Reputation As a New Route to Cooperation in Multi-Agent Reinforcement Learning

Roman Chiva Gil
University of Amsterdam
Amsterdam, Netherlands
r.chivagil@uva.nl

ABSTRACT

While the field of cooperation in mixed-motive multi-agent reinforcement learning (MARL) has seen substantial growth in recent years, fundamental challenges remain. One such challenge is scalability to large and diverse populations of agents. My doctoral project investigates how indirect reciprocity (IR), an extensively studied mechanism in evolutionary game theory, can be adapted to populations of learning agents. IR enables agents to condition their cooperative behavior on reputations assigned by third-party observers according to social norms. In my thesis I start by studying the dynamics of Markov games augmented with reputation systems through minimalist MARL models, prioritizing interpretability while retaining representative emergent behaviors. Addressing previous pessimistic findings regarding the effectiveness of reputation systems in MARL, we demonstrated that the choices of learning algorithm and state representation play a crucial role in enabling cooperation. Through the systematic investigation of these design choices, we identified key principles governing when and how reputation mechanisms can effectively promote cooperation in populations of learning agents. Building on these insights we plan to expand our scope to more complex environments while considering the practical limitations of applying reputation systems to real-world problems. A central focus is the relaxation of common assumptions such as perfect observability, centralized reputation assignment and homogeneous populations with the goal of developing practical, robust mechanisms for fostering cooperation in realistic multi-agent systems.

KEYWORDS

Multi-Agent Reinforcement Learning; Social Dilemmas; Cooperation; Game Theory; Reputation Systems

ACM Reference Format:
Roman Chiva Gil. 2026. Reputation As a New Route to Cooperation in Multi-Agent Reinforcement Learning. 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/BJNB9076

1 INTRODUCTION

Human cooperation has been extensively studied from a game theoretic perspective across various disciplines. It is generally agreed upon that, although we do possess some intrinsic motivations to cooperate (e.g., empathy)[9], cooperation is primarily sustained by evolved social mechanisms and institutions that facilitate trust and mutual aid[2, 13]. However, the effectiveness of these institutions now being tested as we integrate adaptive autonomous agents into systems previously formed exclusively by humans. As these domains transition into hybrid or possibly fully autonomous environments, the social mechanisms that governed human behavior are being outpaced. Autonomous agents can operate at scales and speeds that defy traditional human oversight [12]. Thus, like human societies evolved specific social mechanisms and institutions to sustain cooperation, we must now develop their analogs to ensure robust cooperation in autonomous multi-agent systems[3, 6, 21]. This has been one of the central research areas in multi-agent RL over the last decade.

We propose that Indirect Reciprocity (IR) and reputation systems offer a promising novel framework for sustaining cooperation in large, decentralized populations of autonomous agents. In human societies, these mechanisms serve as promoters of cooperation and trust and are prevalent in forms ranging from informal gossip networks where word spreads about who is trustworthy and who is a cheater, to the sophisticated rating systems used by e-commerce platforms like eBay[19, 22]. Our motivation is that by incorporating a reputation system into environments with interacting autonomous agents, we can create a self-regulating system. In such a system, agents are incentivized to cooperate not through direct oversight or intrinsic motivations, but because their reputation becomes a valuable asset.

2 BACKGROUND

Indirect Reciprocity: IR enables cooperation through reputation-based discrimination, where agents cooperate based on their partner's reputation rather than direct interaction history[18]. Social norms govern the system by mapping an agent's action and their recipient's reputation to a new reputation assignment for the actor, typically encoded as 4-bit binary strings covering the four cases: defecting/cooperating against bad/good reputation recipients ($D \to B$, $D \to G$, $C \to B$, $C \to G$). This creates incentives where maintaining a good reputation yields long-term benefits exceeding immediate cooperation costs. For example, the Stern Judging norm (1001)[20] assigns good reputations when agents cooperate with good-reputation partners or defect against bad-reputation partners, and bad reputations otherwise, effectively distinguishing justified punishment from unjustified defection and thereby sustaining cooperation.

Multi-Agent Reinforcement Learning: Sustaining cooperation in mixed-motive environments is a difficult challenge in MARL.

This work is licensed under a Creative Commons Attribution International 4.0 License.

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/BJNB9076

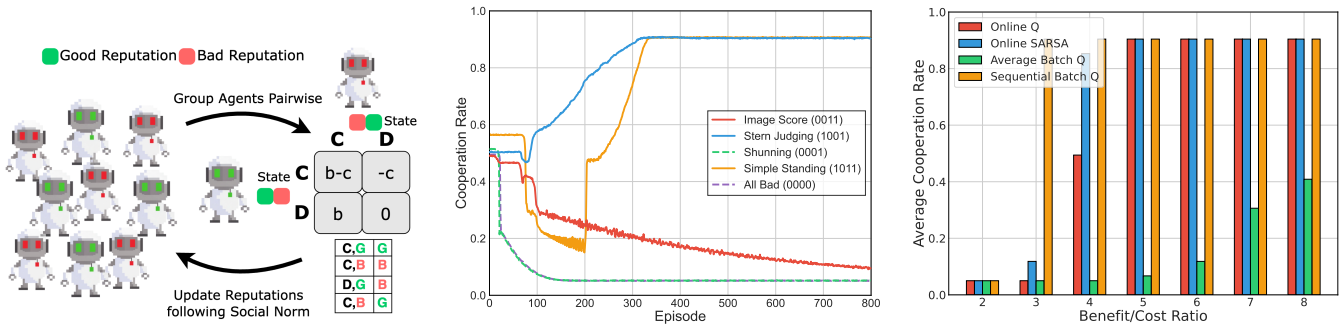

Figure 1: (a) Schematic of the indirect reciprocity environment (b) Learning curves under different social norms for 500 Q-learners (c) Cooperation rates across benefit-to-cost ratios for different algorithmic variants under Stern Judging

The field has seen significant progress over the last decade [8], with methods such as Opponent Shaping [5, 10, 15] or Intrinsic Reward design [16, 17, 25] demonstrating great promise in fostering cooperation. However, many open questions remain [7]. Of these, we point out how reputation systems could provide an answer to scalability to large populations without assuming that agents are inherently rewarded for being prosocial. Instead of requiring control over an agent's reward, the system broadcasts reputation as external, public information and treats cooperation as a strategic choice. Agents can choose to integrate this information into their decision making if it provides a competitive advantage. By researching how to most effectively broadcast information, this approach has the potential of creating a self-regulating environment where cooperation becomes the most rational strategy for self-interested agents.

3 PRELIMINARY RESULTS

The first objective of our project was to develop a detailed understanding of cooperation and reputation dynamics under RL adaptation and what factors determine the effectiveness of reputations in promoting cooperation under RL. In doing so, we demonstrated that reputation systems can be substantially more effective in promoting cooperation than previously reported [1, 24]. We developed a minimalist population game environment that abstracts away environmental complexity representing a typical indirect reciprocity setting[23]: A large pool of agents engages in one-shot interactions with random members of the population. Each agent is assigned a reputation, and at each timestep, agents are paired to play a one-shot prisoner's dilemma game. Agents observe their own reputation and their opponent's reputation to condition their actions. A social norm centrally assigns reputations by judging agents based on their actions and their opponent's reputation. A schematic of the environment is provided in Figure 1a.

Figure 1b shows learning curves illustrating how different social norms impact the emergence of cooperation in a population of 500 Q-learners using an ϵ-greedy policy. Consistent with EGT models[23], Stern Judging and Simple Standing enable agents to converge to cooperative equilibria where they adopt discriminatory strategies, cooperating selectively based on reputation. Under other norms, populations consistently collapse to mutual defection. Both successful norms exhibit tipping-point behavior: initially, discriminatory and defection strategies compete, but once a critical mass of agents adopts discriminatory strategies, the population rapidly shifts toward cooperation. The threshold for this tipping point is norm-dependent, with Stern Judging requiring fewer discriminators to trigger the transition.

We find that in addition to the social norm, the likelihood of converging to the cooperative equilibrium largely determined by how the agents' learning dynamics navigate the competing incentives of immediate defection payoffs and the long-term benefits of maintaining a good reputation before defectors dominate. By comparing several algorithmic variants, we identified specific features that more effectively promote reaching this tipping point. Figure 1c shows a selection of tested variants under the Stern Judging norms across benefit-to-cost ratios. We found that stochastic parameter updates and frequency-adjusted updates (where learning rates effectively depend on state-action visitation frequencies, as in Sequential Batch Q) better facilitate coordination. These findings align with prior work on cooperation emergence among RL agents in two-player iterated prisoner's dilemma [4, 14].

4 FUTURE STEPS

While our initial work used stylized models to align with EGT, we plan to relax these assumptions to bridge the gap with realistic systems. Specifically, we are addressing the challenge of environmental complexity and temporally extended interactions, which is effectively captured by Sequential Social Dilemma (SSD) benchmarks like CoinGame or Cleanup [11]. In these environments, cooperation vs defection is no longer a binary choice but a time-extended process. This requires the development of sophisticated social norms capable of evaluating sequences of actions within their environmental context. Furthermore, since SSDs require agents to master environmental competence alongside strategic interaction, we must investigate how to best integrate reputation data into agent observations to provide a clear and effective learning signal. Finally, we intend to relax the assumption of a centralized reputation system and decentralizing the evaluation process. By introducing a communication layer where agents exchange messages about peer behavior, we explicitly incorporate the dynamics of gossip into the model. This shift allows us to investigate how cooperation can survive when reputation information is shared through noisy, subjective, or even potentially dishonest channels.

ACKNOWLEDGMENTS

This PhD project is funded through project "Reputation as a new route to cooperation in multi-agent reinforcement learning" with file number OCENW.M.22.322 of the research programme Open Competition Domain Science which is (partly) financed by the Dutch Research Council (NWO).

REFERENCES

[1] Nicolas Anastassacos, Julian García, Stephen Hailes, and Mirco Musolesi. 2021. Cooperation and Reputation Dynamics with Reinforcement Learning. https://doi.org/10.48550/arXiv.2102.07523 arXiv:2102.07523 [cs].

[2] Robert Axelrod and William D. Hamilton. 1981. The Evolution of Cooperation. *Science* 211, 4489 (March 1981), 1390–1396. https://doi.org/10.1126/science.7466396

[3] Wolfram Barfuss, Jessica Flack, Chaitanya S. Gokhale, Lewis Hammond, Christian Hilbe, Edward Hughes, Joel Z. Leibo, Tom Lenaerts, Naomi Leonard, Simon Levin, Udari Madhushani Sehwag, Alex McAvoy, Janusz M. Meylahn, and Fernando P. Santos. 2025. Collective cooperative intelligence. *Proceedings of the National Academy of Sciences* 122, 25 (June 2025), e2319948121. https://doi.org/10.1073/pnas.2319948121

[4] Wolfram Barfuss and Janusz M. Meylahn. 2023. Intrinsic fluctuations of reinforcement learning promote cooperation. *Scientific Reports* 13, 1 (Jan. 2023), 1309. https://doi.org/10.1038/s41598-023-27672-7

[5] Tim Cooijmans, Milad Aghajohari, and Aaron Courville. 2023. Meta-Value Learning: a General Framework for Learning with Learning Awareness. https://doi.org/10.48550/arXiv.2307.08863 arXiv:2307.08863 [cs].

[6] Allan Dafoe, Edward Hughes, Yoram Bachrach, Tantum Collins, Kevin R. McKee, Joel Z. Leibo, Kate Larson, and Thore Graepel. 2020. Open Problems in Cooperative AI. https://doi.org/10.48550/arXiv.2012.08630 arXiv:2012.08630 [cs].

[7] Yali Du, Joel Z. Leibo, Usman Islam, Richard Willis, and Peter Sunehag. 2023. A Review of Cooperation in Multi-agent Learning. http://arxiv.org/abs/2312.05162 arXiv:2312.05162 [cs].

[8] Shaheen Fatima, Nicholas R. Jennings, and Michael Wooldridge. 2024. Learning to Resolve Social Dilemmas: A Survey. *Journal of Artificial Intelligence Research* 79 (March 2024), 895–969. https://doi.org/10.1613/jair.1.15167

[9] Ernst Fehr and Urs Fischbacher. 2003. The nature of human altruism. *Nature* 425, 6960 (Oct. 2003), 785–791. https://doi.org/10.1038/nature02043

[10] Jakob N. Foerster, Richard Y. Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. 2018. Learning with Opponent-Learning Awareness. https://doi.org/10.48550/arXiv.1709.04326 arXiv:1709.04326 [cs].

[11] Zihao Guo, Shuqing Shi, Richard Willis, Tristan Tomilin, Joel Z. Leibo, and Yali Du. 2025. SocialJax: An Evaluation Suite for Multi-agent Reinforcement Learning in Sequential Social Dilemmas. https://doi.org/10.48550/arXiv.2503.14576 arXiv:2503.14576 [cs].

[12] Lewis Hammond, Alan Chan, Jesse Clifton, Jason Hoelscher-Obermaier, Akbir Khan, Euan McLean, Chandler Smith, Wolfram Barfuss, Jakob Foerster, Tomáš Gavenčiak, The Anh Han, Edward Hughes, Vojtěch Kovařík, Jan Kulveit, Joel Z. Leibo, Caspar Oesterheld, Christian Schroeder de Witt, Nisarg Shah, Michael Wellman, Paolo Bova, Theodor Cimpeanu, Carson Ezell, Quentin Feuillade-Montixi, Matija Franklin, Esben Kran, Igor Krawczuk, Max Lamparth, Niklas Lauffer, Alexander Meinke, Sumeet Motwani, Anka Reuel, Vincent Conitzer, Michael Dennis, Iason Gabriel, Adam Gleave, Gillian Hadfield, Nika Haghtalab, Atoosa Kasirzadeh, Sébastien Krier, Kate Larson, Joel Lehman, David C. Parkes, Georgios Piliouras, and Iyad Rahwan. 2025. Multi-Agent Risks from Advanced AI. https://doi.org/10.48550/arXiv.2502.14143 arXiv:2502.14143 [cs].

[13] Garrett Hardin. 2000. The tragedy of the commons. In *Environmental Ethics*. Routledge. Num Pages: 12.

[14] Michael Kaisers and Karl Tuyls. 2010. Frequency adjusted multi-agent Q-learning. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1 - Volume 1 (AAMAS '10)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 309–316. https://dl.acm.org/doi/10.5555/1838206.1838250

[15] Alistair Letcher, Jakob Foerster, David Balduzzi, Tim Rocktäschel, and Shimon Whiteson. 2021. Stable Opponent Shaping in Differentiable Games. https://doi.org/10.48550/arXiv.1811.08469 arXiv:1811.08469 [cs].

[16] Kevin R. McKee, Ian Gemp, Brian McWilliams, Edgar A. Duéñez-Guzmán, Edward Hughes, and Joel Z. Leibo. 2020. Social diversity and social preferences in mixed-motive reinforcement learning. https://doi.org/10.48550/arXiv.2002.02325 arXiv:2002.02325 [cs].

[17] Kevin R. McKee, Edward Hughes, Tina O. Zhu, Martin J. Chadwick, Raphael Koster, Antonio Garcia Castaneda, Charlie Beattie, Thore Graepel, Matt Botvinick, and Joel Z. Leibo. 2023. A multi-agent reinforcement learning model of reputation and cooperation in human groups. https://doi.org/10.48550/arXiv.2103.04982 arXiv:2103.04982 [cs].

[18] Martin A Nowak and Karl Sigmund. 1998. The Dynamics of Indirect Reciprocity. *Journal of Theoretical Biology* 194, 4 (Oct. 1998), 561–574. https://doi.org/10.1006/jtbi.1998.0775

[19] Martin A. Nowak and Karl Sigmund. 2005. Evolution of indirect reciprocity. *Nature* 437, 7063 (Oct. 2005), 1291–1298. https://doi.org/10.1038/nature04131

[20] Jorge M. Pacheco, Francisco C. Santos, and Fabio A. C. C. Chalub. 2006. Stern-Judging: A Simple, Successful Norm Which Promotes Cooperation under Indirect Reciprocity. *PLOS Computational Biology* 2, 12 (Dec. 2006), e178. https://doi.org/10.1371/journal.pcbi.0020178

[21] Fernando P. Santos. 2024. Prosocial dynamics in multiagent systems. *AI Magazine* 45, 1 (2024), 131–138. https://doi.org/10.1002/aaai.12143 _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/aaai.12143.

[22] Fernando P. Santos, Jorge M. Pacheco, and Francisco C. Santos. 2021. The complexity of human cooperation under indirect reciprocity. *Philosophical Transactions of the Royal Society B: Biological Sciences* 376, 1838 (Nov. 2021), 20200291. https://doi.org/10.1098/rstb.2020.0291

[23] Fernando P. Santos, Francisco C. Santos, and Jorge M. Pacheco. 2016. Social Norms of Cooperation in Small-Scale Societies. *PLOS Computational Biology* 12, 1 (Jan. 2016), e1004709. https://doi.org/10.1371/journal.pcbi.1004709

[24] Martin Smit and Fernando P. Santos. 2024. Learning Fair Cooperation in Mixed-Motive Games with Indirect Reciprocity. In *Proceedings of the Thirty-ThirdInternational Joint Conference on Artificial Intelligence*. 220–228. https://doi.org/10.24963/ijcai.2024/25 arXiv:2408.04549 [cs].

[25] Jane X. Wang, Edward Hughes, Chrisantha Fernando, Wojciech M. Czarnecki, Edgar A. Duenez-Guzman, and Joel Z. Leibo. 2019. Evolving intrinsic motivations for altruistic behavior. https://doi.org/10.48550/arXiv.1811.05931 arXiv:1811.05931 [cs].