Investigating Symbiosis in Robotic Ecosystems: A Case Study for Multi-Robot Reinforcement Learning Reward Shaping

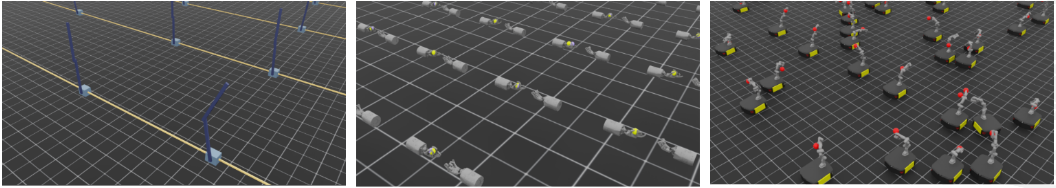
Xuezhi Niu & Didem Gürdür Broo

Cyber-Physical Systems Lab, Department of Information Technology,
Uppsala University

June 28, 2025
2025 9th International Conference on Robotics and Automation Sciences

Agenda

Introduction & Motivation



Multi-Robot Teams Training in Simulation

- **Heterogeneous robots** (wheeled, arms, aerial) can complement one another in complex tasks [1].
- Coordination across unequal agents remains challenging [2].

- MARL struggles with **credit assignment** and **non-stationarity**, especially in mixed-capability teams [3].

- Existing reward shaping is often **heuristic** [4] and fragile:
  - Unstable training
  - Poor coordination
  - Breaks under strong heterogeneity

Inspiration from Nature: How do biological systems achieve seamless cooperation among diverse entities? Perhaps nature's playbook holds clues (e.g., symbiosis in ecosystems [5]).

Biological Inspiration - Mutualism as a Reward Signal

Symbiosis!

A biological relationship where two or more organisms interact for continuous existence.

- Mycorrhizal networks between trees and fungi: sharing resources and information to support collective survival [6]

Figure 1: Mycorrhizal networks between trees and fungi.

Mutualism $\neq$ Altruism

Contributions

- **Formal framework** for modeling mutualism in multi-robot systems (MRS)
- **Reward shaping method** inspired by ecological cooperation, promoting robust coordination under limited task knowledge

Symbiosis!

Let $H = \{a_1, \ldots, a_n\}$ be a set of heterogeneous robots. Each a_i has:

- Capability set C_i
- Resource vector D_i
- Performance function P_i

Symbiotic interaction between a_i and a_j could be defined as:

$$I(a_i, a_j) > \max\{P_i, P_j\} - \delta \quad (\delta \geq 0)$$

Total system performance for a subset $S \subseteq H$ is:

$$P_{\text{total}}(S) = \sum_{a_i \in S} P_i + \sum_{(a_i, a_j) \in E(S)} I(a_i, a_j). \tag{1}$$

Taxonomy of Interaction Types

Modeling Inter-Agent Symbiosis:

- Mutualism: $\Delta P(a_i, a_j) > 0$ and $\Delta P(a_j, a_i) > 0$
- Commensalism: $\Delta P(a_i, a_j) > 0$ and $\Delta P(a_j, a_i) = 0$
- Parasitism: $\Delta P(a_i, a_j) > 0$ and $\Delta P(a_j, a_i) < 0$

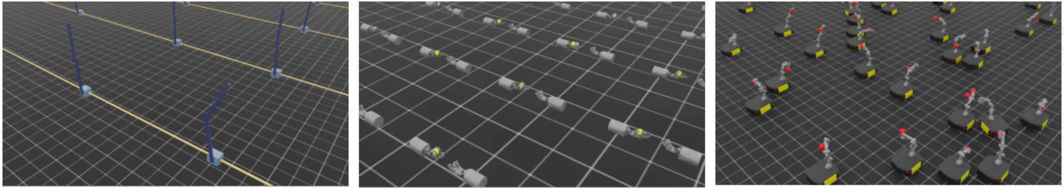
Examples in Mobile Manipulation:

- **Mutualism:** base positions for better arm reach; arm assists base with manipulation
- **Commensalism:** arm acts independently; base reuses trajectory
- **Parasitism:** arm moves aggressively, destabilizing the base

Goal: Promote **mutualism**, suppress harmful asymmetries via structured reward shaping

Environments for Evaluation

Figure 2: Benchmark training in Isaac Sim 4.5.0 using Isaac Lab, showcasing a screenshot with 512 parallel environments: (a) Double pendulum dynamics, (b) Shadow Hand object passing, and (c) Mobile manipulation tasks.



(a) (b) (c)

- Basic
- Dexterous
- Heterogeneous

Agent Observations and Actions

$$\mathbf{O}_{\text{cart}} \in \mathbb{R}^4,\ \mathbf{O}_{\text{cart}} \in \mathbb{R}^3$$
$$\mathbf{A}_{\text{cart}} \in \mathbb{R}^1,\ \mathbf{A}_{\text{cart}} \in \mathbb{R}^1$$

Cart Pendulum

- Cart: **observes** position, velocity, pole angle, pole velocity; **acts** on force
- Pendulum: **observes** pole angle, pendulum angle and velocity; **acts** on torque

$$\mathbf{O}_i \in \mathbb{R}^{157}$$
$$\mathbf{A}_{\text{cart}} \in \mathbb{R}^{20}$$

Shadow Hand

- Each hand: **observes** joint poses and velocities, fingertip poses and velocities, object and goal poses and velocities, and object to goal difference; **acts** on joint angle commands

$$\mathbf{O}_{\text{base}} \in \mathbb{R}^{15},\ \mathbf{O}_{\text{arm}} \in \mathbb{R}^{33}$$
$$\mathbf{A}_{\text{base}} \in \mathbb{R}^{3},\ \mathbf{A}_{\text{arm}} \in \mathbb{R}^{7}$$

Mobile Manipulation

- Base: **observes** base positions and velocities, finger positions, target position; **acts** on position
- Arm: **observes** arm positions and velocities, finger positions, target position; **acts** on joint states

Results

Reward formulation:

$$R_i = \alpha\, P_i + \beta \sum_{j \neq i} \Delta P(a_i, a_j)$$

Cart Pendulum

- $P_{\text{cart}} = \epsilon_{\text{pole pos}} \|\theta_{\text{pole}}\|_2 + \epsilon_{\text{pole vel}} |\dot{\theta}_{\text{pole}}|$
- $P_{\text{pendulum}} = \epsilon_{\text{pendulum pos}} \|\theta_{\text{pole}} + \theta_{\text{pendulum}}\|_2 + \epsilon_{\text{pendulum vel}} |\dot{\theta}_{\text{pendulum}}|$
- $\Delta P_{\text{cart}} = \epsilon_{\text{alive}}(1 - \delta_{\text{reset}}) + \epsilon_{\text{terminated}} \delta_{\text{reset}} + \epsilon_{\text{cart vel}} |\dot{x}_{\text{cart}}|$
- $\Delta P_{\text{pendulum}} = \epsilon_{\text{alive}}(1 - \delta_{\text{reset}}) + \epsilon_{\text{terminated}} \delta_{\text{reset}}$

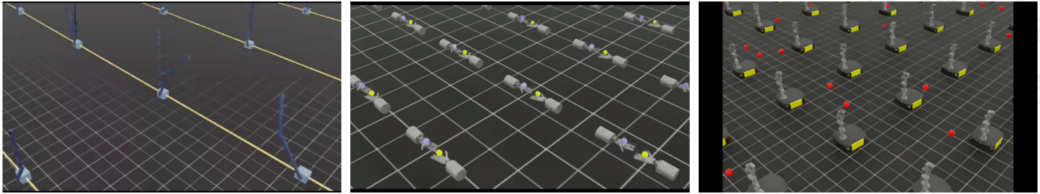
Shadow Hand

- $P_i = 2\, e^{(-20\, d)}$, with $d = \|\mathbf{p}_{\text{object}} - \mathbf{p}_{\text{goal}}\|_2$
- $\Delta P_{\text{right}} = \epsilon_{\text{release}}(1 - \delta_{\text{fail}})$
- $\Delta P_{\text{left}} = \epsilon_{\text{catch}}(1 - \delta_{\text{drop}})$

Mobile Manipulation

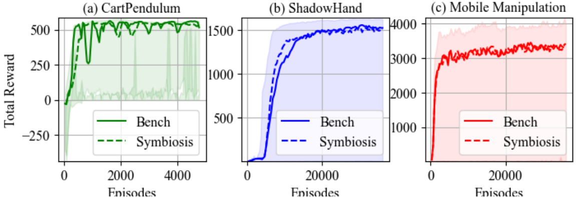
- $P_{\text{base}} = 5\, e^{(-2\, \|\mathbf{p}_{\text{obj}} - \mathbf{p}_{\text{goal}}\|_2)} - \epsilon_{\text{vel}} \|\mathbf{v}_{\text{base}}\|_2$
- $P_{\text{arm}} = 5\, e^{(-2\, \|\mathbf{p}_{\text{obj}} - \mathbf{p}_{\text{goal}}\|_2)} - \epsilon_{\text{vel}} \|\dot{\mathbf{q}}_{\text{arm}}\|_2$
- $\Delta P_{\text{base}} = -\epsilon_{\text{pos}} \|\mathbf{p}^{xy}_{\text{ee}} - \mathbf{p}^{xy}_{\text{target}}\|_2$
- $\Delta P_{\text{arm}} = -\epsilon_{\text{pos}} \|\mathbf{q}_{\text{arm}} - \mathbf{q}_{\text{target}}\|_2$

Results in Simulations



Results

Figure 3: Training results: total reward per episode with mean (solid) and variation (shaded). Evaluated with five random seeds. (a) Cart Pendulum, (b) Shadow Hand, (c) Mobile Manipulation.



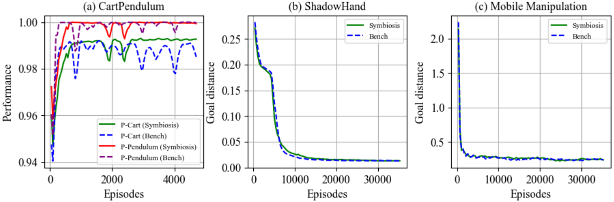


Figure 4: Mean performance comparison across tasks. Dashed lines indicate baselines. (a) Cart Pendulum, (b) Shadow Hand, (c) Mobile Manipulation.

1. Introduction
2. Methodology
3. Results
4. **Conclusions**

Discussion: Mutualism in Practice

Structured shaping via mutualism:

- Agents: $H = \{a_1, a_2, \ldots, a_n\}$ with capabilities C_i, resources D_i, performance P_i
- Mutual benefit: $I(a_i, a_j) > \max\{P_i, P_j\} - \delta$
- Shaping guides coordination without distorting task goals

Empirical findings:

- **Cart Pendulum:** minimal gains, but improved stability
- **Shadow Hand / Mobile Manipulation:** smoother learning, faster convergence, lower variance
- Benefits grow with task complexity and coordination demands

Implication: Reward portability $\rightarrow$ structure generalizes across tasks, reduces tuning effort

Conclusions & Future Works

The source code could be found at github.com/Cyber-physical-Systems-Lab/RewMARL

Summary:

- Formal framework for modeling symbiosis in multi-robot systems
- Mutualism-based reward shaping improves coordination in MARL
- Benefits: training stability, policy transfer, robustness

Next steps:

- Learn adaptive interaction functions $I(a_i, a_j)$
- Scale to larger, more diverse robot teams
- Combine with intrinsic rewards for open-ended tasks
- Extend to commensalism and parasitism dynamics

References

Back to start

[1] A. A. Nguyen, F. Jabbari, and M. Egerstedt, “Mutualistic interactions in heterogeneous multi-agent systems,” in *2023 62nd IEEE Conference on Decision and Control (CDC)*. IEEE, 2023, pp. 411–418.

[2] R. M. D’Souza, M. di Bernardo, and Y.-Y. Liu, “Controlling complex networks with complex nodes,” *Nature Reviews Physics*, vol. 5, no. 4, pp. 250–262, 2023.

[3] H. Zhang, X. Zhang, Z. Feng, and X. Xiao, “Heterogeneous multi-robot cooperation with asynchronous multi-agent reinforcement learning,” *IEEE Robotics and Automation Letters*, vol. 9, no. 1, pp. 159–166, 2023.

[4] H. Ma, K. Sima, T. V. Vo, D. Fu, and T.-Y. Leong, “Reward shaping for reinforcement learning with an assistant reward agent,” in *Proceedings of the 41st International Conference on Machine Learning*, ser. Proceedings of Machine Learning Research, R. Salakhutdinov, Z. Kolter, K. Heller, A. Weller, N. Oliver, J. Scarlett, and F. Berkenkamp, Eds., vol. 235. PMLR, 21–27 Jul 2024, pp. 33 925–33 939. [Online]. Available: https://proceedings.mlr.press/v235/ma24l.html

[5] A. A. Nguyen, M. Rodriguez Curras, M. Egerstedt, and J. N. Pauli, “Mutualisms as a framework for multi-robot collaboration,” *Frontiers in Robotics and AI*, vol. 12, p. 1566452, 2025.

[6] S. W. Simard, K. J. Beiler, M. A. Bingham, J. R. Deslippe, L. J. Philip, and F. P. Teste, “Mycorrhizal networks: mechanisms, ecology and modelling,” *Fungal Biology Reviews*, vol. 26, no. 1, pp. 39–60, 2012.