# Fairness and Cooperation between Independent Reinforcement Learners through Indirect Reciprocity

Extended Abstract

Jacobus Smit
University of Amsterdam
Amsterdam, The Netherlands
jacobus.smit@uva.nl

Fernando P. Santos
University of Amsterdam
Amsterdam, The Netherlands
f.p.santos@uva.nl

## ABSTRACT

In a multi-agent setting, altruistic cooperation is costly yet socially desirable. As such, agents adapting through independent reinforcement learning struggle to converge to efficient, cooperative policies. Indirect reciprocity (IR) constitutes a possible mechanism to encourage cooperation by introducing reputations, social norms and the possibility that agents reciprocate based on past actions. IR has been mainly studied in homogeneous populations. In this paper, we introduce a model that allows for both reputation and group-based cooperation, and analyse how specific social norms (i.e. rules to assign reputations) can lead to varying levels of cooperation and fairness. We investigate how a finite population of independent Q-learning agents perform under different social norms. We observe that while norms such as *Stern-Judging* sustain both cooperation and fairness in populations of learning agents, other norms used to judge in- or out-group interactions can lead to unfair outcomes.

## KEYWORDS

Indirect reciprocity; Prosocial dynamics; Cooperation; Fairness; Multi-agent Reinforcement learning; Reputation systems

**ACM Reference Format:**
Jacobus Smit and Fernando P. Santos. 2024. Fairness and Cooperation between Independent Reinforcement Learners through Indirect Reciprocity: Extended Abstract. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), Auckland, New Zealand, May 6 – 10, 2024*, IFAAMAS, 3 pages.

## 1 INTRODUCTION

Cooperation is a fundamental research topic across disciplines [6, 17]. While cooperative populations tend to thrive, individuals are tempted to act selfishly, receiving the benefits of others' cooperation without exerting the effort themselves. The conundrum underlying this interaction is evident if we formally translate it into the so-called *donation game*, whereby a donor decides whether to pay a cost $c$ to offer a benefit $b$ to a recipient. Assuming that $b > c > 0$, this simple interaction illustrates the ubiquitous social dilemma of altruistic cooperation. Understanding how to engineer cooperation in these mixed-motive settings is a central challenge in multi-agent systems research [3, 4, 15].



*Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2024), N. Alechina, V. Dignum, M. Dastani, J.S. Sichman (eds.), May 6 – 10, 2024, Auckland, New Zealand.* © 2024 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

Indirect reciprocity (IR) is one mechanism used to sustain cooperation [11, 13, 18]. With IR, agents are assumed to discriminate and provide benefits based on previous interactions and reputations of others. The rules that determine how such reputations are updated (so-called *social norms*) encapsulate the moral judgements of what constitutes a good or a bad action [12].

Although being a promising mechanism, applying IR in systems of learning agents comes with some challenges. First, only a small set of social norms (i.e., rules followed to assign reputations) are able to stabilise cooperation in populations of homogeneous agents [1, 12]. Secondly, it is not uncommon for distinct groups to exist or emerge in a population, particularly when cooperation is conditional on reputations [7, 19]. When social norms display in-group bias or out-group prejudice, we may observe parochial cooperation leading to inequality, as is widespread in many human systems [5]. It remains unclear how discrimination based on reputations might be affected by discrimination based on group identities. Here we address the following question: *How to apply indirect reciprocity to group-structured populations of learning agents to achieve high levels of both cooperation and fairness?*

We model a population of independent agents learning through Q-learning. We find that fairness and cooperation are possible to achieve if the right norm is chosen to judge actions.

## 2 MODEL

We consider well-mixed population of agents stratified into a minority and majority group, with size $Z_m$ and $Z_M$ respectively. Agents interact by playing the *donation game* introduced previously.

To encourage cooperation we consider reputations and social norms. Social norms are rules that determine agents' reputations after each game takes place. In our model, these rules may depend on 1) the action taken by the donor, 2) the current reputation of the recipient, and 3) whether the two agents are in the same group. Following prior works on indirect reciprocity, all of these inputs are binary: a donor has two actions, reputations are either "good" or "bad" [11], and agents can be in the same or a different group.

An example of a social norm is `SternJudging`, which deems that it is good to either defect against a bad agent or cooperate with a good one, and that doing the opposite action in either case is bad [14]. Another is `SimpleStanding`, stating that the only bad thing to do is to defect against a good agent [16]. `ImageScore` assumes that Cooperation is good, and defection is bad [10]. `Shunning` postulates that only cooperation with good opponents is good, and all else is bad. Neither of these norms take into account the group identity of agents. A norm that judges in-group interactions with `SternJudging` and out-group interactions with `SimpleStanding`

would imply a greater degree of strictness when judging interactions where both agents are members of the same group.

We assume that an agent who intends to cooperate will sometimes defect with some probability $\epsilon$, that third-party observers using social norms to assign reputations can also err and assign the opposite reputation than intended with probability $\delta$ and that agents err likewise when retrieving others' reputations with rate $\chi$.

Agents decide to cooperate or defect according to a strategy/policy (learned as detailed in Section 2.1). Strategies define, for each combination of reputation and in/out-group, a corresponding action. As such, the space of strategies consists of functions $\sigma : \{0,1\}^2 \rightarrow \{0,1\}$. Notable strategies include `AllD`, which unconditionally refuses to donate, `Disc`, which conditionally cooperates with good agents and defects against bad ones, and `AllC`, which unconditionally donates. So called "unfair" strategies may, for example, play `AllC` with in-group members and play `Disc` with everyone else.

## 2.1 Reinforcement learning model

We model agents as independent tabular Q-learners. Formally, if agent $i$ with meets agent $j$ who has group relation $x$ to agent $i$ and reputation $y$, then the action taken by $i$ ($a^*$) is determined by $a^* = \arg\max_{a \in \{0,1\}} Q_i[x, y, a]$, where $Q_i \in \mathbb{R}^3$ is the Q-table of agent $i$ and actions 0 and 1 correspond to defection and cooperation respectively. In doing so, agent $j$ will receive payoff $a^*b$ and agent $i$ will pay cost $a^*c$. Agent $j$ will attribute this payoff to the last action they took ($\hat{a}$) in context $\hat{x}$ and $\hat{y}$

$$Q_i^{t+1}[x, y, a^*] \leftarrow (1-\alpha) Q_i^t[x, y, a^*] - \alpha a^* c \quad (1)$$

$$Q_j^{t+1}[\hat{x}, \hat{y}, \hat{a}] \leftarrow (1-\alpha) Q_j^t[\hat{x}, \hat{y}, \hat{a}] + \alpha \hat{a} b \quad (2)$$

We start our simulations from randomly initialized $Q_i$ values and let agents adapt over time. Each simulation runs for 250,000 interactions and is repeated 50 times with different seeds.

# 3 RESULTS

We fix each possible social norm in a population of independent reinforcement learning agents and we measure the cooperativeness and fairness levels obtained in the long run, after agents adapt their policies. The *cooperativeness* of a system is the probability that in a uniformly sampled interaction the donor will cooperate. The *fairness* of a system is the ratio between the average payoffs of the worse off and better off group. This is akin to the *demographic parity* ratio from fairness in supervised learning applications [9]. As shown in Figure 1, under a small set of norms agents can learn to adopt policies such that high levels of cooperativeness and fairness emerge, even in a population composed by a minority and majority group. Stern-judging, a norm previously identified as leading to high levels of cooperation [1, 14], can also here support high levels of cooperation and fairness. Using Stern-judging to assess in-group interactions and the more simple Image-score to assess out-group interactions (i.e., Cooperation is Good, all else is Bad [10]) can still steer agents to learn policies that are both cooperation and fair; often, however, populations can end up in states where unfair cooperation is learned and only the majority groups receives the benefits of cooperation.

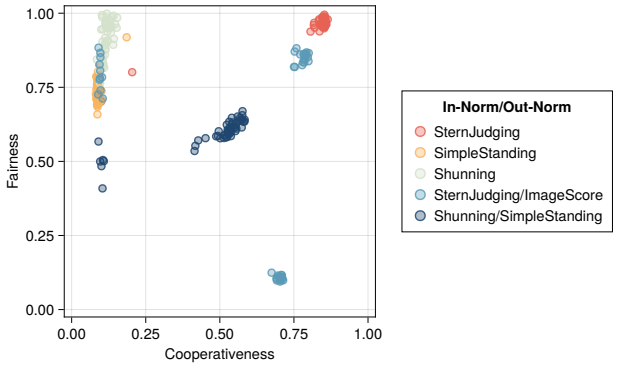


**Figure 1: With the proposed model we can classify each norm regarding the long-term average cooperativeness and fairness sustained in a population of agents independently adapting through Q-Learning. Here we show representative norms that, when applied to assess in- and out-group interactions, can lead to different levels of cooperativeness and fairness. By manipulating a social norm determining how reputations are assigned, we can affect not only the overall levels of cooperation but also the level of fairness in a population (i.e., the balance in average payoffs obtained by two groups). Parameters: $b = 10$, $c = 1$, $Z_M = 45$, $Z_m = 5$, $\alpha = 0.1$, $\epsilon = \delta = \chi = 0.01$.**

# 4 CONCLUSION

In this paper we show that indirect reciprocity allows for fair cooperation amongst group-structured agents. For this to happen, one has to judiciously select social norms. Norms play a large part in determining the stability and learnability of policies leading to a fair and cooperative outcome. Stricter norms (like `SternJudging`) allow for high levels of cooperativeness and fairness when agents adapt through reinforcement learning.

By using the minimal and generic donation game, we have demonstrated a proof of concept whose results may inform the application of indirect reciprocity to more elaborate multi-agent systems. That said, our simple framework motivates future work: First, it has previously been shown that ad hoc group formation may occur in spatial mixed-motive models [2] or complex networks [7, 19]; here we assume static group identities, however investigating learning dynamics with a changing group labels can, in the future, inform how to sustain fair cooperation in scenarios where group membership is dynamic. Second, our simple setup has the advantage of allowing to perform an exhaustive analysis of each norm. In more complex, sequential games where interactions are indirect and outcomes extend in time [8], it is difficult to immediately assign a cost and benefit to actions, making it unclear which actions deserve punishment. Future work can use the base model presented here and adapt it to analyse cooperation and fairness in more complex sequential environments.


# ACKNOWLEDGMENTS

Work partly funded by the Civic AI Lab (ICAI) and the Dutch Research Council (NWO) with project HUMAINER AI (KIVI.2019.006).

# REFERENCES


[1] Nicolas Anastassacos, Julian García, Stephen Hailes, and Mirco Musolesi. 2021. Cooperation and Reputation Dynamics with Reinforcement Learning. In *Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems, (AAMAS'21)*. International Foundation for Autonomous Agents and Multiagent Systems, 115–123. https://arxiv.org/abs/2102.07523

[2] Jacques Bara, Fernando P. Santos, and Paolo Turrini. 2023. The Role of Space, Density and Migration in Social Dilemmas. In *Proceedings of the 22nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS '23)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 625–633.

[3] Vincent Conitzer and Caspar Oesterheld. 2023. Foundations of Cooperative AI. *Proceedings of the AAAI Conference on Artificial Intelligence* 37, 13 (June 2023), 15359–15367. https://doi.org/10.1609/aaai.v37i13.26791

[4] Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. 2021. Cooperative AI: machines must learn to find common ground. *Nature* 593, 7857 (May 2021), 33–36. https://doi.org/10.1038/d41586-021-01170-0

[5] Jim A. C. Everett, Nadira S. Faber, and Molly Crockett. 2015. Preferences and beliefs in ingroup favoritism. *Frontiers in Behavioral Neuroscience* 9 (Feb. 2015), 15. https://doi.org/10.3389/fnbeh.2015.00015

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

[7] Jörg Gross and Carsten K. W. De Dreu. 2019. The rise and fall of cooperation through reputation and group polarization. *Nature Communications* 10, 1 (Feb. 2019), 776. https://doi.org/10.1038/s41467-019-08727-8

[8] Joel Z. Leibo, Vinicius Zambaldi, Marc Lanctot, Janusz Marecki, and Thore Graepel. 2017. Multi-agent Reinforcement Learning in Sequential Social Dilemmas. In *Proceedings of the 16th Conference on Autonomous Agents and MultiAgent Systems (AAMAS '17)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 464–473.

[9] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. 2021. A Survey on Bias and Fairness in Machine Learning. *Comput. Surveys* 54 (July 2021), 1–35. https://doi.org/10.1145/3457607

[10] Martin A. Nowak and Karl Sigmund. 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393, 6685 (June 1998), 573–577. https://doi.org/10.1038/31225 Number: 6685 Publisher: Nature Publishing Group.

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

[12] Hisashi Ohtsuki and Yoh Iwasa. 2004. How should we define goodness?—reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology* 231, 1 (Nov. 2004), 107–120. https://doi.org/10.1016/j.jtbi.2004.06.005

[13] Isamu Okada. 2020. A Review of Theoretical Studies on Indirect Reciprocity. *Games* 11, 3 (Sept. 2020), 27. https://doi.org/10.3390/g11030027 Number: 3 Publisher: Multidisciplinary Digital Publishing Institute.

[14] 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

[15] Ana Paiva, Fernando P. Santos, and Francisco C. Santos. 2018. Engineering pro-sociality with autonomous agents. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence and Thirtieth Innovative Applications of Artificial Intelligence Conference and Eighth AAAI Symposium on Educational Advances in Artificial Intelligence (AAAI'18/IAAI'18/EAAI'18)*. AAAI Press, New Orleans, Louisiana, USA, 7994–7999.

[16] Karthik Panchanathan and Robert Boyd. 2003. A Tale of Two Defectors: The Importance of Standing for Evolution of Indirect Reciprocity. *Journal of Theoretical Biology* 224 (Oct. 2003), 115–26. https://doi.org/10.1016/S0022-5193(03)00154-1

[17] David G. Rand and Martin A. Nowak. 2013. Human cooperation. *Trends in Cognitive Sciences* 17, 8 (Aug. 2013), 413–425. https://doi.org/10.1016/j.tics.2013.06.003

[18] 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 Publisher: Royal Society.

[19] Fernando P. Santos, Francisco C. Santos, Jorge M. Pacheco, and Simon A. Levin. 2021. Social Network Interventions to Prevent Reciprocity-driven Polarization. In *Proceedings of the 20th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS '21)*. International Foundation for Autonomous Agents and Multiagent Systems, Richland, SC, 1643–1645.