Learning Empathy in Multi-Agent Social Dilemmas: Cooperation, Inequality, and Stability

Christina Christodoulou*[1] Simone De Giorgi*[1] Lavinia Skandali*[1]

[1]Bocconi Students for Machine Learning, Bocconi University, Milan, Italy
{simone.degiorgi}@studbocconi.it

December 21, 2025

Abstract

Empathy is often proposed as a mechanism to promote cooperation and fairness in multi-agent reinforcement learning (MARL), yet its interaction with learning dynamics remains poorly understood. We study empathy-weighted reward shaping across canonical social dilemmas—Prisoner's Dilemma (PD), Stag Hunt (SH), and a Renewable Resource Sharing (RS) commons—and compare fixed empathy regimes with agents that learn empathy endogenously. We adopt a convention where $\alpha = 0$ denotes maximal empathy and $\alpha = 1$ maximal selfishness. Across PD and SH, high empathy simultaneously maximizes cooperation, social welfare, and equality. However, agents that learn empathy fail to converge to this optimal regime, instead stabilizing at intermediate values. In RS, empathy stabilizes shared resources and prevents collapse, revealing a genuine stability–efficiency trade-off. Our results show that empathy can be globally optimal, but standard learning dynamics are insufficient to reliably discover and sustain it.

1 Introduction

Multi-agent reinforcement learning (MARL) provides a framework for studying strategic interaction in social dilemmas, where individually rational behavior can lead to collectively suboptimal outcomes. Classical results in evolutionary game theory and behavioral economics identify cooperation-supporting mechanisms such as reciprocity, repeated interaction, and social preferences [1, 7].

Empathy—broadly defined as valuing others' outcomes—has been proposed as a key driver of cooperation and fairness. In MARL, empathy is commonly implemented via reward shaping or inequity-averse objectives [5]. Yet it remains unclear whether empathy improves or harms efficiency, whether it induces a fairness–efficiency trade-off, and whether empathy can be learned endogenously.

This paper addresses these questions through a comprehensive empirical study across multiple environments and population sizes ($N = 2, 4$), measuring cooperation, strict cooperation, total reward, inequality, stock dynamics, collapse frequency, and learned empathy.

Contributions.

- We show that high empathy can be globally optimal in social dilemmas, maximizing cooperation, welfare, and equality.
- We demonstrate that learning empathy fails to recover this optimum, converging instead to suboptimal intermediate regimes.
- We identify a genuine stability–efficiency trade-off in commons dilemmas.
- We show that limitations arise from learning dynamics rather than empathy itself.

2 Related Work

Social preferences and fairness. Models of inequity aversion formalize how agents trade off

*Equal contribution, the ordering is alphabetical.

personal and social outcomes [3, 2], motivating empathy-based objectives in MARL.

Cooperation in social dilemmas. The emergence of cooperation has been extensively studied in evolutionary systems [1, 7]. Sequential social dilemmas extend these ideas to learning agents [6].

Prosocial MARL. Prosocial reward shaping can improve outcomes in some cooperative settings [8, 5], but results are sensitive to scale and learning dynamics.

Learning instability. Deep MARL is known to suffer from sensitivity and local optima [4], motivating multi-seed evaluation.

3 Methodology

3.1 Empathy-Weighted Reward

Let $r_i(t)$ be the environment reward of agent i at time t, and N the number of agents:

$$\tilde{r}_i(t) = \alpha_i\, r_i(t) + (1 - \alpha_i)\frac{1}{N-1}\sum_{j \neq i} r_j(t), \qquad (1)$$

where $\alpha_i \in [0, 1]$ denotes **selfishness**.

Convention. Throughout the paper:

- $\alpha = 0$ corresponds to maximal empathy,
- $\alpha = 1$ corresponds to pure selfishness.

3.2 Fixed vs Learned Empathy

We compare fixed-α baselines with agents that learn $\alpha_i = \sigma(\theta_i)$ jointly with policy parameters. We report the population mean $\bar{\alpha}$.

3.3 Environments and Metrics

We evaluate PD, SH, and RS. Metrics include cooperation fraction, strict cooperation, total reward, inequality (reward variance), stock level and collapse frequency (RS), and learned empathy.

4 Results

4.1 Prisoner's Dilemma (N=2)

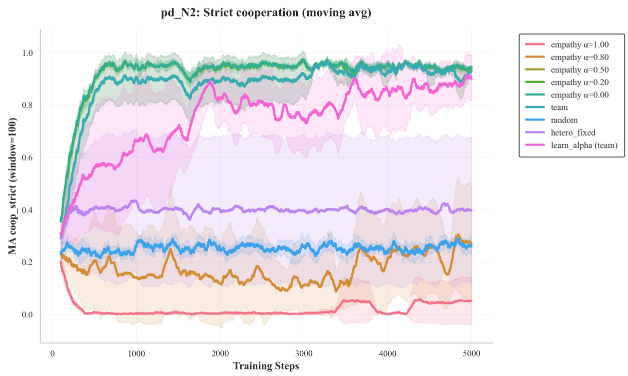

Figure 1: PD ($N = 2$): strict cooperation. Maximal empathy ($\alpha = 0$) yields near-perfect cooperation, while selfish agents ($\alpha = 1$) collapse to zero.

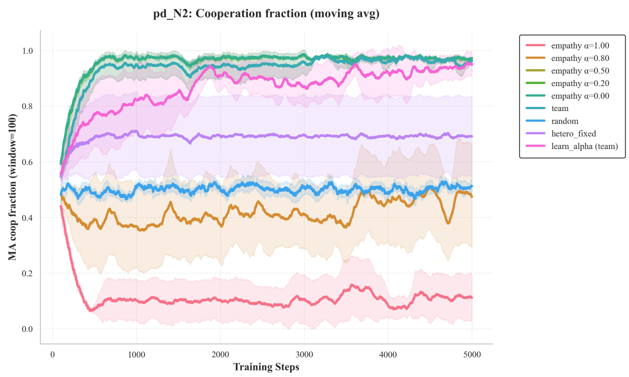

Figure 2: PD ($N = 2$): cooperation fraction. Cooperation decreases monotonically with increasing selfishness.

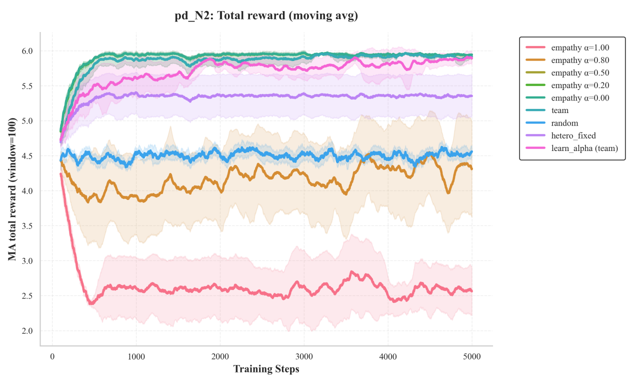

Figure 3: PD ($N = 2$): total reward. Empathy simultaneously maximizes social welfare and cooperation.

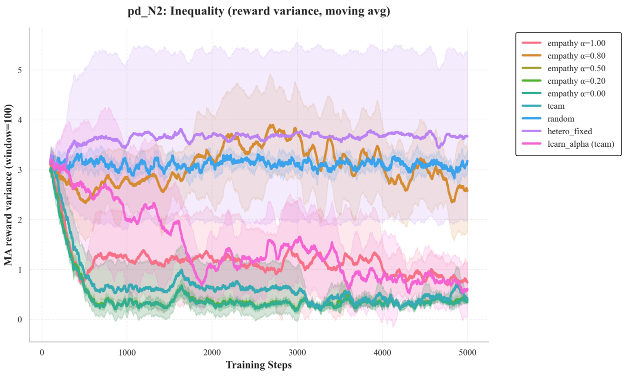

Figure 4: PD ($N = 2$): inequality. Increasing empathy sharply reduces reward variance.

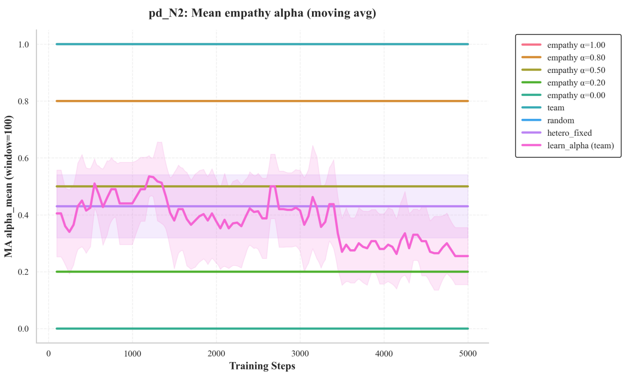

Figure 5: PD ($N = 2$): learned empathy. Despite $\alpha = 0$ being globally optimal, learning converges to intermediate values.

4.2 Prisoner's Dilemma (N=4)

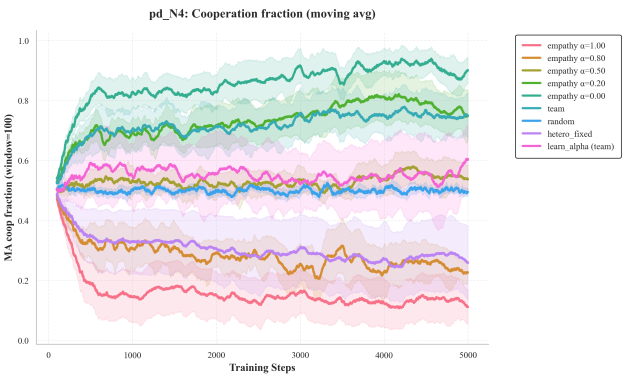

Figure 6: PD ($N = 4$): cooperation fraction. Empathy dominates across all levels of cooperation.

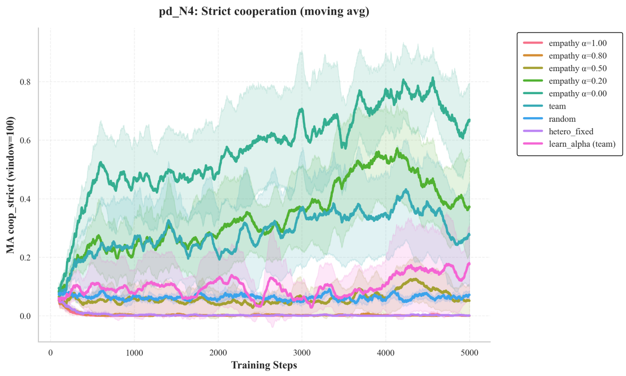

Figure 7: PD ($N = 4$): strict cooperation. Selfish agents fail to sustain cooperation at scale.

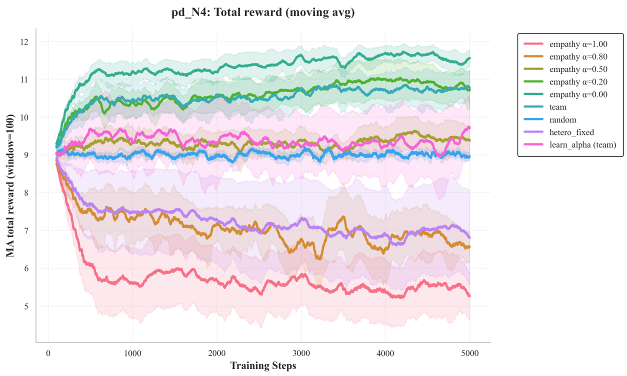

Figure 8: PD ($N = 4$): total reward. Empathy maximizes welfare, while selfishness yields the worst outcomes.

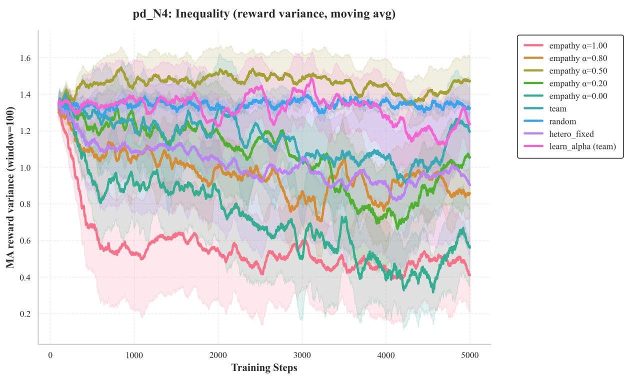

Figure 9: PD ($N = 4$): inequality. Reward variance increases monotonically with selfishness.

Figure 10: PD ($N = 4$): learned empathy. Learning stabilizes at suboptimal intermediate regimes.

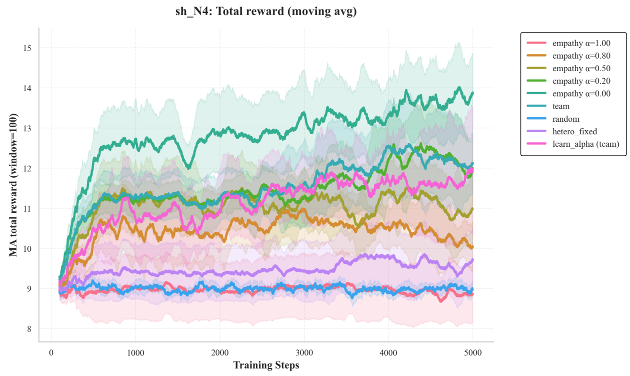

Figure 13: SH ($N = 4$): total reward. Empathy yields the highest collective payoff.

4.3 Stag Hunt (N=4)

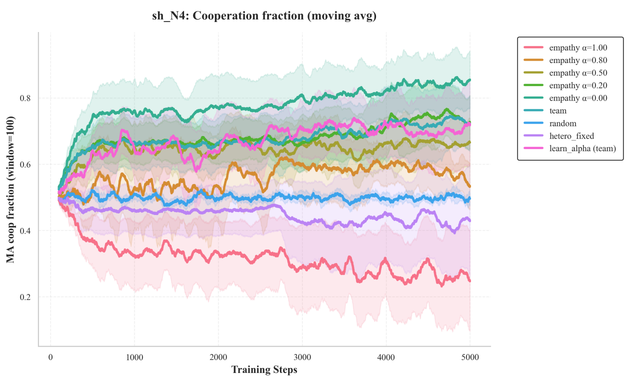

Figure 11: SH ($N = 4$): cooperation fraction. Empathy improves coordination on the payoff-dominant equilibrium.

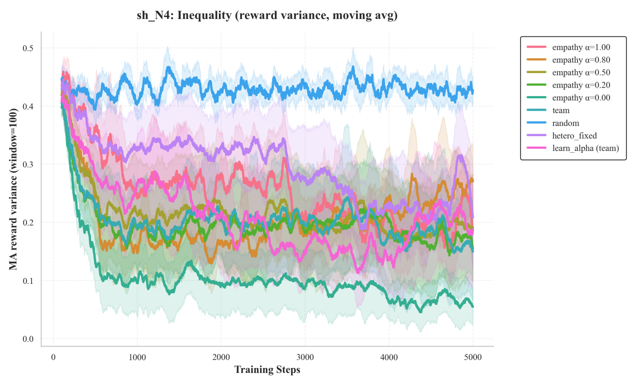

Figure 14: SH ($N = 4$): inequality. Empathy compresses payoff disparities without sacrificing welfare.

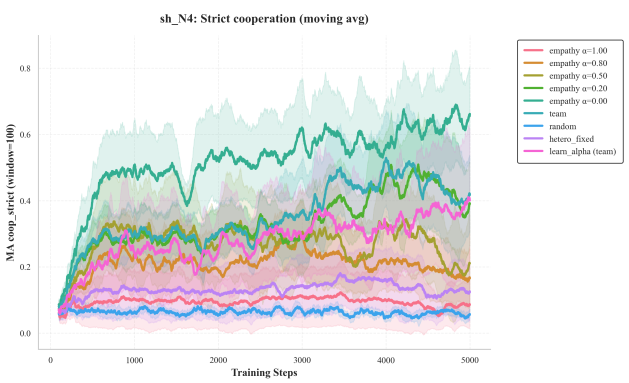

Figure 12: SH ($N = 4$): strict cooperation. High empathy promotes stable coordination.

4.4 Renewable Resource Sharing (N=4)

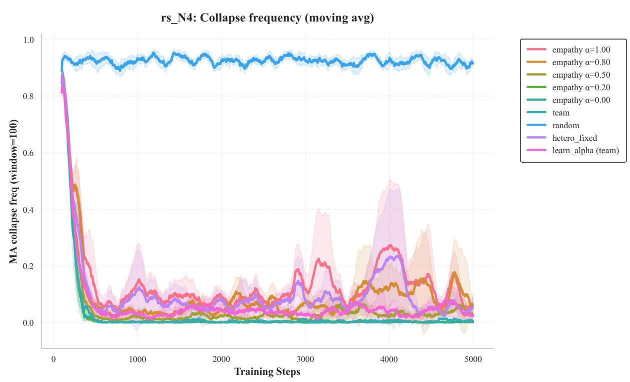

Figure 15: RS ($N = 4$): collapse frequency. Empathy nearly eliminates collapses, while selfishness induces systemic failure.

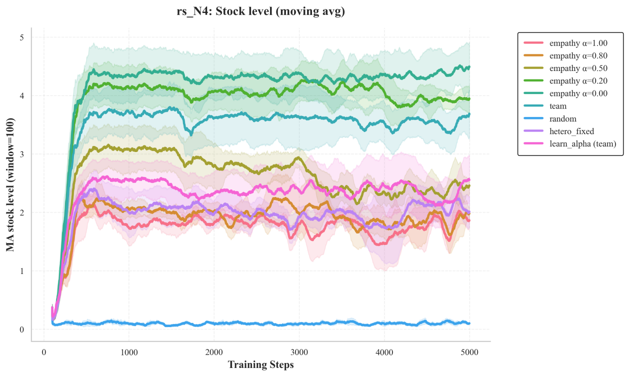

Figure 16: RS ($N = 4$): stock level. Empathy sustains higher resource levels over time.

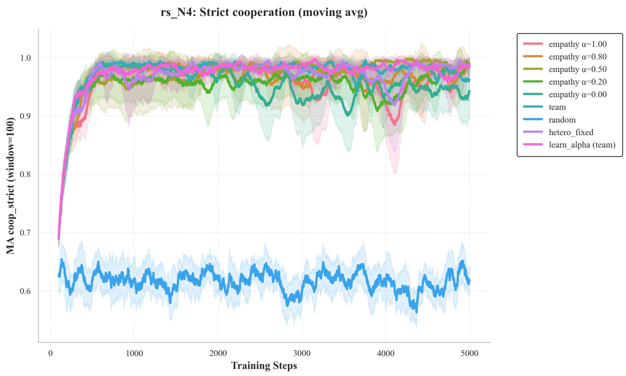

Figure 17: RS ($N = 4$): strict cooperation. Empathy promotes coordinated restraint in resource extraction.

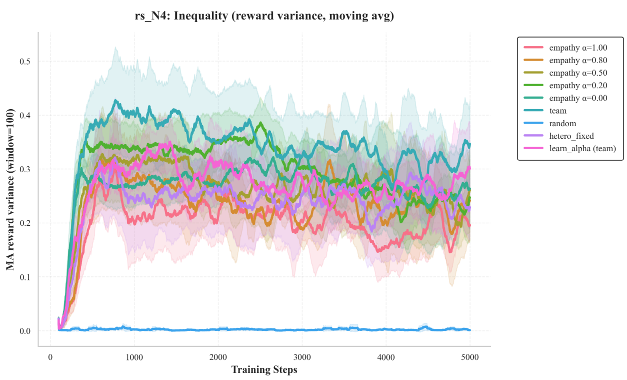

Figure 18: RS ($N = 4$): inequality. Empathy reduces payoff dispersion while stabilizing the commons.

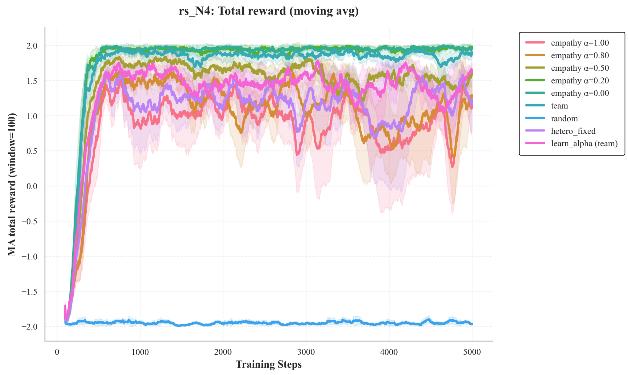

Figure 19: RS ($N = 4$): total reward. Intermediate selfishness can maximize reward, revealing a stability–efficiency trade-off.

5 Discussion

Across Prisoner's Dilemma and Stag Hunt, empathy emerges as globally optimal: it maximizes cooperation, strict cooperation, total reward, and equality. These results show that empathy does not merely trade efficiency for fairness—it can align individual incentives with collective optimality.

However, agents that learn empathy endogenously consistently fail to converge to this regime. Learned empathy stabilizes at intermediate values even when full empathy strictly dominates across all metrics, indicating a fundamental limitation of gradient-based learning dynamics.

In Renewable Resource Sharing, empathy stabilizes the system and prevents collapse, but conservative exploitation introduces a genuine stability–efficiency trade-off. Here, empathy remains essential for avoiding catastrophic outcomes, even when not maximizing short-term reward.

Overall, the primary bottleneck is not empathy as a social preference, but the inability of standard learning mechanisms to discover and sustain globally optimal empathic behavior.

6 Conclusion

We presented a comprehensive empirical study of empathy in multi-agent reinforcement learning. High empathy is shown to be globally optimal in social dilemmas and coordination games, maximizing cooperation, welfare, and equality. Nevertheless, agents that learn empathy fail to recover this optimal regime, converging instead to suboptimal compromises.

In commons dilemmas, empathy stabilizes shared resources and prevents collapse, exposing a stability–efficiency trade-off absent in simpler games. These findings suggest that future MARL systems must move beyond scalar, state-independent empathy parameters and develop learning mechanisms capable of discovering and sustaining globally optimal social preferences.

References

[1] Robert Axelrod and William D. Hamilton. The evolution of cooperation. *Science*, 211(4489):1390–1396, 1981.

[2] Gary E. Bolton and Axel Ockenfels. Erc: A theory of equity, reciprocity, and competition. *American Economic Review*, 90(1):166–193, 2000.

[3] Ernst Fehr and Klaus M. Schmidt. A theory of fairness, competition, and cooperation. *Quarterly Journal of Economics*, 114(3):817–868, 1999.

[4] Peter Henderson, Riashat Islam, Philip Bachman, Joelle Pineau, Doina Precup, and David Meger. Deep reinforcement learning that matters. *Proceedings of the AAAI Conference on Artificial Intelligence*, 32(1), 2018.

[5] Edward Hughes, Joel Z. Leibo, Murray Phillips, Karl Tuyls, and Edgar Duéñez-Guzmán. Inequity aversion improves cooperation in multi-agent reinforcement learning. *arXiv preprint arXiv:1805.09243*, 2018.

[6] Joel Z. Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. Multi-agent reinforcement learning in sequential social dilemmas. *arXiv preprint arXiv:1702.03037*, 2017.

[7] Martin A. Nowak. Five rules for the evolution of cooperation. *Science*, 314(5805):1560–1563, 2006.

[8] Alexander Peysakhovich and Adam Lerer. Prosocial learning agents solve coordination problems. *arXiv preprint arXiv:1703.04908*, 2018.