Enhancing Cooperation in Heterogeneous Multi-Robot Systems

Symbiosis-Inspired Interaction of Cooperation

Xuezhi Niu

Supervisor: Didem Gürdür Broo

Co-supervisor: Ginevra Castellano
Chair: Per Mattsson

Department of Information Technology, Uppsala University

xuezhi.niu@it.uu.se

26 August 2026
Half-Time Seminar

UPPSALA
UNIVERSITET

Heterogeneous multi-robot systems

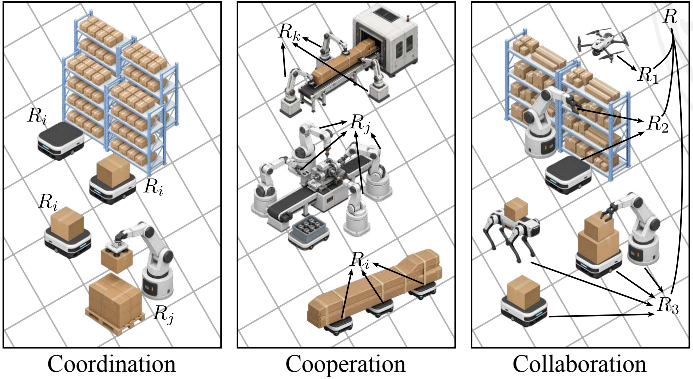
Robots with different capabilities are increasingly deployed together

But sharing a task does not make them a **mutual team**.

The hard part is not making robots different

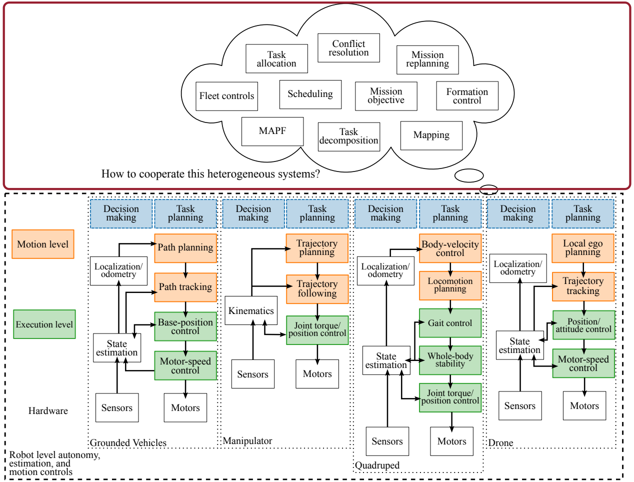
It is making their differences useful while tasks, resources, and partners change.

Coordination, cooperation, collaboration



The same pair can be irrelevant, helpful, or costly as tasks and resources change.

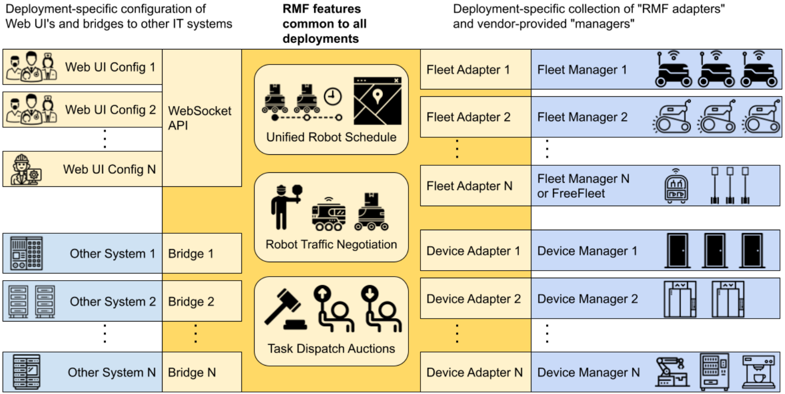
Coordination in heterogeneous robot teams



Autonomy beyond task execution, not just makespan

- **Centralized** planning and task allocation guarantee convergence once the graph & cost are fixed in advance
- **Decentralized** consensus and auctions encode effects indirectly through costs & constraints fixed at allocation

Existing coordination middleware



Mature middleware (Open-RMF, Intrinsic) routes and schedules heterogeneous fleets — it does not represent whether a partner helps, is neutral, or costs.

Heterogeneous robots depend on one another in different ways

Partial heterogeneity

Thesis setting

Identical capabilities — Overlapping but incomplete capabilities — Fully distinct capabilities

- Each robot can perform multiple functions.
- No robot satisfies all task requirements.
- Robot types do not operate in isolation.

Interdependence is unavoidable

- Act independently when individual capabilities are sufficient.
- Cooperate when capabilities must be combined.
- Share or compete for scarce resources.
- Preserve team function under task, resource, or robot changes.

Team-level requirements

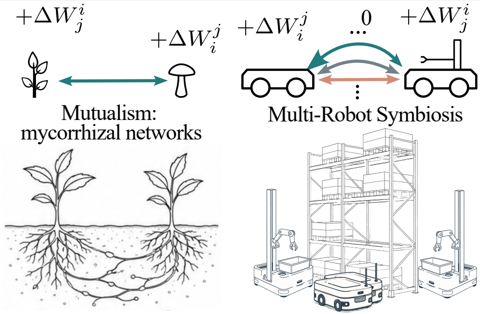
Task completion | Resource sharing | Resilience

Core lens: Symbiosis

What is symbiosis?

Definition

A biological relationship where two or more organisms interact for continuous existence, including **mutualism** (+,+), **commensalism** (+,0), **parasitism** (+,−), and **neutralism** (0,0).

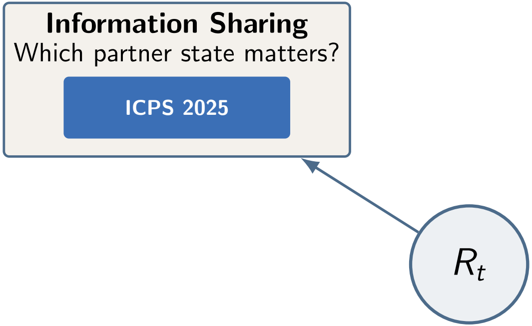


Symbiosis classifies effects, not intentions

Symbiosis does not mean that all agents help one another.
It describes the **sign and direction** of persistent interaction effects.

Ecology	Industrial ecology	Multi-robot system	Formalisation
Organism	Production unit	Robot or role	$i \in \mathcal{N}$
Fitness	Utility or continuity	Long-term value	$W_i^{\pi}(s_t)$
Environment	Resource conditions	Task and resource state	s_t
Effect $i \to j$	Net exchange benefit	Partner contribution	$\Delta W_{i \to j,t}$
Interaction type	Benefit–cost pattern	Directed dependence	$(\operatorname{sgn} \Delta W_{ij}, \operatorname{sgn} \Delta W_{ji})$
Interaction network	Exchange network	Dependence graph	$\mathbf{W}_t = [w_{i \to j,t}]$

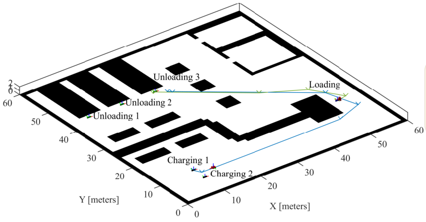
Symbiosis scope



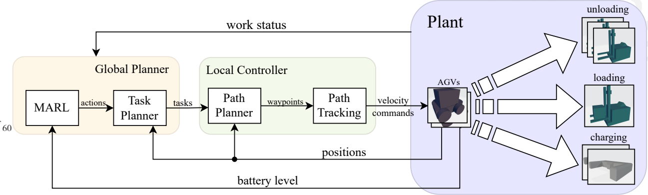
Symbiosis enters the decision layer through what robots **observe**, **optimize**, and **adapt**.

A warehouse as a first testbed

A shared warehouse is a natural first case: robots compete for aisles and charging points, yet can also help one another. MARL offers a way to learn an adaptive, partially decentralized symbiosis, since Dec-POMDP formulations otherwise assume a single shared global reward.

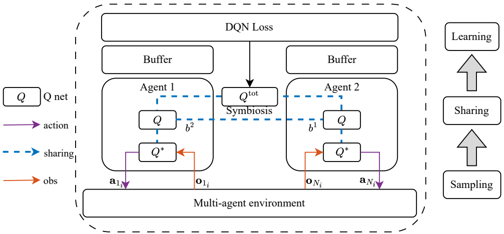


Simulated in Matlab/Simulink; two identical AGVs.



Centralized training and decentralized execution (CTDE).

Expose on partner state



Partner battery is added to the learner's information path (ICPS 2025).

From double DQN: individual Q_i and a team-optimal Q that sees all batteries from Q^{tot}. **Problem:** battery state hidden — charging conflicts invisible.

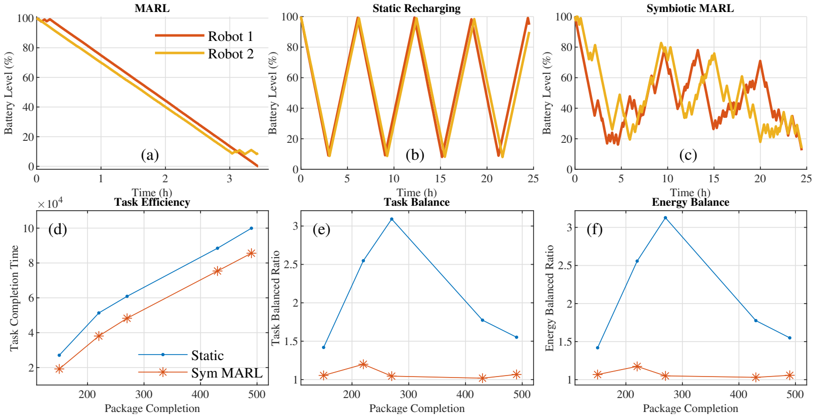
$$\tilde{o}_i = [o_i, b_j]$$

$$r^i_{\text{local}}(t) = \epsilon_1 p_i(t) - \epsilon_2\, e^{20(0.1 - b_i(t))} 1[b_i(t) < 0.1]$$

$$r_{\text{global}}(t) = T_{\text{total}} - \frac{1}{N}\sum_{i=1}^{N}(T_i - \bar{T}) - 10\, 1[n_a(a_i(t))]$$

Battery is the shared resource state that makes symbiosis observable to the learner.

ICPS findings: it works, but the reward is fragile

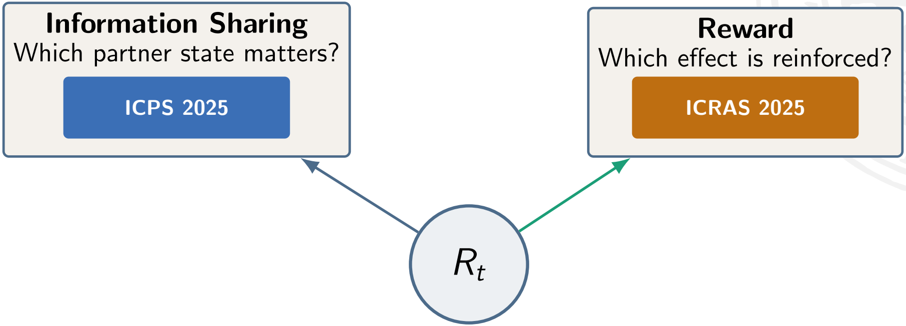


Sharing partner battery state improved coordination outcomes, but the rewards driving it are sparse and event-based, so training stays sensitive to reward tuning.

We exposes partner battery state

- Partner battery inferred to the observation and learning path
- Faster task completion, better workload and energy balance
- Reward signal stays sparse since it only fires at completion and low-battery events

Symbiosis scope

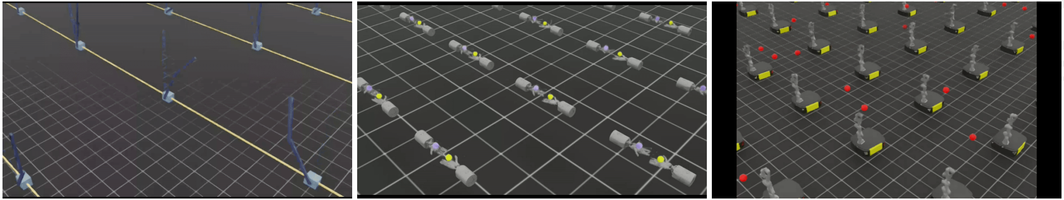


Symbiosis enters the decision layer through what robots **observe**, **optimize**, and **adapt**.

Task-specific mutual-benefit rewards

Symbiosis into **reward shaping** (ICRAS 2025):

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



$$\Delta P_{\text{cart}} = \epsilon_{\text{alive}}(1 - \delta_{\text{reset}}) + \epsilon_{\text{term}}\delta_{\text{reset}} + \epsilon_{\text{vel}}|\dot{x}_{\text{cart}}|,$$

$$\Delta P_{\text{pend}} = \epsilon_{\text{alive}}(1 - \delta_{\text{reset}}) + \epsilon_{\text{term}}\delta_{\text{reset}}$$

$$\Delta P_{\text{right}} = \epsilon_{\text{release}}(1 - \delta_{\text{fail}}),$$

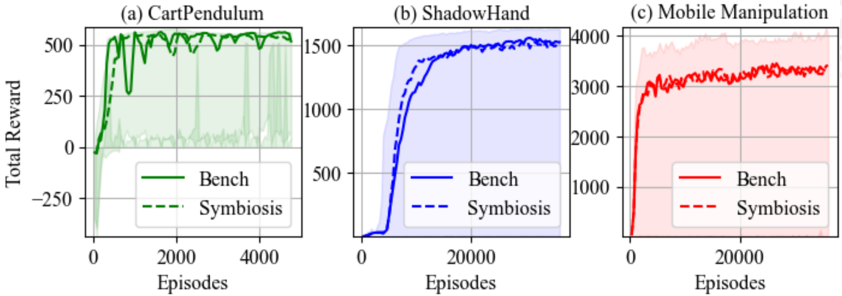
$$\Delta P_{\text{left}} = \epsilon_{\text{catch}}(1 - \delta_{\text{drop}})$$

$$\Delta P_{\text{base}} = -\epsilon_{\text{pos}}||\mathbf{p}^{xy}_{\text{ee}} - \mathbf{p}^{xy}_{\text{tgt}}||_2,$$

$$\Delta P_{\text{arm}} = -\epsilon_{\text{pos}}||\mathbf{q}_{\text{arm}} - \mathbf{q}_{\text{tgt}}||_2$$

MAPPO training under mutual-benefit shaping

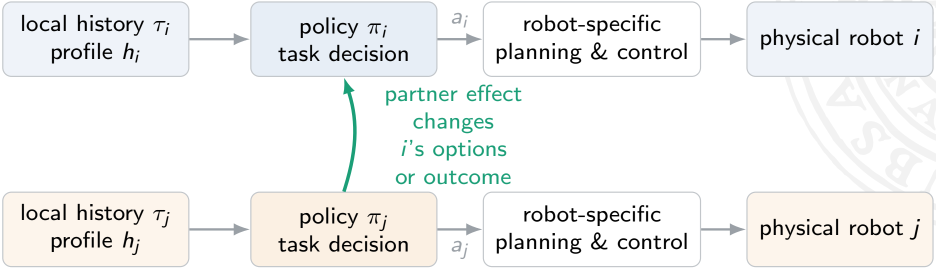
MAPPO with heterogeneous credit assignment lets mutual-benefit symbiosis emerge during training, but so far the evidence is qualitative and limited to a single robot pair.



We shapes rewards from partner effects

- Reward term reinforces one robot's effect on its partner
- Mutual-benefit behaviour emerges during MAPPO training
- Contribute to heterogeneous robots credit assignment

Local policies, coupled outcomes



$$\mathcal{G} = \langle N, S, \{A_i\}, T, \{\Omega_i\}, O, \{r_i\}, \gamma \rangle, \qquad a_{i,t} \sim \pi_i(\cdot \mid \tau_{i,t})$$

Heterogeneous: actions, observations, capabilities, and controllers may differ.

Partially observed: no robot sees the full joint state during execution.

Coupled: $T, \{r_i\}$ depend on the joint state-action even though each π_i conditions only on local τ_i.

Symbiosis is studied at the task-decision layer, might also inside planning & control.

Problem statement

The causal effect of one robot's actions on another's outcome is hard to model, and expensive to simulate or collect data for.

Working question:

How can inter-robot symbiosis be **represented, regulated, and adapted** to sustain collective function?

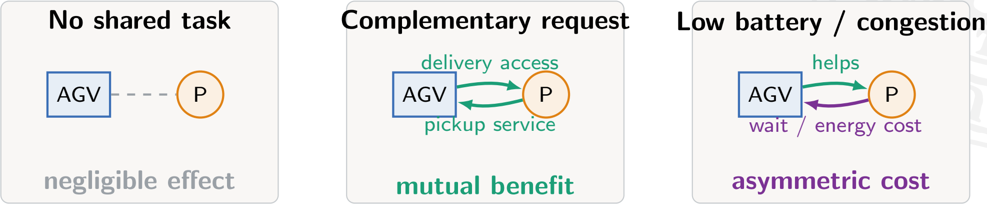
↓ **Formalised as directed partner dependence**

Thesis question:

How can **directed partner dependence** be represented and used to improve decentralised cooperation under changing roles, resources, and partners?

Partner relevance changes even the same tasks

One robot pair, three moments, same tasks:

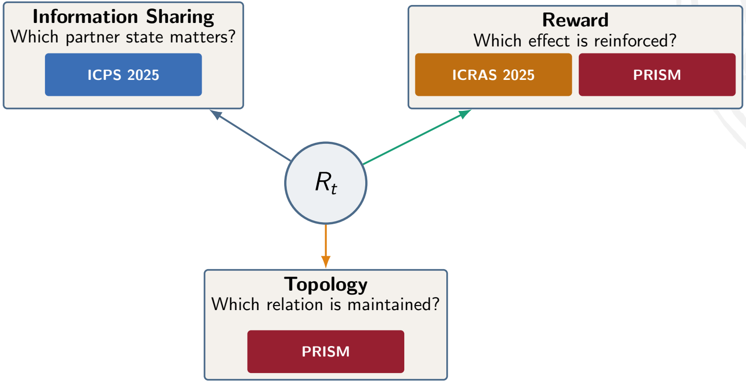


Shared reward: team did well, or not.

Not retained: who changed whose options, at whose cost.

The relation belongs to a robot pair, function, and context; it is not a permanent team label.

Operationalizing symbiosis



One compact directed relation enters three decentralized decisions.

Partner state changes local decisions

Local observation
own task, pose, and energy

↓ add partner state

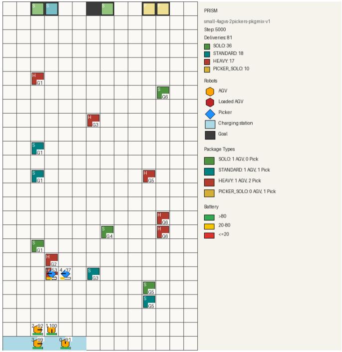
Relation-aware observation
local state + partner battery

Observed in the evaluated warehouse setting

- higher task completion;
- improved workload and energy balance.

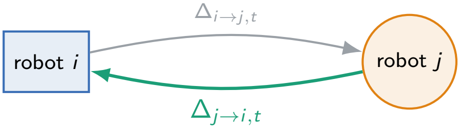
The useful variable is the one that changes attainable task value.

Task-Assignment Multi-Robot Warehouse



PRISM, IPPO policy: 4 AGVs, 2 pickers, mixed package types SOLO(1,0) PICKER_SOLO(0,1) STANDARD(1,1) HEAVY(1,2) — 81 deliveries over 5000 steps.

Directed partner dependence



$$\Delta_{j\to i,t} = J_i^H(t; a_t^j) - J_i^H(t; \bar{a}_t^j)$$

$$W_i(\pi) = \mathbb{E}_\pi\left[\sum_{t=0}^{\infty} \gamma^t r_i^{\text{task}}(s_t, a_t)\right].$$

Δ: counterfactual target; W_i: the fitness

Mutualism $(+,+)$	**Commensalism** $(+,0)$
Parasitism $(+,-)$	**Neutralism** $(0,0)$

Modelled as a heterogeneous, cooperative partially observed Markov game: MARL learns a joint policy π maximising team return, each $r_i(t)$ stays local but aligned via the symbiosis potential.

$$\widehat{\Delta W}_{i,t}^{(j)} \leftarrow (1-\alpha_{\text{ema}})\widehat{\Delta W}_{i,t-1}^{(j)} + \alpha_{\text{ema}}\underbrace{\psi_{ij}(t)}_{\text{proxy}}$$

PRISM makes relational effects learnable

Relation potential

$$\Phi_i(t) = \frac{1}{|\mathcal{N}_i^{\text{cross}}|} \sum_{j \in \mathcal{N}_i^{\text{cross}}} \underbrace{\phi(\mathsf{rel}_{ij}^t)\, \delta_{ij}^t}_{\widehat{\Delta W}_{i,t}^{(j)}}$$

Potential-difference shaping

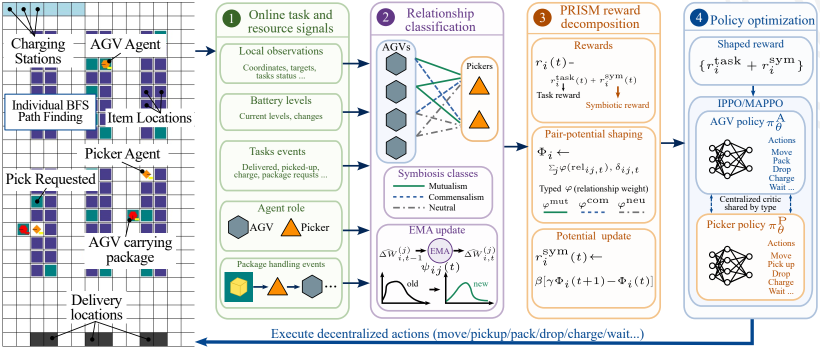
$$r_i'(t) = r_i^{\text{task}}(t) + \beta\left[\gamma \Phi_i(t+1) - \Phi_i(t)\right]$$

$\mathbf{rel}_{ij}$	typed interaction sign
δ_{ij}	estimated partner effect
ϕ	type-dependent treatment
$\Delta\Phi_i$	reinforces relational transitions

Task reward defines the objective; relational potential shapes how it is learned.

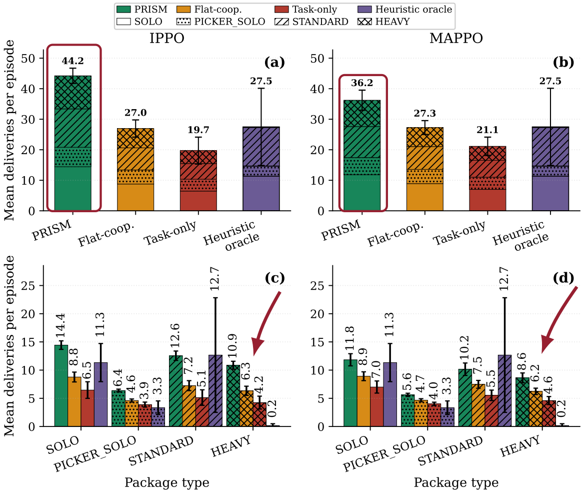
PRISM: events to pairwise proxy to shaped reward

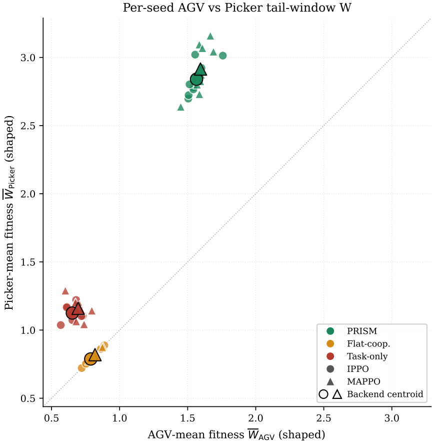
Events → typed, bounded proxy → shaped reward (PRISM, under submission).



$N = 6$: 4 AGVs, 2 pickers; a mutual environment, but with asymmetric contributions.

PRISM request decomposition results

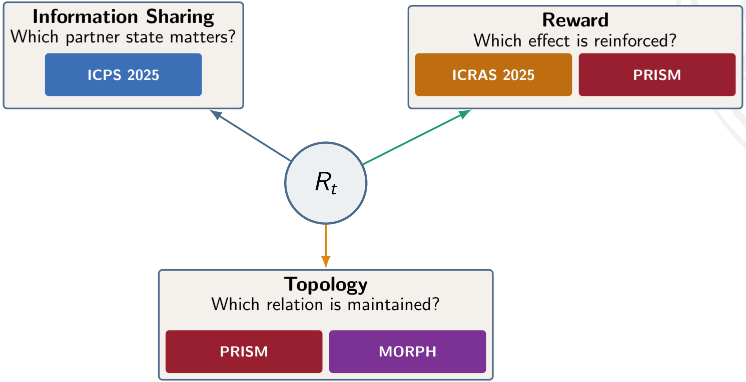




PRISM turns partner events into reward shaping

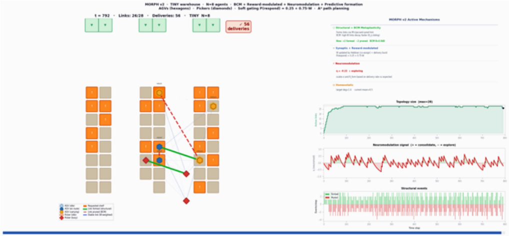
- Define robotic symbiosis through directed counterfactual fitness effects and provide an online relationship classifier
- Operationalizes fitness as a potential Φ_i — agents are rewarded by its change, not its level
- Fully decentralized IPPO (57.1 del/ep) beats MAPPO's centralized critic (52.1): decentralization itself helps more
- Recovers the task's own role specialisation automatically

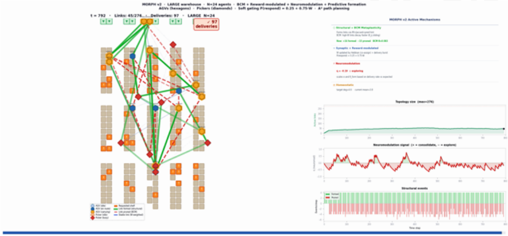
Research question and validation axes



One compact directed relation enters three decentralized decisions.

MORPH:Warehouse environment





TINY warehouse, N=8 agents LARGE warehouse, N=24 agents

TA-RWARE environment at two fleet scales, showing the adaptive topology (links formed/pruned), neuromodulation signal, and structural events unfold live over the episode.

MORPH: neural plasticity into cooperation

Four rules adapted from neural plasticity update the cooperation network online — no offline training or gradient derivation, just a warm start.

Rule	Plasticity form	Biological basis	What it updates
R1	Synaptic Plasticity	Hebbian LTP + dopamine gating (Hebb 1949; Dayan & Abbott 2001)	Preference W_{ij}
R2	Homeostatic Plasticity	Synaptic scaling (Turrigiano & Nelson 2004)	Per-agent degree bias h_i
R3	Structural Plasticity	Synaptogenesis + pruning (Holtmaat & Svoboda 2009)	Structural links $A_{ij} \in \{0, 1\}$
R4	Metaplasticity	BCM threshold (Bienenstock et al. 1982)	Sliding threshold θ^{bcm}, arousal η

MORPH: a smaller graph remains competitive

At $N = 18$ and 24: 10% to 13% of links, 102% to 105% of Full-Graph throughput, and 80% to 81% of Oracle.

Scale (N)	Method	Deliveries	%AP	%Or	Links (%max)
Tiny (8)	Proximity	55.9 ± 9.8	101%	95%	22.2 (79%)
	TSG	53.3 ± 8.8	96%	91%	2.6 (9%)
	MORPH	$\mathbf{53.3 \pm 9.6}$	**96%**	**91%**	**21.5 (77%)**
	Full-Graph	55.4 ± 8.6	100%	94%	28.0 (100%)
	Oracle	58.8 ± 11.0	106%	100%	—
Small (12)	Proximity	68.6 ± 9.7	106%	92%	19.6 (30%)
	TSG	64.9 ± 3.7	100%	87%	3.7 (6%)
	MORPH	$\mathbf{64.1 \pm 5.0}$	**99%**	**86%**	**22.8 (34%)**
	Full-Graph	64.6 ± 6.5	100%	86%	66.0 (100%)
	Oracle	74.9 ± 6.5	116%	100%	—
Medium (18)	Proximity	85.4 ± 12.8	101%	78%	24.0 (16%)
	TSG	86.3 ± 5.5	102%	78%	5.7 (4%)
	MORPH	$\mathbf{88.4 \pm 2.1}$	**105%**	**80%**	20.3 (13%)
	Full-Graph	84.2 ± 7.5	100%	77%	153.0 (100%)
	Oracle	$\mathbf{110.0 \pm 5.3}$	**131%**	**100%**	—
Large (24)	Proximity	104.3 ± 7.6	105%	83%	30.8 (11%)
	TSG	100.4 ± 3.7	101%	80%	7.9 (3%)
	MORPH	$\mathbf{101.3 \pm 6.5}$	**102%**	**81%**	**28.7 (10%)**
	Full-Graph	99.2 ± 3.4	100%	79%	276.0 (100%)
	Oracle	125.6 ± 3.1	127%	100%	—

MORPH adapts directed partner preferences

- Pairwise preferences and links adapt online — no offline training
- Competitive throughput using a fraction of the links
- Trails the Hungarian oracle since it re-solves a fresh, fully-connected assignment every step, while MORPH commits to a temporary sparse graph

Future Work!

Ongoing works:

- (Real robot) Learning? to Cooperate with New Partners in Open Heterogeneous Robot Teams
- (Warehouse fleets) Joint Task and Charging Coordination under Directed Inter-Robot Dependencies
- (Human-robot symbiosis) Human-Aware Adaptation of Robot Roles under Effectiveness and Efficiency Failures

Publications

Presented work

- X. Niu, N. C. Barajas and D. G. Broo, "Enabling Symbiosis in Multi-Robot Systems Through Multi-Agent Reinforcement Learning," 2025 IEEE 8th International Conference on Industrial Cyber-Physical Systems (ICPS), Emden, Germany, 2025, pp. 1-7.
- X. Niu and D. G. Broo, "Investigating Symbiosis in Robotic Ecosystems: A Case Study for Multi-Robot Reinforcement Learning Reward Shaping," 2025 9th International Conference on Robotics and Automation Sciences (ICRAS), Osaka, Japan, 2025, pp. 112-117.
- X. Niu, D. G. Broo, "PRISM: Policy-Shaping via Reward Decomposition for Inter-Agent Symbiosis in Multi-Robot Cooperation," manuscript under review at The International Journal of Robotics Research (IJRR), 2026.
- X. Niu, D. G. Broo, "MORPH: Self-Organising Multi-Robot Task Allocation via Neuroplasticity-Inspired Adaptive Topology," accepted to the 23rd European Conference on Multi-Agent Systems (EUMAS) 2026.

Publications

Other work supports the broader research context but is not included

- A. Rouchitsas, X. Niu, D. G. Broo, "Dynamics of Human Adaptability to Robot Effectiveness and Efficiency Malfunctions in Collaborative Manufacturing," manuscript under review at Robotics and Computer-Integrated Manufacturing, 2026.
- J. Xu, X. Niu, D. Gürdür Broo, K. Hjort, "TouchDrive: Electronics-Free Tactile Sensing Interface for Assistive Grasping," accepted for RoboTac workshop, IEEE ICRA, 2026.
- J. Xu, X. Niu, J. Andersson, D. G. Broo, K. Hjort, "Miniaturized Multifunctional Valves for Intelligent Pneumatic Systems in Soft Robotics," manuscript under review at Advanced Intelligent Systems, 2026.
- A. Rouchitsas, X. Niu, G. Castellano, D. G. Broo, " What do I do now?: Spontaneous Human Responses to Robot Effectiveness and Efficiency Malfunctions in Collaborative Robotics," In Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), ACM, 2026.

Ending

Thank you for listening!

Questions?