - ➢ Handle additional optimization vs statistical challenges

Theorem 2: upper bound with breaking curse of multiagency

Using the TV distance, for any RMGs with fictitious uncertainty set and any $\varepsilon \le \sqrt{\min\{H, \frac{1}{\min_{1 \le i \le n} \sigma_i}\}}$ *robust-Q-FTRL can output an* ε-approximate robust-CCE $\hat{\xi}$ as long as the total number of samples acquired in the learning process exceeds

$$\widetilde{O}\left(\frac{SH^6 \sum_{1 \le i \le n} A_i}{\varepsilon^4} \min\left\{H, \frac{1}{\min_{1 \le i \le n} \sigma_i}\right\}\right)$$

- **Discussions**
  - **Lower bound:** $\widetilde{O}\left(\frac{SH^3 \max_{1 \le i \le n} A_i}{\varepsilon^2} \min\left\{H, \frac{1}{\min_{1 \le i \le n} \sigma_i}\right\}\right)$
  - **Comparisons with prior works on general RMGs**

Algorithm	Uncertainty set	Equilibria	Sample complexity
P²MPO (Blanchet et al., 2024)	$(s,\boldsymbol{a})$-rectangularity	robust NE	$S^4\left(\prod_{i=1}^n A_i\right)^3 H^4/\varepsilon^2$
DR-NVI (Shi et al., 2024)	$(s,\boldsymbol{a})$-rectangularity	robust NE/CE/CCE	$\frac{SH^3\prod_{i=1}^n A_i}{\varepsilon^2}\min\left\{H, \frac{1}{\min_{1 \le i \le n}\sigma_i}\right\}$
Robust-Q-FTRL **(this work)**	fictitious (s,a_i)-rectangularity	robust CCE	$\frac{SH^6\sum_{1 \le i \le n} A_i}{\varepsilon^4}\min\left\{H, \frac{1}{\min_{1 \le i \le n}\sigma_i}\right\}$

- **Technical insights:** Prior approaches for breaking curse in standard MARL can't apply rely on linear value functions (w.r.t. transition kernel) for error cancellation, but RMGs' nonlinearity prevent this. There is a tradeoff between statistical (data) efficiency and tight regret bound in online optimization induced by nonlinearity.