An Axiomatic Framework for Reinforcement Learning in Open Multi-Agent Systems

Learning individual values for cooperation with unfamiliar teammates

Jianhong Wang, Yang Li, Samuel Kaski, Jonathan Lawry

University of Bristol Shanghai Jiao Tong University ELLIS Institute Finland
Aalto University University of Manchester

Cooperation with unfamiliar teammates

n-agent ad hoc teamwork (NAHT)

N

Controlled agents C

We learn their policies.

M − N

Uncontrolled teammates U

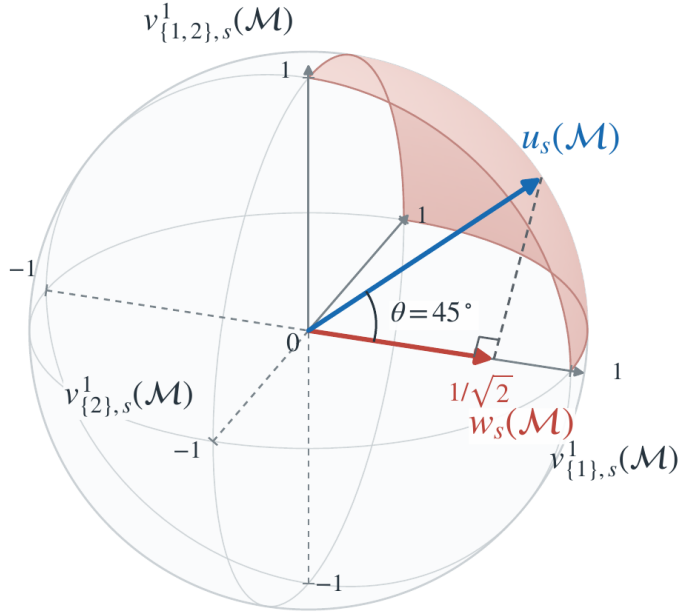
Their policies are fixed.
Team compositions can change.

Full team: C ∪ U, with M agents

Objective: maximise expected discounted team rewards

Internal cooperation and external teamwork

$$v_s(\mathcal{M}) = \underbrace{\sum_{C_I \in \mathcal{P}_+(\mathcal{C})} k_{C_I} v^1_{C_I,s}(\mathcal{M})}_{\text{Internal Cooperation}} + \underbrace{\sum_{C_E \in \mathcal{P}_+(\mathcal{M}) \setminus \mathcal{P}_+(\mathcal{C})} k_{C_E} v^1_{C_E,s}(\mathcal{M})}_{\text{External Teamwork}}.$$



Two agents: C = {1}, U = {2}

Internal cooperation

Coalitions of controlled agents only

External teamwork

Every coalition involving an uncontrolled agent

Axioms guide individual value learning

$V_i(s) = \varphi_i(v_s)$ φ maps cooperative games to individual values

Linearity

$\varphi(av + bw) = a\varphi(v) + b\varphi(w)$

Symmetry

Equal contributions receive equal values

Efficiency

$\Sigma_i \varphi_i(v) = v(C \cup U)$

Truncated TD(λ)-style targets

Weighted bootstrapped returns

Value function structure

Consistent treatment under agent relabelling

Reward shaping and a regularizer

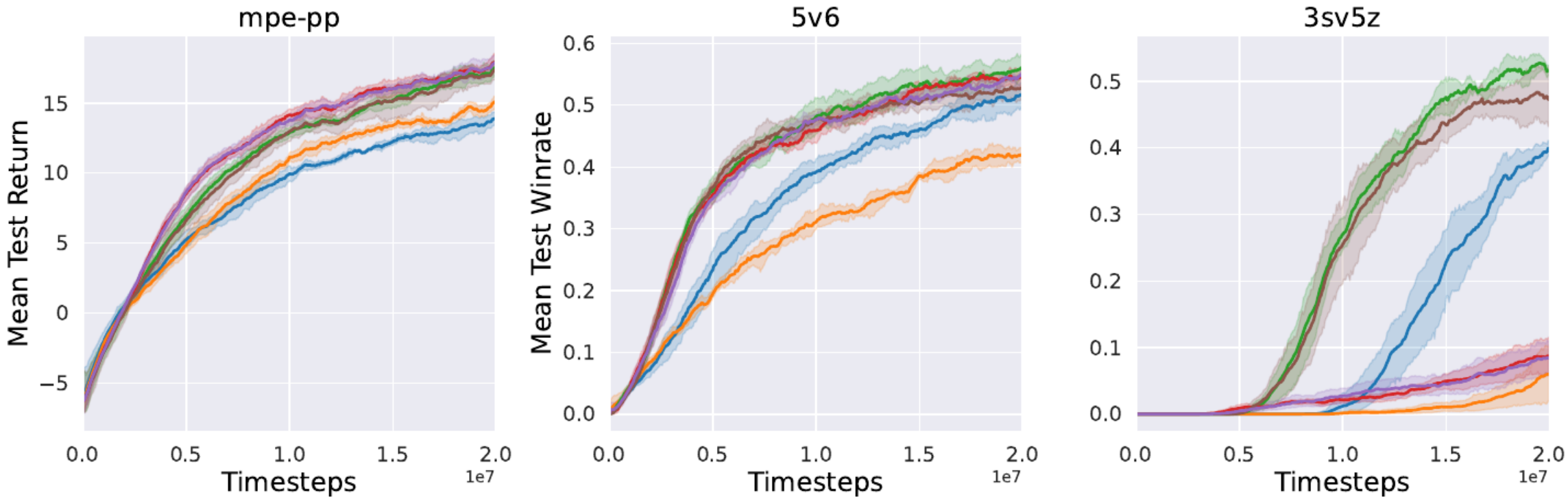
Allocations account for the full team value

SM uses all three axioms. BM omits Efficiency.

Both augment POAM or IPPO-NAHT using architecture and learning losses.

Axiomatic variants improve learning

Selected tasks from MPE-PP and SMAC



POAM **IPPO-NAHT** **SM-POAM** **SM-IPPO** **BM-POAM** **BM-IPPO**

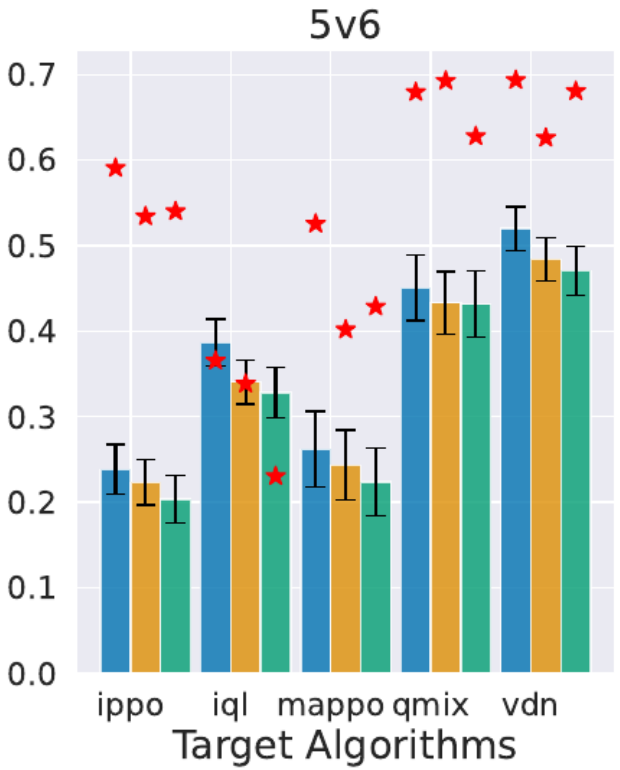
The benefit of Efficiency varies by task.

Five seeds. 128 evaluation episodes per metric. 95% confidence intervals.

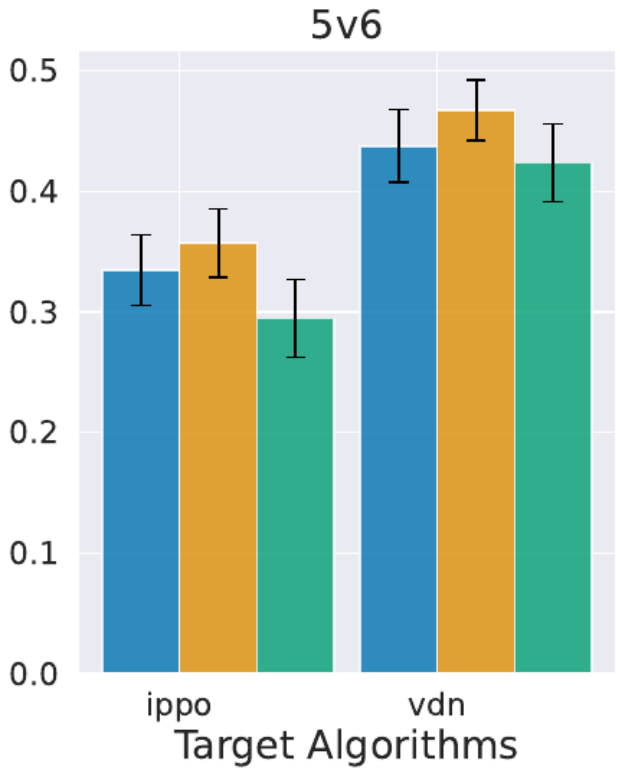
Generalisation depends on the teammate shift

5v6: mean test win rate under two teammate shifts

Unseen conventions



Unseen agent types



SM-POAM generally helps with unseen conventions.

BM-POAM can be competitive or better with unseen agent types.

SM-POAM **BM-POAM**

POAM

Red stars: in-convention baselines

Cooperative game structure guides learning with unfamiliar teammates.

Paper: arxiv.org/abs/2611.03301

Code: github.com/hsvgbkhgbv/axiomatic-rl-naht