DeepMind

Modelling Cooperation in Network Games with Spatio-Temporal Complexity

Michiel A. Bakker*, Richard Everett*, Laura Weidinger, Iason Gabriel, William S. Isaac, Joel Z. Leibo, Edward Hughes

Full paper: https://arxiv.org/abs/2102.06911

* Core Contributors

Introduction

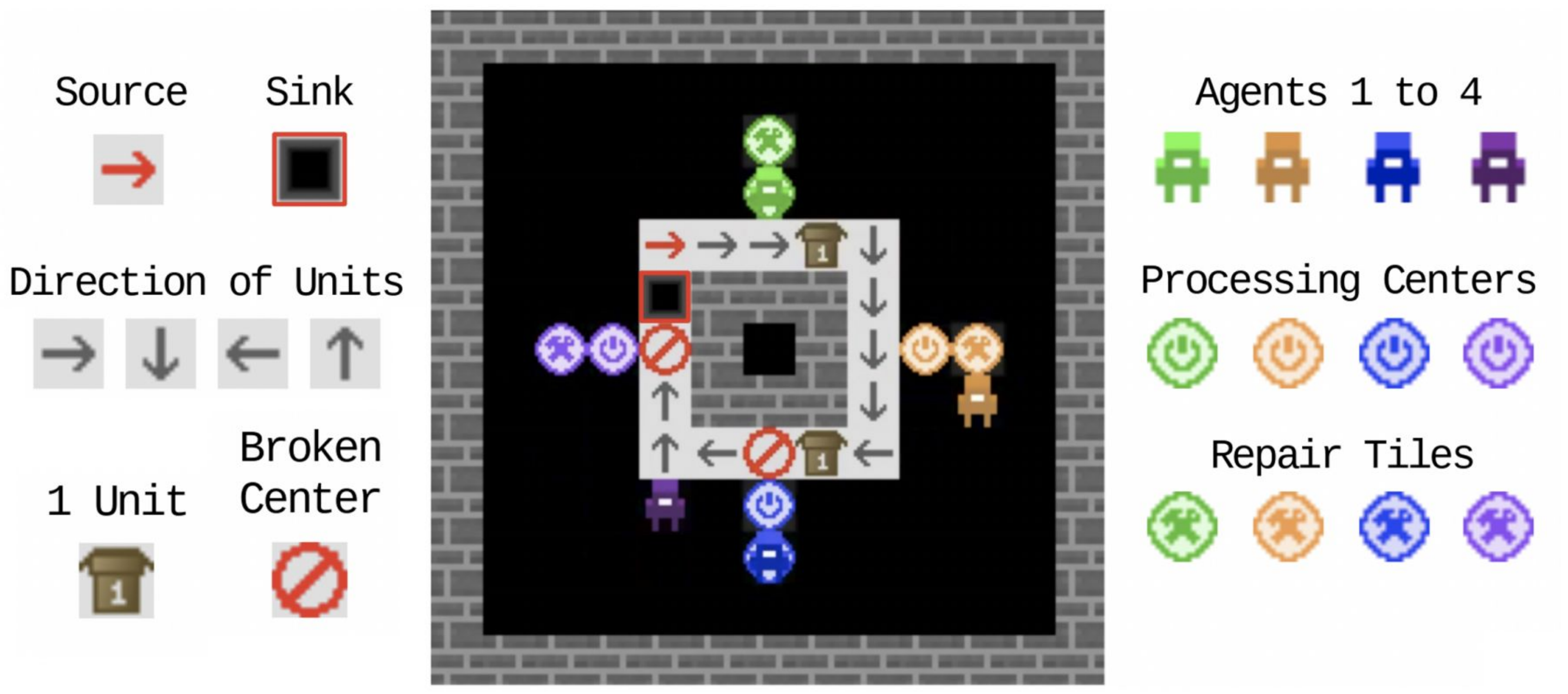
Introduction

- In real-world collective action problems, incentives are often misaligned and multi-agent cooperation is necessary to ensure beneficial outcomes.
- These real-world collective action problems often have a latent graph structure, like computer networks, irrigation systems and road networks.
- The field of network games studies interactions between agents in graph structures but currently abstracts away important details such as geometry and time.
- System designers benefit from models that predict or explain the effects of planned interventions, such as altering the layout of a supply chain.

Contributions

- We apply multi-agent reinforcement learning to spatially and temporally extended collective action problems with an embedded graph structure.
- Using a set of new analytical tools to measure social outcomes, we study the implications of different interventions in the environment and the agent population.
- We vary the topology of the world, as in traditional network games, but also the geometry, maintenance cost, and agent specialization (please see the full ArXiv paper for all results).

The Supply Chain environment



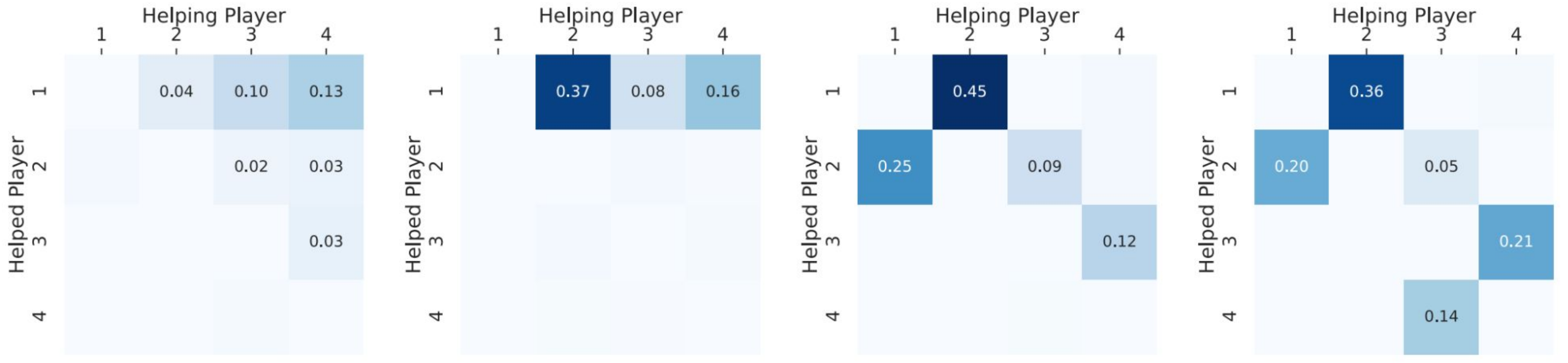
The environment is a 2D gridworld with an underlying graph structure in which agents must process units while maintaining their own individual processing centers with the help of other agents.

- Units enter at the source tile, follow the edges of the supply chain, and finally leave the environment at the sink tiles. On each of the 1000 steps in an episode there is 10% chance of a unit entering.
- Agents process units by standing on their processing tile which gives them +1 reward.
- If a unit moves while the next space is occupied, a unit gets discarded, leading to lost opportunity for reward.
- Upon processing, there is a 25% chance that a processing center breaks down, after which two agents are necessary to repair the center.

Collective action is required to maintain the processing centers: each agent prefers other agents to take on the responsibility for fixing broken centers, since leaving their station to repair comes at an opportunity cost. However, if all agents refuse to cooperate, all workstations end up being broken and the agents receive low collective reward.

Emergence of Care

- Agents need to cooperate with other agents to ensure their own processing centers are repaired. However, repairing does not benefit agents directly and is thus a complex behavior to learn.
- In the figure, we show that emergence of care happens in four distinct phases. First only upstream care emerges, after which reciprocal care emerges first between agents 1 and 2, and then between agents 3 and 4.
- One might have expected that a model-based intervention is necessary to solve the social dilemma. However, we observe that the underlying graph structure promotes the emergence of reciprocity. Note that reciprocity (tit-for-tat) is a fundamentally temporal strategy that would not emerge in a one-shot network game.

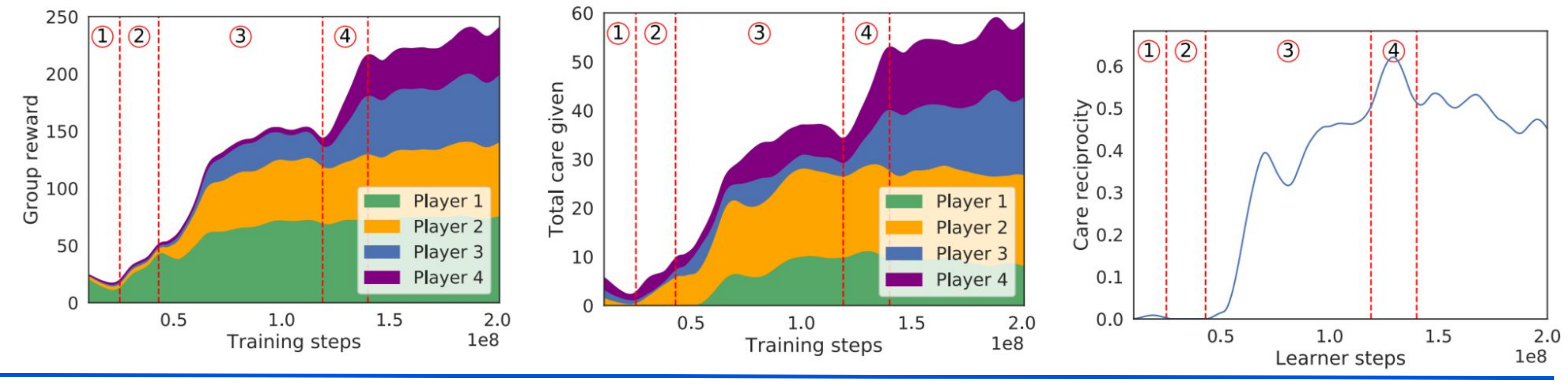


Phase 1. Learning to process units

Phase 2. First upstream care

Phase 3. First reciprocal care

Phase 4. Full reciprocal care



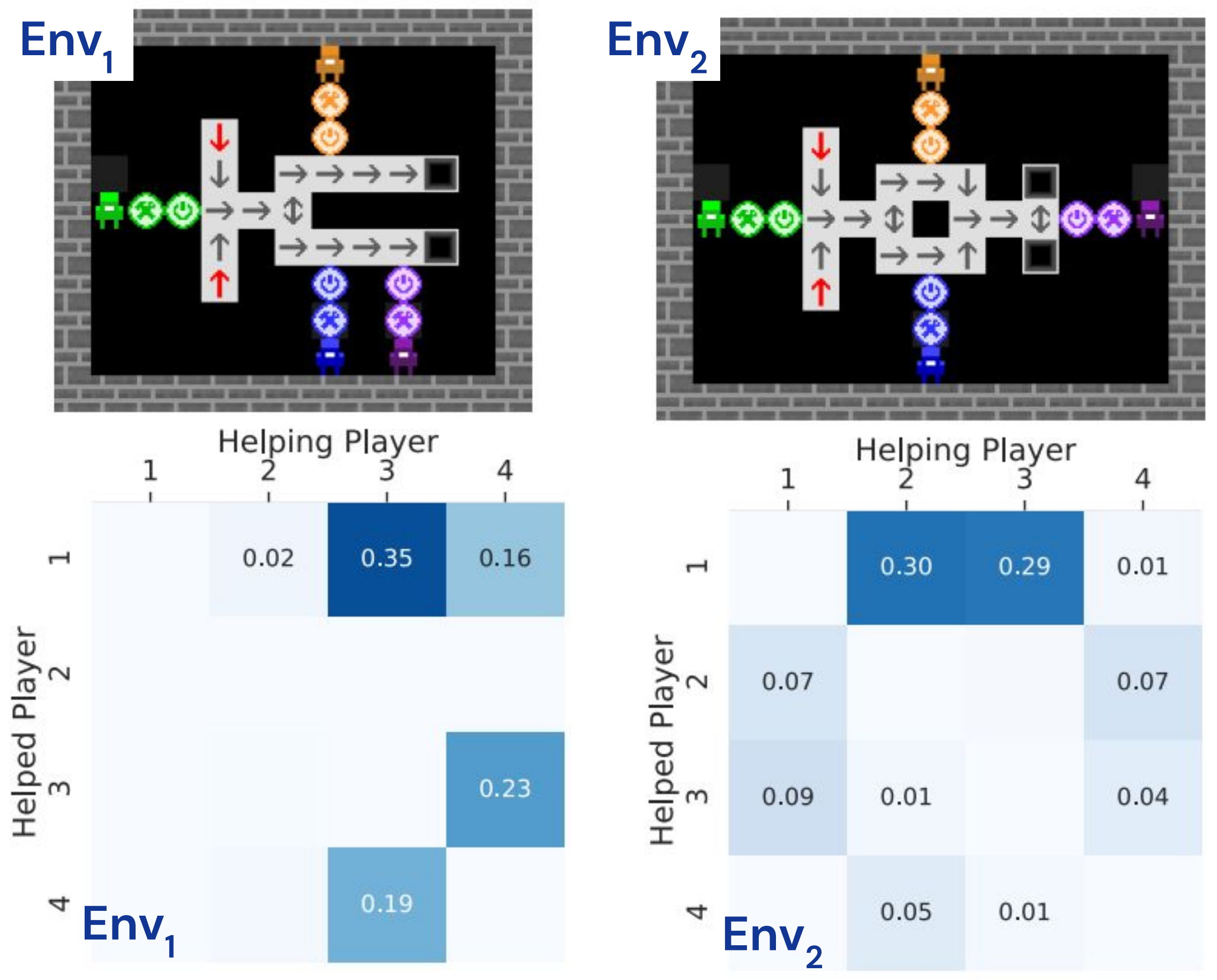
Topological intervention

We investigate the effect of different underlying graph structures on the social outcomes.

- **Environment 1 (left) –** In the top branch, agent 2 earns little reward as no other agent has an incentive to care for its workstation. In contrast, in the bottom branch, the agents receive similar reward as they learn to reciprocate care.
- **Environment 2 (right) –** All agents earn reward but the amount varies strongly as there are multiple stable outcomes. For example, agent 1 can establish reciprocity with agent 2 or agent 3.

A mechanism designer might be interested in maximizing efficiency of the supply chain (outgoing units relative to incoming units), highest for environment 1.

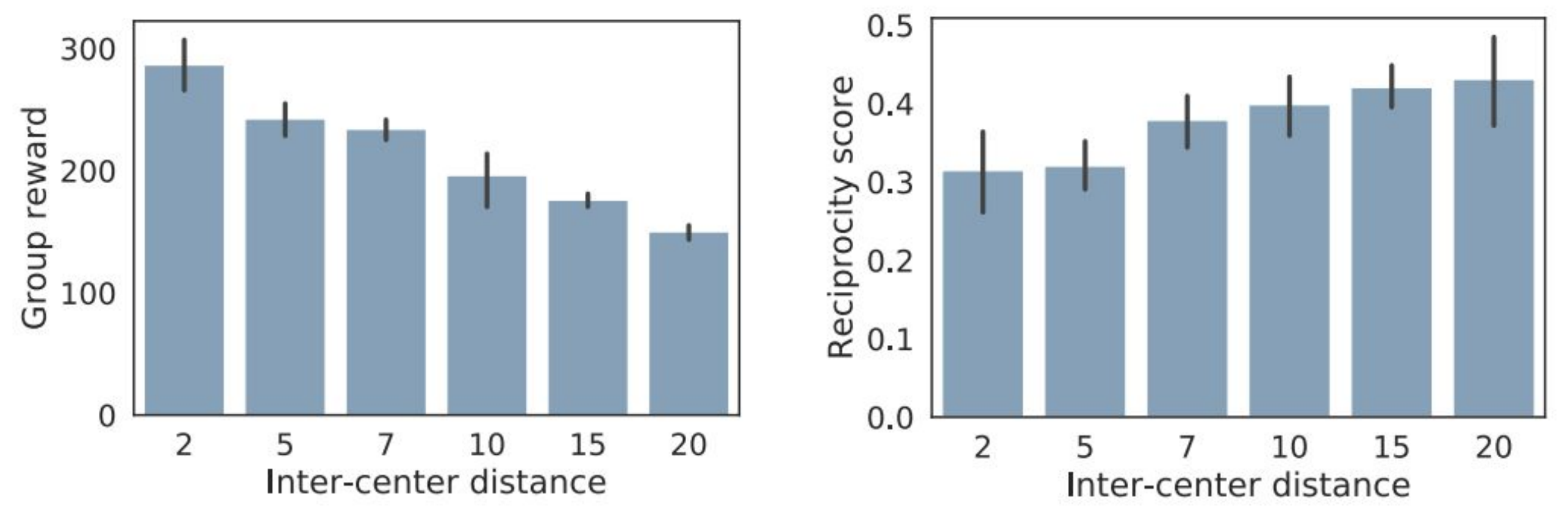
	Env_1	Env_2
R_1	125 ± 2	119 ± 5
R_2	4.0 ± 0.5	31 ± 15
R_3	55 ± 0.9	31 ± 15
R_4	48 ± 2	17 ± 14
$\Sigma_i R_i$	234 ± 4	198 ± 21
Eff.	26% ± 2%	8% ± 6%



Geometric intervention

In contrast to pure network analysis, multi-agent RL provides tools for measuring the impact of geometric changes. We vary the distance between processing centers between 2 and 20 tiles.

- Naturally, group reward decreases as the distance increases.
- Interestingly, the distance also influences the dynamics of care. Longer distances increase the effective "cost" of caring which makes reciprocity more important.



Methods

Social outcome metrics

- Care matrix with elements C_{ij} tracks care between agents i and j normalized by the number of breakages.
- Care reciprocity measures how symmetric the care matrix is.
- Care direction measures whether care is, on average, more upstream (D=1) or more downstream (D=-1).

Agents

Advantage actor-critic (A3C) with 400 parallel environments and a population of 8 agents randomly assigned to 4 processing centers.

Architecture

For each 13x13 RGB observation, the neural network computes a policy and value. It consists of a 2D conv net, a dense layer and an LSTM.