Towards Multiagent Coordination under Multiple Objectives and Sparse Rewards

Doctoral Consortium

Raghav Thakar
Collaborative Robotics and Intelligent Systems (CoRIS) Institute
Oregon State University
Corvallis, Oregon, USA
thakarr@oregonstate.edu

ABSTRACT

Real-world deployment demands that autonomous agents coordinate effectively, balance conflicting objectives, and learn from sparse feedback. However, current research typically addresses these challenges in isolation. My thesis proposes a unified framework to bridge this gap, building upon two key contributions: the Multi-Objective Difference Evaluation (D_{MO}) operator, and the Mixed Advantage Pareto Extraction (MAPEX) algorithm. By leveraging these methods, I aim to establish a general framework for multi-objective multiagent learning in sparse-reward environments.

KEYWORDS

multi-objective reinforcement learning; multiagent reinforcement learning; reward shaping; sparse rewards; evolutionary algorithms

ACM Reference Format:
Raghav Thakar. 2026. Towards Multiagent Coordination under Multiple Objectives and Sparse Rewards: Doctoral Consortium. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 3 pages. https://doi.org/10.65109/PGON3968

1 INTRODUCTION

Recent AI breakthroughs have mastered complex domains, from defeating champions in Go [16], and achieving superhuman performance in car racing [19], to mastering diverse suites of arcade games [3]. Yet, these successes often rely on privileges unavailable in the real world: operation in isolation, singular, well-defined goals, and dense and informative feedback. Real-world deployment, conversely, demands agents that can coordinate, balance conflicting objectives, and learn from sparse feedback. Consequently, my thesis focuses on enabling multiagent learning in environments characterised by multiple objectives and varying degrees of reward sparsity.

Prior work has enhanced multiagent reinforcement learning (MARL) algorithms for the sparse-reward setting, notably through shaped rewards that provide intrinsic motivation to explore [2, 5, 14], promote joint-policy entropy [20], or assign potentials to states [10]. For multi-objective problems, prior work has extended the notable MARL algorithm Q-MIX [15] to produce a Pareto front of joint-policies [11]. More recently, efforts have been made to reconcile the learning challenges of multiagent systems and multi-objective settings in continuous control [6].

While these works provide essential foundations, they are results of isolated pursuits. There is a lack of integrated approaches that leverage a combined perspective on these sub-fields and produce general, coordinated agents capable of navigating the intersection of multiple objectives and sparse feedback.

My thesis approaches this broader problem by building the individual components necessary to bridge this gap. Thus far, my contributions have been:

(1) the **Multi-Objective Difference Evaluation (D_{MO}) Operator** [17], which addresses the multiagent credit assignment problem in multi-objective settings, and
(2) the preliminary, single-agent **Mixed Advantage Pareto Extraction (MAPEX) Algorithm** [18], which extracts policies that balance multiple objectives, from starting policies that were trained on singular, disjoint objectives.

The remainder of my thesis work will focus on the synthesis of these prior contributions. Specifically, I seek to develop one of the first sparse-reward multi-objective reinforcement learning (MORL) algorithms, leveraging the algorithmic flexibility of MAPEX. Subsequently, I aim to propose a unified framework for multi-objective multiagent learning under reward sparsity. This framework would combine the recent advances in sparse-reward MARL, with the ability of MAPEX to produce multi-objective solutions from single-objective starting points, with the help of D_{MO} to fine-tune these solutions further.

2 BACKGROUND

Multi-Objective Optimisation. Multi-objective problems return a vector of rewards, requiring us to find a set of *Pareto-optimal* policies rather than a single solution. A solution v *Pareto-dominates* u (denoted $v \succ_p u$) if v is strictly better in at least one objective and no worse in others. The set of non-dominated solutions forms the *Pareto front*, often evaluated using the *hypervolume* indicator, which measures the volume of objective space covered by the set relative to a reference point.

Multiagent Credit Assignment. The *credit assignment problem* requires isolating an individual agent's contribution to global team performance. Difference Evaluation (D) addresses this by estimating an agent's marginal impact [1, 8, 9]. Formally, D_i for agent i compares global performance $G(z)$ to a counterfactual $G(z_{-i} \cup c_i)$

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/PGON3968

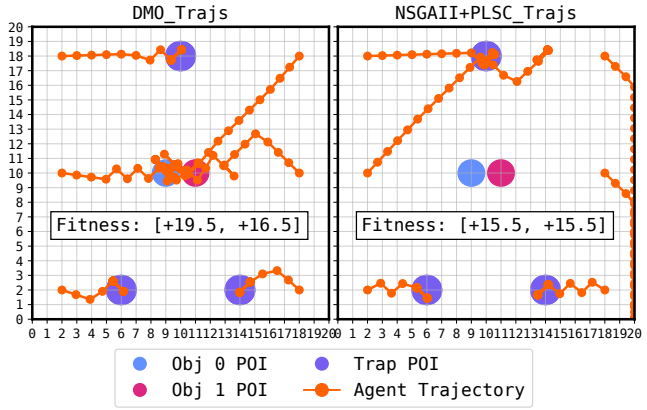

Figure 1: D_{MO} agent trajectories compared to an ablated baseline in the rover exploration domain. The central targets are high-reward, but require synchronised observation by each agent. D_{MO} successfully discovers these harder targets and shows optimal behaviour.

where agent i is replaced by a default action c_i. This calculation isolates the agent's specific contribution to the system.

3 MULTI-OBJECTIVE CREDIT ASSIGNMENT

Multiagent credit assignment is challenging in multi-objective settings due to credit having to be computed for each objective. Existing methods typically rely on *a priori* scalarisation, collapsing the multi-objective reward vector into a scalar, which imposes preferences and may overlook nuanced non-linear trade-offs [13, 21].

To address this, I developed the Multi-Objective Difference Evaluation (D_{MO}) operator [17]. Rather than relying on scalarised rewards, D_{MO} derives agent-specific credit by measuring the agent's marginal contribution to the hypervolume H of the Pareto front:

$$D_{\text{MO}}(\pi_i, \pi, \mathcal{T}) = H(\mathcal{T}) - H(\mathcal{T}') \tag{1}$$

where $\mathcal{T}$ is the set of joint trajectories generated by a population of joint-policies, and $\mathcal{T}'$ is the modified set where the trajectory π_i of agent i is replaced by a counterfactual default. The greater the D_{MO} value, the greater is agent i's contribution to the hypervolume.

Empirical evaluations on the Multi-Objective Rover Exploration domain demonstrated that D_{MO} delivers up to a 20% performance improvement in sparse-reward scenarios, validating the utility of accurate credit assignment in complex coordination tasks. Figure 1 shows D_{MO} successfully learning optimal team coordination while baselines converge to suboptimal behaviours.

4 MIXED ADVANTAGE PARETO EXTRACTION

In many real-world scenarios, multi-objective preferences arise retroactively, often after an agent has already been trained to master its primary task. Current MORL methods typically require training from scratch to learn a Pareto front, failing to leverage pre-trained 'specialist' policies [4, 7, 12]. To resolve this inefficiency, I developed Mixed Advantage Pareto Extraction (MAPEX) [18], an offline MORL method that constructs a frontier of policies by reusing disjoint, pre-trained specialist policies, critics, and replay buffers.

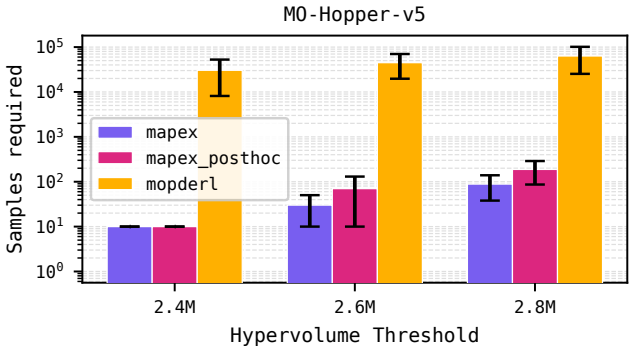

Figure 2: Number of samples consumed by Pareto extraction algorithms to produce Pareto fronts of various threshold hypervolumes from the same starting policies.

MAPEX operates on the insight that optimal trade-off behaviours can be synthesised by blending expert behaviours from single-objective specialists. First, we identify gaps in the current Pareto front estimate and derive a target weight vector w_{target} to fill that gap. We then construct a static 'hybrid buffer' by sampling transitions from the specialists' buffers in proportion to these weights. We then compute a *mixed advantage* signal, A_{mixed}, for these transitions:

$$A_{mixed}(s, a) = w_{target}^{\top} \cdot A(s, a) \tag{2}$$

where $A(s, a)$ is an advantage vector, comprising evaluations of each single-objective critic. This mixed advantage is used to weight a supervised regression loss, cloning actions that contribute to the target trade-off. Evaluations on continuous control MuJoCo benchmarks show that MAPEX produces Pareto fronts comparable to established baselines while reducing the sample cost by three orders of magnitude (0.001% of the baseline) when starting from the same pre-trained specialists (Figure 2).

5 PROPOSED WORK

The research conducted thus far has established two critical pillars: D_{MO} for effective multi-objective credit assignment, and MAPEX for efficient Pareto extraction. The remainder of this thesis focuses on the synthesis of these components to address the intersection of multi-objective, multiagent, and sparse-reward learning.

I aim to develop a unified framework that leverages the algorithmic flexibility of MAPEX to circumvent the difficulties of learning complex trade-offs from scratch in sparse-reward environments. Rather than dealing with the compounding difficulty of sparse rewards and learning multi-objective trade-off behaviours simultaneously, I propose to train policies on singular, sparse objectives using existing methods. With MAPEX, these 'specialists' can then be blended to construct a Pareto front, decoupling the challenge of feedback sparsity from the challenge of multi-objective learning.

Subsequently, I will extend this methodology to the multiagent domain. This entails adapting the extraction mechanism to operate over joint-policies and leveraging D_{MO} to fine-tune the resulting solutions, ensuring an even coverage of the objective space. Consequently, my thesis seeks to provide the first general multi-objective multiagent learning framework robust to reward sparsity.

REFERENCES

[1] Adrian Agogino, Kagan Tumer, and Risto Miikkulainen. 2005. Efficient credit assignment through evaluation function decomposition. In *Proceedings of the 7th Annual Conference on Genetic and Evolutionary Computation* (Washington DC, USA) *(GECCO '05)*. Association for Computing Machinery, New York, NY, USA, 1309–1316. https://doi.org/10.1145/1068009.1068221

[2] Ayhan Alp Aydeniz, Enrico Marchesini, Robert Loftin, and Kagan Tumer. 2023. Entropy Maximization in High Dimensional Multiagent State Spaces. In *2023 International Symposium on Multi-Robot and Multi-Agent Systems (MRS)*. 92–99. https://doi.org/10.1109/MRS60187.2023.10416789

[3] Adrià Puigdomènech Badia, Bilal Piot, Steven Kapturowski, Pablo Sprechmann, Alex Vitvitskyi, Daniel Guo, and Charles Blundell. 2020. Agent57: outperforming the Atari human benchmark. In *Proceedings of the 37th International Conference on Machine Learning (ICML'20)*. JMLR.org, Article 48, 11 pages.

[4] Toygun Basaklar, Suat Gumussoy, and Umit Ogras. 2023. PD-MORL: Preference-Driven Multi-Objective Reinforcement Learning Algorithm. In *The Eleventh International Conference on Learning Representations*. https://openreview.net/forum?id=zS9sRyaPFlJ

[5] Yuri Burda, Harri Edwards, Deepak Pathak, Amos Storkey, Trevor Darrell, and Alexei A. Efros. 2019. Large-Scale Study of Curiosity-Driven Learning. In *International Conference on Learning Representations*. https://openreview.net/forum?id=rJNwDjAqYX

[6] Adam Callaghan, Karl Mason, and Patrick Mannion. 2026. MOMA-AC: A preference-driven actor-critic framework for continuous multi-objective multi-agent reinforcement learning. *Neurocomputing* 664 (2026), 132032. https://doi.org/10.1016/j.neucom.2025.132032

[7] Xi Chen, Ali Ghadirzadeh, Mårten Björkman, and Patric Jensfelt. 2019. Meta-Learning for Multi-objective Reinforcement Learning. In *2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)* (Macau, China). IEEE Press, 977–983. https://doi.org/10.1109/IROS40897.2019.8968092

[8] Mitchell Colby and Kagan Tumer. 2012. Shaping fitness functions for coevolving cooperative multiagent systems, Vol. 1. 425–432.

[9] Joshua Cook, Kagan Tumer, and Tristan Scheiner. 2023. Leveraging Fitness Critics To Learn Robust Teamwork. In *Proceedings of the Genetic and Evolutionary Computation Conference* (Lisbon, Portugal) *(GECCO '23)*. Association for Computing Machinery, New York, NY, USA, 429–437. https://doi.org/10.1145/3583131.3590497

[10] Sam Devlin and Daniel Kudenko. 2011. Theoretical considerations of potential-based reward shaping for multi-agent systems. In *The 10th International Conference on Autonomous Agents and Multiagent Systems - Volume 1* (Taipei, Taiwan) *(AAMAS '11)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 225–232.

[11] Tianmeng Hu, Biao Luo, Chunhua Yang, and Tingwen Huang. 2023. MO-MIX: Multi-Objective Multi-Agent Cooperative Decision-Making With Deep Reinforcement Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 10 (Oct. 2023), 12098–12112. https://doi.org/10.1109/TPAMI.2023.3283537

[12] Erlong Liu, Yu-Chang Wu, Xiaobin Huang, Chengrui Gao, Ren-Jian Wang, Ke Xue, and Chao Qian. 2025. Pareto set learning for multi-objective reinforcement learning. In *Proceedings of the Thirty-Ninth AAAI Conference on Artificial Intelligence and Thirty-Seventh Conference on Innovative Applications of Artificial Intelligence and Fifteenth Symposium on Educational Advances in Artificial Intelligence (AAAI'25/IAAI'25/EAAI'25)*. AAAI Press, Article 2095, 9 pages. https://doi.org/10.1609/aaai.v39i18.34068

[13] Patrick Mannion, Karl Mason, Sam Devlin, Jim Duggan, and Enda Howley. 2016. Multi-Objective Dynamic Dispatch Optimisation using Multi-Agent Reinforcement Learning: (Extended Abstract). In *Proceedings of the 2016 International Conference on Autonomous Agents & Multiagent Systems* (Singapore, Singapore) *(AAMAS '16)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1345–1346.

[14] Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. 2017. Curiosity-driven exploration by self-supervised prediction. In *Proceedings of the 34th International Conference on Machine Learning - Volume 70* (Sydney, NSW, Australia) *(ICML'17)*. JMLR.org, 2778–2787.

[15] Tabish Rashid, Mikayel Samvelyan, Christian Schroeder De Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. 2020. Monotonic value function factorisation for deep multi-agent reinforcement learning. *J. Mach. Learn. Res.* 21, 1, Article 178 (Jan. 2020), 51 pages.

[16] David Silver, Aja Huang, Christopher Maddison, Arthur Guez, Laurent Sifre, George Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, and Demis Hassabis. 2016. Mastering the game of Go with deep neural networks and tree search. *Nature* 529 (01 2016), 484–489. https://doi.org/10.1038/nature16961

[17] Raghav Thakar, Gaurav Dixit, Siddarth Iyer, and Kagan Tumer. 2025. Multiagent Credit Assignment for Multi-Objective Coordination. In *Proceedings of the Genetic and Evolutionary Computation Conference* (NH Malaga Hotel, Malaga, Spain) *(GECCO '25)*. Association for Computing Machinery, New York, NY, USA, 663–672. https://doi.org/10.1145/3712256.3726445

[18] Raghav Thakar, Gaurav Dixit, and Kagan Tumer. 2026. Post Hoc Extraction of Pareto Fronts for Continuous Control. arXiv:2603.02628 [cs.LG] https://arxiv.org/abs/2603.02628

[19] Peter Wurman, Samuel Barrett, Kenta Kawamoto, James MacGlashan, Kaushik Subramanian, Thomas Walsh, Roberto Capobianco, Alisa Devlic, Franziska Eckert, Florian Fuchs, Leilani Gilpin, Piyush Khandelwal, Varun Kompella, HaoChih Lin, Patrick MacAlpine, Declan Oller, Takuma Seno, Craig Sherstan, Michael Thomure, and Hiroaki Kitano. 2022. Outracing champion Gran Turismo drivers with deep reinforcement learning. *Nature* 602 (02 2022), 223–228. https://doi.org/10.1038/s41586-021-04357-7

[20] Pei Xu, Junge Zhang, and Kaiqing Huang. 2023. Exploration via joint policy diversity for sparse-reward multi-agent tasks. In *Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence* (Macao, P.R.China) *(IJCAI '23)*. Article 37, 9 pages. https://doi.org/10.24963/ijcai.2023/37

[21] Logan Yliniemi and Kagan Tumer. 2014. Multi-objective Multiagent Credit Assignment Through Difference Rewards in Reinforcement Learning. In *Simulated Evolution and Learning*, Grant Dick, Will N. Browne, Peter Whigham, Mengjie Zhang, Lam Thu Bui, Hisao Ishibuchi, Yaochu Jin, Xiaodong Li, Yuhui Shi, Pramod Singh, Kay Chen Tan, and Ke Tang (Eds.). Springer International Publishing, Cham, 407–418.