Learning Fair Cooperation in Mixed-Motive Games with Indirect Reciprocity

Martin Smit, **Fernando P. Santos**
Informatics Institute, University of Amsterdam
{j.m.m.smit, f.p.santos}@uva.nl

Abstract

Altruistic cooperation is costly yet socially desirable. As a result, agents struggle to learn cooperative policies through independent reinforcement learning (RL). Indirect reciprocity, where agents consider their interaction partner's reputation, has been shown to stabilise cooperation in homogeneous, idealised populations. However, more realistic settings are comprised of heterogeneous agents with different characteristics and group-based social identities. We study cooperation when agents are stratified into two such groups, and allow reputation updates and actions to depend on group information. We consider two modelling approaches: evolutionary game theory, where we comprehensively search for social norms (i.e., rules to assign reputations) leading to cooperation and fairness; and RL, where we consider how the stochastic dynamics of policy learning affects the analytically identified equilibria. We observe that a defecting majority leads the minority group to defect, but not the inverse. Moreover, changing the norms that judge in- and out-group interactions can steer a system towards either fair or unfair cooperation. This is made clearer when moving beyond equilibrium analysis to independent RL agents, where convergence to fair cooperation occurs with a narrower set of norms. Our results highlight that, in heterogeneous populations with reputations, carefully defining interaction norms is fundamental to tackle both dilemmas of cooperation and of fairness.

1 Introduction

Cooperation is a fundamental research topic across disciplines [Rand and Nowak, 2013; Fehr and Fischbacher, 2003]. While cooperative populations tend to thrive, individuals are tempted to act selfishly, receiving the benefits from the cooperation of others without exerting the effort themselves. The conundrum underlying this interaction is evident if we formally translate it into the so-called *donation game*, where a donor decides whether to pay a cost c to offer a benefit b to a recipient. Assuming $b > c > 0$, this simple interaction illustrates the ubiquitous social dilemma of altruistic cooperation. These dilemmas are known as mixed-motive games as they combine the principles of competitive (i.e., zero-sum games) and cooperative interactions. Understanding how to engineer cooperation in mixed-motive settings is a fundamental scientific challenge [Pennisi, 2005; Rand and Nowak, 2013] and a key frontier in AI research [Paiva *et al.*, 2018; Dafoe *et al.*, 2021; Conitzer and Oesterheld, 2023; Fatima *et al.*, 2024].

In distributed multi-agent systems, research has focused on the design of autonomous systems where cooperation is stable [Genesereth *et al.*, 1986]. In such contexts, it is fundamental to understand how adaptive agents can learn to cooperate over time in a decentralised way [Claus and Boutilier, 1998]. The cooperation mechanisms observed in human societies [Rand and Nowak, 2013] have inspired formal methods to stabilise cooperation in groups of artificial agents.

One particularly effective mechanism to sustain cooperation among humans is indirect reciprocity (IR) [Nowak and Sigmund, 2005; Santos *et al.*, 2021; Okada, 2020]. Within such a framework, agents are assumed to discriminate and provide benefits based on the social standing of others; this mechanism relies on the availability of reputations. 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 [Ohtsuki and Iwasa, 2004].

Despite success in human populations, the application of IR in systems of learning agents is technically challenging. As social norms assess the "goodness" of every action in every possible context, the number of possible norms grows combinatorially in the number of actions and states [Santos *et al.*, 2018]. Moreover, small differences in otherwise similar norms can have unpredictable effects on reputation dynamics. This makes predicting which norms will lead to a cooperative system a difficult task. A central challenge is therefore determining how reputations should be assigned for cooperation to be maximised. Previous work has shown that only a small set of social norms are able to stabilise cooperation in populations of homogeneous agents [Ohtsuki and Iwasa, 2004].

Additionally, it is common for groups to exist or emerge in a population, possibly as a byproduct of existing reputation systems [Rosenblat *et al.*, 2017; Gross and De Dreu, 2019], and for agents to consider group affiliation alongside reputations when making decisions. Notably, when social norms governing reputations display in-group bias or out-group prejudice, even group-blind decisions through reputations can lead to inequality. Interactions in online marketplaces, where reputations are key, offer a paradigmatic example. The short-stay rental platform Airbnb, for instance, displays ratings and profile pictures for both hosts and guests. Recent findings show that guests with a distinctly African American name have a 12% lower acceptance rate [Edelman *et al.*, 2017].

Although we observe cooperation dependent on the joint effect of reputation and group identity, the interaction of the two mechanisms has not yet been formalised. In this paper, we fill this gap by answering whether independent agents that consider both reputations and group affiliations can learn to cooperate in an equitable manner despite biased social norms. To answer this question, we employ two models which provide complementary perspectives: one focuses on analytically identifying norms that stabilise cooperative and fair states; the other focuses on how populations of reinforcement learning agents can reach the stable states identified. As detailed below, we find that fairness and cooperation can be achieved if the right norm is chosen to judge actions.

1.1 Structure of Paper and Contributions

After introducing related literature on cooperation in mixed-motive games and indirect reciprocity (**section 2**), subsequent sections offer the main contributions of our paper:

- **Formalisation of a new model to systematically study cooperation and fairness under indirect reciprocity (section 3).** We introduce group identities into an existing evolutionary game theory model to formalise differing treatment of in and out-group interactions by social norms. Our model allows us to study the evolutionary stability of cooperative and fair strategies under different norms **(3.1)** and learning dynamics in a population of identical, but independent, Q-learning agents (**3.2**).
- **Stability analysis of norm-strategy combinations (subsection 4.1).** We show that if the majority-identity defects, then the minority-identity cannot sustain cooperation even amongst themselves, but the inverse does not hold. Moreover, social norms that favour in-group interactions can sustain relatively high fairness and cooperation given a "counterweight" strategic bias. We go on to explore the effectiveness of well-studied norms when paired up to judge in-group/out-group interactions.
- **Analysis of cooperation and fairness dynamics of under Q-learning and group-structured populations (subsection 4.2).** We investigate the impact of RL on learning cooperative and fair policies under specific norms. We show that, although cooperation decreases on average in an independent RL setting, agents can learn fair cooperation. Despite this, we show that reaching fair/cooperative states when the benefits of cooperation are low requires an initial fraction of cooperative agents.

Finally, in **section 5**, we discuss the limitations of our results and possible directions for future work. [1]

2 Related Work

2.1 The Cooperation Dilemma

Understanding human cooperation is a fundamental research topic across disciplines [Pennisi, 2005; Fehr and Fischbacher, 2003; Rand and Nowak, 2013]. In AI, there is a growing interest in designing artificial agents to be cooperative and generate cooperation in others. In a recent commentary [Dafoe *et al.*, 2021], the authors argue that AI requires "social understanding" and the ability to cooperate in order to achieve success in tasks that require complex interactions such as navigating pavements, financial markets, and online communication. Many tasks that AI engage with also require cooperation with humans or other AI. Recent works have explored mechanisms to help enable cooperation. The proposed methods include agents with inequality-aversion [Hughes *et al.*, 2018], rewarding causal influence [Jaques *et al.*, 2019], self-play [Anastassacos *et al.*, 2021], gifting [Lupu and Precup, 2020] or introducing non-adaptive, pro-social agents [Santos *et al.*, 2019; Anastassacos *et al.*, 2021; Guo *et al.*, 2023].

2.2 Social Norms

Previous work explores how coordination techniques and social governance can influence the autonomy of agents in a system and, for example, change the system's levels of cooperation. Social norms implemented in computational systems can help in this endeavour [Savarimuthu and Cranefield, 2011]. Two main classes of social norms are identified by [Villatoro *et al.*, 2010]: "essential" norms that seek to solve cooperation dilemmas and collective action problems [Griffith, 2010; Peleteiro *et al.*, 2014] and "conventional" – used to establish a convention, solving coordination dilemmas [Shoham and Tennenholtz, 1997; Sen and Airiau, 2007; Morales *et al.*, 2013]. Our work focuses on *essential* norms. It is also common to divide norms into top-down "legalist" approaches, in which norms are designed offline and imposed by a central authority and bottom-up "interactionist" approaches, where norms are emergent phenomena (e.g., as defined in [Haynes *et al.*, 2017]). Recent work explores the bottom-up creation of norms via so-called norm entrepreneurship [Anavankot *et al.*, 2023] or observing public sanctions [Vinitsky *et al.*, 2023]. Our contribution involves both: while we apply norms top-down, their effectiveness is computed via a bottom-up process, where strategies evolve over time.

2.3 Indirect Reciprocity in Multi-Agent Systems

Indirect reciprocity (IR) has been proposed as a mechanism to elicit cooperation among reinforcement learning agents. Anastassacos *et al.* [2021] examine whether agents with private social norms learned through Q-learning can reach a socially optimal consensus on how reputations should be interpreted and updated. To aid this learning, the researchers propose seeding the population with non-learning agents and through introspective self-play. They find that a combination of both mechanisms can sustain cooperation. Differently from Anastassacos *et al.* [2021], here agents play a *one-sided*

[1] Appendix and code available at: www.github.com/sias-uva/indirect-reciprocity. An extended abstract of this paper appears in the Proceedings of AAMAS'24 [Smit and Santos, 2024].

donation game as opposed to a *two-sided* donation game (a prisoner's dilemma), and we shift learning from the norm space to the strategy space. These decisions were made to align our paper with the extensive pre-existing IR literature.

While the asymmetrical nature of our game merely makes learning slightly less consistent due to delayed rewards, the shift from learning norms to strategies fundamentally changes the focus of the papers. In their paper, the goal is to internalise the reputation mechanism and examine its effects on learning cooperation. In contrast, we take the norm to be exogenous, introduce another variable upon which that agents can discriminate (group identity), and see how inequality can emerge in spite of cooperation.

2.4 Group-structured Populations

Some prior works studied IR in populations where agents explicitly belong to groups, through the lens of evolutionary game theory. In this domain, Kessinger *et al.* [2023] assume that different groups might use different social norms and focus on the effect of different information broadcasting mechanisms, whereby information about individuals can spread only between members of the same groups or publicly (as in traditional models). Contrary to the setting we explore here, Kessinger *et al.* [2023] assume that strategies only discriminate based on reputations and not group identity. The authors find that in such systems, cooperation ultimately depends on the rate of in/out-group interactions, and cooperation can collapse if information remains within the same groups.

In a more recent work, Stewart and Raihani [2023] study how stereotypes might be formed through group reciprocity: the authors find that stereotyping can lead to *negative judgement bias*, in which individuals become pessimistic about the willingness of out-group members to cooperate. Although these works study dynamics of IR under reinforcement learning [Anastassacos *et al.*, 2021] and dynamics of reciprocity associated with group identities [Stewart and Raihani, 2023], the combination of indirect reciprocity, group identity and reinforcement learning remains under-explored. In this paper, we propose a model that contributes to fill this gap.

3 Model

In this paper, a well-mixed population of agents, stratified into two groups, interact by playing a donation game. In the donation game, a player is either a donor or a recipient. The donor must decide whether to pay a cost to confer a benefit to their partner, and we assume that $b > c > 0$.

In the one-shot game, the dominant strategy for the donor is to not donate, i.e. to defect. To encourage cooperation, we consider reputations and social norms. Social norms encode the rules of society that confer agents with a corresponding reputation based on whether these rules are followed or broken. In our model, these rules can depend on the action taken by the donor, the current reputation of the potential recipient (their "goodness"), and whether the two agents are in the same group (their "sameness"). Following prior works on indirect reciprocity, all of these inputs are binary: a donor has two actions, reputations are either "good" or "bad", and agents can be in the same or a different group.

Donor action	0	0	1	1	Input bits
Recipient reputation	0	1	0	1	
Shunning (SH)	0	0	0	1	Reputation
SternJudging (SJ)	1	0	0	1	
ImageScoring (IS)	0	0	1	1	
SimpleStanding (SS)	1	0	1	1	

Table 1: Social norms assign a reputation given different combinations of actions and recipient reputations. The input bits above represent the context in which an action takes place. By convention, 1 means Good/Cooperate and 0 means Bad/Defect. In our model, a norm is composed of two four-bit norms: one to judge in-group interactions and one to judge out-group interactions.

Although we study 2^8 different social norms, in Table 1 we give examples of commonly studied norms. One is *SternJudging*, explored in detail by Pacheco *et al.* [2006], which deems that it is good to defect against a bad agent or cooperate with a good one, and that doing the opposite action in either case is bad. Another is *SimpleStanding*, which says that the only bad thing to do is to defect against a good agent [Panchanathan and Boyd, 2003]. Neither of these norms take into account the group relation of the agents involved, but a norm that judged 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. This specific norm, which could be called *Stern/Stand*, judges in and out-group interactions differently, so we refer to it interchangeably as *unfair* or *discriminatory* and the others mentioned as *fair* or *group-agnostic*. Fair and unfair norms are mutually exclusive.

We assume that agents can make execution errors. An agent who intended to cooperate will sometimes defect with some probability ϵ. Following previous works, agents have zero probability of cooperating when intending to defect. Similarly, we assume that third-party observers using social norms to assign reputations can also err and assign the opposite reputation than intended, with some probability δ.

We assume a public reputation scheme where reputations are common knowledge among all agents. Although we do not explicitly model the process of reputation diffusion, public reputations can exist due to gossiping about what happened over an efficient communication network or a central judge observing interactions and broadcasting reputations.

The way that agents decide how to act is determined by their strategy. Strategies define, for each combination of reputation and in/out-group, a corresponding action. As such, the space of strategies consists of functions σ such that

$$\sigma : \{0, 1\}^2 \to \{0, 1\} \quad (1)$$

In the Q-learning model (see below), the Q-table holds Q-values associated with each action for each binary pair of information. We refer to the (effective) strategy of an agent as the strategy that would result from applying the $\arg\max$ function to the Q-table on each possible input pair.

Some notable strategies include *AllD*, which unconditionally refuses to donate, *Disc*, which conditionally cooperates

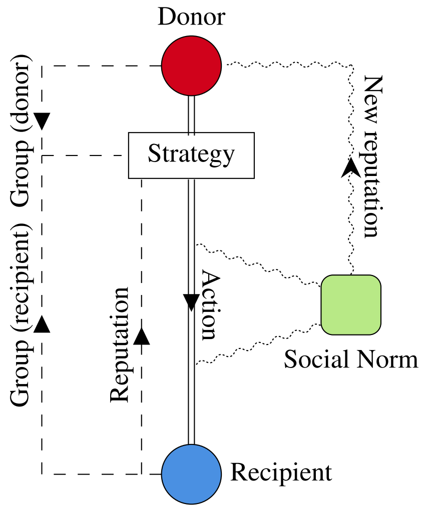

Figure 1: An illustration of an interaction between two agents which is observed by a third party using a social norm to update reputations. The donor's strategy determines the action taken based on the agents' relation to each other, and the reputation of the donor. The third party observes the action and context to assign the donor a new reputation based on a fixed social norm.

with good agents and defects against bad ones, and *AllC*, which unconditionally donates. Unfair strategies may instead, for example, play *AllC* with in-group members and play *Disc* with everyone else.

An example interaction is shown in Figure 1, in which the donor evaluates the recipient's reputation and group relation according to their strategy and takes the corresponding action. A third party observes the action taken by the donor, and its context, and assigns the donor a new reputation.

3.1 Evolutionary Stability

Given the model introduced above, a natural question one can pose is: Given a norm, which strategies are more likely to be played by agents in the long-run? One way this can be answered is with evolutionary game theoretical (EGT) tools. A key concern of EGT is that of the evolutionary stability of strategies: as agents play the donation game, their payoffs rise and fall based on their strategy and its impact on their reputation; a (resident) strategy is evolutionarily stable if agents using an alternative (mutant) strategy are unable to achieve higher average payoffs and invade.

Assume that strategies S_i have proliferated the entire incumbent population of each group i, and that the population is governed by social norm N. By calculating the expected payoff of a player of each group playing their strategy S_i, we can determine whether a mutant strategy in any group could outperform the incumbents of that group.

To introduce IR, we assume a timescale separation between the evolution of reputations and the arrival of mutants, i.e. allowing reputations to converge before introducing mutants. This allows us to derive and analytically solve the differential equations that give the long-term proportions of good agents in each group. Our line of reasoning is identical to that of Ohtsuki and Iwasa [2004], with the added complication of groups with interdependent reputations. The full derivation is available in Section A of the appendix.

After the reputations of players has stabilised, by introducing the utilities associated with each action, we can consider the long-run average payoffs of each population. Given fixed reputations, the average player from any group has a fixed probability of cooperating and being cooperated with every time they partake in an interaction according to each population's strategy, the relative size of each population, and the reputations of each population.

If we associate to each group i a benefit received b_i and cost of cooperating c_i, then we can determine the long term payoffs U_i of each group as

$$U_i = \sum_{j=1}^{2} \mathbb{P}(i \overset{\text{int.}}{\longleftrightarrow} j) \left(b_i \mathbb{P}(j \xrightarrow{\text{don.}} i) - c_i \mathbb{P}(i \xrightarrow{\text{don.}} j) \right) \quad (2)$$

where $\overset{\text{int.}}{\longleftrightarrow}$ means "interacts with", $\xrightarrow{\text{don.}}$ means "donates to", and the probabilities of donating are derived from the stationary reputations, strategies, and error rates of each group.

By doing a similar (yet simpler) derivation of mutant reputations and payoffs, one can determine if the combination of norms and strategies $(N, S_1, \dots, S_K)$ is an *evolutionarily stable state* (ESS).

Stronger than the traditional Nash equilibrium, a strategy S_I is evolutionarily stable on the condition that if any alternative strategy S_M arises in a group that S_I has proliferated and that the proportion of agents playing this alternative is sufficiently small, then this alternative strategy will perform worse than the incumbent strategy S_I and die out. We say that a combination of norm and strategies is an ESS if all of its strategies are evolutionarily stable. If all strategies in a combination are AllD, then the combination is trivially an ESS because any alternative strategy that cooperates with anyone would immediately be worse off.

While EGT and stability analysis informs which strategies are stable under each norm, we need to understand 1) how prevalent each equilibrium point is and 2) whether learning agents converge to a certain ESS. For both purposes, we use multi-agent reinforcement learning.

3.2 Reinforcement Learning

We model agents as independent (tabular) Q-learners. Due to the asymmetry in the donation game, while agents always learn from every interaction, the donor can only be negatively reinforced by possibly donating, and recipients can only be positively reinforced by possibly receiving a donation. Nevertheless, no matter the interaction, the Q-values of both parties decay due to the learning rate. We assume that recipients attributes whatever donations they may receive (or lack thereof) to their most recently taken action as a donor.

Formally, if agent i meets agent j who has x relation to agent i and y reputation, then the action taken by i (a^*) is determined by the equation

$$a^* = \underset{a \in \{0,1\}}{\operatorname{argmax}}\, Q_i[x, y, a] \quad (3)$$

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 $\hat{y}$

$$Q_i^{t+1}[x, y, a^*] \leftarrow (1-\alpha)Q_i^t[x, y, a^*] - \alpha a^* c \quad (4)$$
$$Q_j^{t+1}[\hat{x}, \hat{y}, \hat{a}] \leftarrow (1-\alpha)Q_j^t[\hat{x}, \hat{y}, \hat{a}] + \alpha a^* b \quad (5)$$

In our model, strategies take into account the in/out-group relation, and not which specific group another agent is from. When combined with imbalances in group size, benefit-to-cost ratio, and error rates, this minimal two-group model can capture the most pertinent fairness results while minimising complexity and the computational work necessary.

3.3 Metrics of Social Desirability

We measure the performance of a system with two metrics:

- The *cooperativeness* is the probability that, in a uniformly sampled interaction, the donor will cooperate.
- The *fairness* is the ratio between the average payoffs of the worst off and best off group. This is akin to the *demographic parity* ratio from supervised learning.

3.4 Experimental Setup

In all the following experiments, analytical or agent-based, the reader may assume the following parameters were used unless otherwise stated: The rate of agent execution errors and judgement execution errors is relatively rare at 1%, and the benefit-to-cost ratio in our analytical model is 5 with $c = 1, b = 5$. This represents a scenario where cooperation is highly beneficial (high b/c ratio), yet defection is still a dominant strategy ($c > 0$). Furthermore, the majority group comprises 90% of the population, and agents in different groups are functionally identical. These choices allow us to isolate the effects of discrimination and the effects of inherent differences between agents' in different groups. We refer readers to the appendix for a discussion of alternative parameter setups (group size, benefits, errors).

We run our RL experiments with a population of 50 agents (45 in the majority group and 5 in the minority group). We fix the exploration rate μ and learning rate α to 0.1. Each simulation runs for $250,000$ interactions and we run each simulation 50 times with a different seed. The source code for this paper (models, experiments, and figures) is available on GitHub.

4 Results

4.1 Stability Analysis

First, we evaluate the stability of all possible combinations of a norm and a strategy in each group, of which there are 2^{16} possibilities. We distinguish between defective strategies (Always defect/*AllD*), strategies ignoring group identity (Group-agnostic), and strategies discriminating based on group-identity (Discriminatory). By considering these categories we aim at developing an intuition for the possibility that fair and unfair cooperation emerge which, respectively, relies on stable Group-agnostic and Discriminatory strategies.

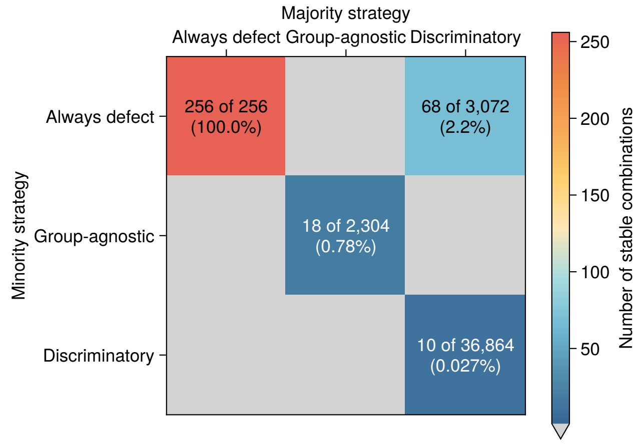

Figure 2: We evaluate all 2^{16} NSS (Norm-Strategy-Strategy) combinations of norms and strategies used by the two groups. We categorise every stable NSS according to the cooperative and discriminatory nature of strategies involved: strategies can defect regardless of the opponent type (Always defect), ignore group identity (Group-agnostic) or discriminate based on group identity (Discriminatory). We observe that group-agnostic strategies are unable to coexist with any other type of strategy, and a defecting majority playing *AllD* leads a minority to defect – but not the inverse. Parameters used: $b/c = 5$, error rate $= 0.01$. In the appendix (Figure 1) we confirm that the number of stable states remains unchanged for a wide range of error rates and b/c.

Figure 2 shows which strategies are stable under this setup. We observe that group-agnostic strategies by one group cannot sustainably coexist with discriminatory strategies in another, which means that a whole group ignoring group identities can make such identity discrimination unstable for everyone. Furthermore, we observe that the majority group can dictate whether cooperation can be stable: if the majority group unconditionally defects, there is no stable strategies where the minority cooperates; the inverse does not hold. As can be also seen in Figure 2, conditional cooperation from the majority group is a prerequisite for cooperation in the minority group.

In Figure 3 we explicitly compute the level of cooperativeness and fairness for each stable combination of strategies, given any norm fixed in the population. We observe that norms assigning reputations independently of group identities, lead to high levels of fairness, but not necessarily high levels of cooperation. Unfair norms, on the other hand, can lead to a cooperative yet unfair system, where a minority group always defects and the majority only cooperates with in-group members. Surprisingly, however, unfair norms can also sustain highly cooperative and fair systems, through the stability of group-agnostic strategies. This means that the observed levels of cooperation and fairness in a system can not trivially be inferred from the fairness level of a norm. Due to execution errors, the cooperativeness of a system can never achieve a perfect score of 100% even when all players always *intend* to cooperate. However, several norms are able

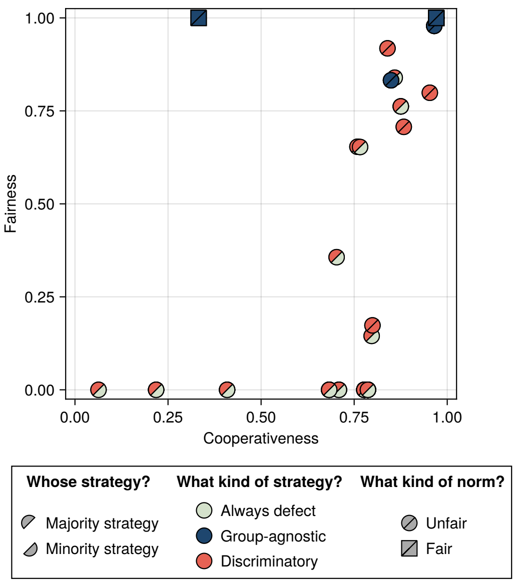

Figure 3: Fair norms (plotted as squares), which assign reputations independently of group identities, lead to fairness, but not necessarily high levels of cooperation (top-left). High cooperation and fairness can be stable with both fair and unfair norms (top-right quadrant). Parameters used: same as Figure 2.

to achieve equivalent levels of cooperation and fairness, and these are superimposed in the very top-right of the figure. When we inspect the common bits between these norms, we find that, regardless of whether the norm treats in-group and out-group the same, the two must be treated by either *SimpleStanding* or *SternJudging* (see Table 1).

Evaluating Well-known Social Norms

A number of "leading" norms, which can consistently stabilise cooperation, have been identified in previous works [Ohtsuki and Iwasa, 2004; Ohtsuki and Iwasa, 2006]. These norms agree that cooperating with good individuals is good, and defecting against good individuals is bad. In Figure 4,

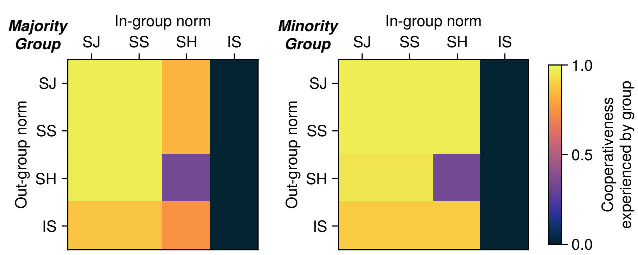

Figure 4: We use abbreviations from Table 1 to refer to the norms. Diagonal entries are previously studied "fair" norms. Overall, the majority group is most affected by the in-group norm, and vice-versa for the minority group. In-group-*ImageScoring* causes a total cooperation breakdown, whereas in-group-*Shunning* is most impacted when paired with out-group-*Shunning*, as agents have fewer ways to recover reputation.

we see the analytical levels of cooperation produced by each combination of norms when used to judge in- and out-group interactions. We find that, as in previous works [Ohtsuki and Iwasa, 2007], *SternJudging* and *SimpleStanding* (defined in Table 1) stabilise cooperation. However, with the addition of groups, we are also able to examine cooperation when interactions with outsiders are judged according to a different social norm. We observe that, as long as in-group interactions are judged according to *SternJudging* or *SimpleStanding*, out-group interactions can be evaluated according to other norms (e.g., the strict norm *Shunning*) without significantly decreasing cooperation. In fact, we observe that one can even use *ImageScoring* to judge out-group interactions and recover high levels of cooperation. This is remarkable given that this norm is simple and only relies on information about an action — a property that can be instrumental in settings where the reputations of out-group members are not widely accessible.

4.2 Learning Fair Cooperation

The previous (EGT) stability analysis informs which norms stabilise cooperation and, as a result, be effective in sustaining fair cooperation in groups of heterogeneous agents. Identifying norms that theoretically stabilise fairness and cooperation is a computationally attractive way of filtering norms that, when implemented in a system of learning agents, can possibly sustain fair cooperation. Learning fair cooperation in a finite population of adaptive agents requires, however, that agents are able to converge to a desirable equilibrium (which is naturally not trivial). Here, we explicitly model a population of independent RL agents. Figure 5 shows that even with a higher benefit-cost ratio ($b/c = 10$), some norms that were previously predicted to stabilise fair cooperation are not able to do so consistently under RL. Importantly, Figure 5 reveals that moving from an EGT to an RL analysis can affect one or both of cooperation and fairness, and to a different extent, depending on the norm in question. Furthermore, this effect is worsened when the dilemma is more difficult, and cooperation depends on the initial Q-values, as shown in Figure 6.

Although some norms might lead to fair and cooperative stable states, they differ in the size of the attraction basins that lead to such states. Furthermore, finite populations of learning agents are subject to stochastic effects that might prevent reaching fair and cooperative states.

Despite the challenges of converging to cooperative/fair equilibria, indirect reciprocity constitutes a promising mechanism to sustain cooperation and fairness in independent RL domains. We conclude that social norms such as *SternJudging* (and close variants) can be applied to multi-agent reinforcement learning domains to, in group-structured populations, to ensure both cooperation and fairness in the long-run. The effectiveness of this norm can be augmented by resorting to seeding agents, as previous works suggest [Anastassacos *et al.*, 2021]). We observe that a combination of indirect reciprocity and seeded agents can be harnessed to, not only sustain cooperation [Anastassacos *et al.*, 2021], but also supporting fairness in a heterogeneous group-structured population.

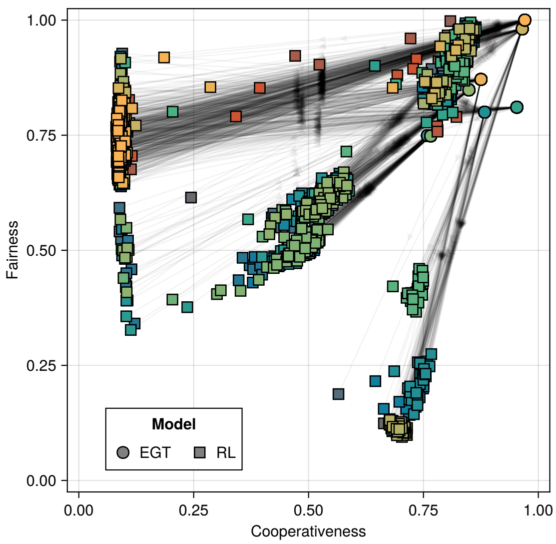

Figure 5: Some norms can sustain fair and cooperative equilibria, yet may be ineffective at guiding a population of independent RL agents to converge towards such states even with an elevated b/c ratio of 10, a trend which this figure demonstrates. We use different colours to represent different highly cooperative and fair norms (from the top-right quadrant of Figure 3). Circles represent EGT results, while squares represent RL results. For some norms, RL agents are unable to converge to the strategies that theoretically form an equilibrium, which leads to lower levels of fairness, cooperation, or both. Some norms, such as *SternJudging* (and variations) are impacted very little, and thus indicate that strictness is required for independent RL agents to sustain fairness and cooperation. In Section C of the appendix, we explore which strategies are learned to lead to these outcomes.

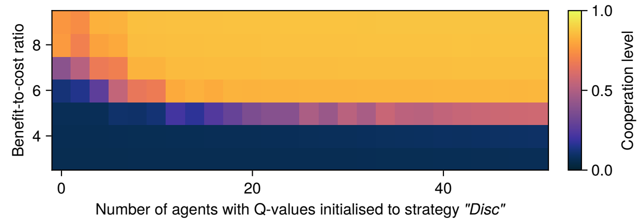

Figure 6: *SternJudging* can sustain high levels of cooperation in a multi-agent RL setting, but for stricter dilemmas (lower b/c ratios), convergence to a cooperative equilibrium is less likely and highly dependent on the initial distribution of Q-values. To demonstrate this, the x-axis shows that some agents were initialised with Q-values corresponding to the socially optimal strategy: ignoring groups and cooperating only with good individuals.

5 Conclusion

In this paper, we have shown that indirect reciprocity allows for fair cooperation among 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. We showed that a large variety of norms and strategies can be stable, with varying levels of cooperation and fairness. We showed that well-known norms (like *SternJudging*) perform well when agents adapt through reinforcement learning, being able to consistently achieve near to the idealised levels of cooperation and fairness predicted by our analytical model.

By using the minimal and generic donation game, we offer a proof of concept whose results may inform the application of indirect reciprocity to more elaborate multi-agent systems and motivates future work along several dimensions: While we make the common EGT assumption of a well-mixed population, a preference for interacting with similar individuals has been observed among humans [Fu *et al.*, 2012]. Furthermore, groups in our model are static. It has been shown that ad hoc group formation may occur in spatial mixed-motive models [Bara *et al.*, 2023] or in complex networks [Gross and De Dreu, 2019] where diverse local conventions can evolve [Hu and Leung, 2017]; investigating learning dynamics with a changing interaction structure or group labels could inform how to sustain cooperation and fairness in scenarios where group membership is dynamic.

Prior work focuses on norm emergence and dynamics [De *et al.*, 2017; Savarimuthu and Cranefield, 2011]. In the context of indirect reciprocity, work in this domain indicates that decentralised reputation systems with private information can hinder cooperation [Hilbe *et al.*, 2018], especially when the norm itself must also be learned [Anastassacos *et al.*, 2021]. This requires extra mechanisms to retain cooperation under private reputations [Krellner and Han, 2022; Kawakatsu *et al.*, 2024]. It may also be the case in a fully decentralised setting that the norm changes over time, something that may lead to an effective norm being chosen [Pacheco *et al.*, 2006] or a complete breakdown in cooperation [Xu *et al.*, 2019]. It would be interesting to explore how norms could be modified over time to maintain stable, fair cooperation between groups.

Despite these directions for future work, the model we formalise, and following comprehensive study, already sheds light on the advantages and challenges of indirect reciprocity in a minimal setting of a group structured population. With this modelling approach, we offer a link between algorithmic fairness – typically considered in the context of supervised [Mehrabi *et al.*, 2021] or unsupervised [Chierichetti *et al.*, 2017] learning – and multi-agent systems where reputations exist and social dilemmas of cooperation need to be solved. Our model provides a base framework to identify norms that, in such a context, sustain high levels of cooperation and fairness, thereby guaranteeing that universal cooperation is attained and parochialism avoided.

Acknowledgements

Co-funded by the European Union (ERC, RE-LINK, 101116987). Views and opinions expressed are however those of the author(s) only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

References

[Anastassacos *et al.*, 2021] Nicolas Anastassacos, Julian García, Stephen Hailes, and Mirco Musolesi. Cooperation and reputation dynamics with reinforcement learning. In *Proc. of the 20th International Conference on Autonomous Agents and MultiAgent Systems*. IFAAMAS, February 2021.

[Anavankot *et al.*, 2023] Amritha Menon Anavankot, Stephen Cranefield, and Bastin Tony Roy Savarimuthu. Towards norm entrepreneurship in agent societies. In *Advances in Practical Applications of Agents, Multi-Agent Systems, and Cognitive Mimetics. The PAAMS Collection*, Lecture Notes in Computer Science. Springer Nature Switzerland, 2023.

[Bara *et al.*, 2023] Jacques Bara, Fernando P. Santos, and Paolo Turrini. The role of space, density and migration in social dilemmas. In *Proc. of the 2023 International Conference on Autonomous Agents and Multiagent Systems*, AAMAS '23. IFAAMAS, May 2023.

[Chierichetti *et al.*, 2017] Flavio Chierichetti, Ravi Kumar, Silvio Lattanzi, and Sergei Vassilvitskii. Fair clustering through fairlets. *Advances in Neural Information Processing Systems, NeurIPS'17*, 30, 2017.

[Claus and Boutilier, 1998] Caroline Claus and Craig Boutilier. The dynamics of reinforcement learning in cooperative multiagent systems. In *Proc. of the Fifteenth National Conference on Artificial Intelligence*, AAAI '98. AAAI Press, 1998.

[Conitzer and Oesterheld, 2023] Vincent Conitzer and Caspar Oesterheld. Foundations of cooperative ai. *Proc. of the AAAI Conference on Artificial Intelligence*, 37, June 2023.

[Dafoe *et al.*, 2021] Allan Dafoe, Yoram Bachrach, Gillian Hadfield, Eric Horvitz, Kate Larson, and Thore Graepel. Cooperative ai: Machines must learn to find common ground. *Nature*, 593, May 2021.

[De *et al.*, 2017] Soham De, Dana S Nau, and Michele J Gelfand. Understanding norm change: An evolutionary game-theoretic approach. In *Proc. of the 16th Conference on Autonomous Agents and Multiagent Systems, AAMAS 2017*, 2017.

[Edelman *et al.*, 2017] Benjamin Edelman, Michael Luca, and Dan Svirsky. Racial discrimination in the sharing economy: Evidence from a field experiment. *American Economic Journal: Applied Economics*, 9, April 2017.

[Fatima *et al.*, 2024] Shaheen Fatima, Nicholas R. Jennings, and Michael Wooldridge. Learning to resolve social dilemmas: A survey. *Journal of Artificial Intelligence Research*, 79, March 2024.

[Fehr and Fischbacher, 2003] Ernst Fehr and Urs Fischbacher. The nature of human altruism. *Nature*, 425, October 2003.

[Fu *et al.*, 2012] Feng Fu, Martin A. Nowak, Nicholas A. Christakis, and James H. Fowler. The evolution of homophily. *Scientific Reports*, 2, November 2012.

[Genesereth *et al.*, 1986] Michael R. Genesereth, Matthew L. Ginsberg, and Jeffrey S. Rosenschein. Cooperation without communication. In *Proc. of the Fifth AAAI National Conference on Artificial Intelligence*, AAAI'86. AAAI Press, 1986.

[Griffith, 2010] Meghan Griffith. Why agent-caused actions are not lucky. *American Philosophical Quarterly*, 47, 2010.

[Gross and De Dreu, 2019] Jörg Gross and Carsten K. W. De Dreu. The rise and fall of cooperation through reputation and group polarization. *Nature Communications*, 10, February 2019.

[Guo *et al.*, 2023] Hao Guo, Chen Shen, Shuyue Hu, Junliang Xing, Pin Tao, Yuanchun Shi, and Zhen Wang. Facilitating cooperation in human-agent hybrid populations through autonomous agents. *iScience*, 26, 2023.

[Haynes *et al.*, 2017] Chris Haynes, Michael Luck, Peter McBurney, Samhar Mahmoud, Tomáš Vítek, and Simon Miles. Engineering the emergence of norms: A review. *The Knowledge Engineering Review*, 32, January 2017.

[Hilbe *et al.*, 2018] Christian Hilbe, Laura Schmid, Josef Tkadlec, Krishnendu Chatterjee, and Martin A. Nowak. Indirect reciprocity with private, noisy, and incomplete information. *Proc. of the National Academy of Sciences*, 115, November 2018.

[Hu and Leung, 2017] Shuyue Hu and Ho-Fung Leung. Local norm phenomena in multi-agent systems under community networks. In *Proc. of the 16th Conference on Autonomous Agents and Multiagent Systems, AAMAS'17*, 2017.

[Hughes *et al.*, 2018] Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, Heather Roff, and Thore Graepel. Inequity aversion improves cooperation in intertemporal social dilemmas. In *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018.

[Jaques *et al.*, 2019] Natasha Jaques, Angeliki Lazaridou, Edward Hughes, Caglar Gulcehre, Pedro Ortega, Dj Strouse, Joel Z. Leibo, and Nando De Freitas. Social influence as intrinsic motivation for multi-agent deep reinforcement learning. In *Proc. of the 36th International Conference on Machine Learning*. PMLR, May 2019.

[Kawakatsu *et al.*, 2024] Mari Kawakatsu, Taylor A Kessinger, and Joshua B Plotkin. A mechanistic model of gossip, reputations, and cooperation. *Proceedings of the National Academy of Sciences*, 121, 2024.

[Kessinger *et al.*, 2023] Taylor A. Kessinger, Corina E. Tarnita, and Joshua B. Plotkin. Evolution of norms for judging social behavior. *Proc. of the National Academy of Sciences*, 120, June 2023.

[Krellner and Han, 2022] Marcus Krellner and The Anh Han. Pleasing enhances indirect reciprocity-based cooperation under private assessment. *Artificial Life*, 27, March 2022.

[Lupu and Precup, 2020] Andrei Lupu and Doina Precup. Gifting in multi-agent reinforcement learning. In *Proc. of the 19th International Conference on autonomous agents and multiagent systems*, 2020.

[Mehrabi *et al.*, 2021] Ninareh Mehrabi, Fred Morstatter, Nripsuta Saxena, Kristina Lerman, and Aram Galstyan. A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54, July 2021.

[Morales *et al.*, 2013] Javier Morales, Maite Lopez-Sanchez, Juan A. Rodriguez-Aguilar, Michael Wooldridge, and Wamberto Vasconcelos. Automated synthesis of normative systems. In *Proc. of the 2013 International Conference on Autonomous Agents and Multi-Agent Systems*, May 2013.

[Nowak and Sigmund, 2005] Martin A. Nowak and Karl Sigmund. Evolution of indirect reciprocity. *Nature*, 437, October 2005.

[Ohtsuki and Iwasa, 2004] Hisashi Ohtsuki and Yoh Iwasa. How should we define goodness?—reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology*, 231, November 2004.

[Ohtsuki and Iwasa, 2006] Hisashi Ohtsuki and Yoh Iwasa. The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology*, 239, April 2006.

[Ohtsuki and Iwasa, 2007] Hisashi Ohtsuki and Yoh Iwasa. Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. *Journal of Theoretical Biology*, 244, February 2007.

[Okada, 2020] Isamu Okada. A review of theoretical studies on indirect reciprocity. *Games*, 11, September 2020.

[Pacheco *et al.*, 2006] Jorge M. Pacheco, Francisco C. Santos, and Fabio A. C. C. Chalub. Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Computational Biology*, 2, December 2006.

[Paiva *et al.*, 2018] Ana Paiva, Fernando P. Santos, and Francisco C. Santos. Engineering pro-sociality with autonomous agents. *Proc. of the AAAI Conference on Artificial Intelligence*, 32, April 2018.

[Panchanathan and Boyd, 2003] Karthik Panchanathan and Robert Boyd. A tale of two defectors: The importance of standing for evolution of indirect reciprocity. *Journal of Theoretical Biology*, 224, October 2003.

[Peleteiro *et al.*, 2014] Ana Peleteiro, Juan C. Burguillo, and Siang Yew Chong. Exploring indirect reciprocity in complex networks using coalitions and rewiring. In *Proc. of the 2014 International Conference on Autonomous Agents and Multi-Agent Systems*, AAMAS '14. IFAAMAS, 2014.

[Pennisi, 2005] Elizabeth Pennisi. How did cooperative behavior evolve? *Science (New York, N.Y.)*, 309, July 2005.

[Rand and Nowak, 2013] David G. Rand and Martin A. Nowak. Human cooperation. *Trends in Cognitive Sciences*, 17, 2013.

[Rosenblat *et al.*, 2017] Alex Rosenblat, Karen E.C. Levy, Solon Barocas, and Tim Hwang. Discriminating tastes: Uber's customer ratings as vehicles for workplace discrimination. *Policy & Internet*, 9, 2017.

[Santos *et al.*, 2018] Fernando P. Santos, Jorge M. Pacheco, and Francisco C. Santos. Social norms of cooperation with costly reputation building. *Proc. of the AAAI Conference on Artificial Intelligence*, 32, April 2018.

[Santos *et al.*, 2019] Fernando P. Santos, Jorge M. Pacheco, Ana Paiva, and Francisco C. Santos. Evolution of collective fairness in hybrid populations of humans and agents. In *Proc. of the Thirty-Third AAAI Conference on Artificial Intelligence*, AAAI'19. AAAI Press, 2019.

[Santos *et al.*, 2021] Fernando P. Santos, Jorge M. Pacheco, and Francisco C. Santos. The complexity of human cooperation under indirect reciprocity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 376, 2021.

[Savarimuthu and Cranefield, 2011] Bastin Tony Roy Savarimuthu and Stephen Cranefield. Norm creation, spreading and emergence: A survey of simulation models of norms in multi-agent systems. *Multiagent and Grid Systems*, 7, January 2011.

[Sen and Airiau, 2007] Sandip Sen and Stéphane Airiau. Emergence of norms through social learning. In *Proc. of the 20th International Joint Conference on Artifical Intelligence*, IJCAI'07. Morgan Kaufmann Publishers Inc., 2007.

[Shoham and Tennenholtz, 1997] Yoav Shoham and Moshe Tennenholtz. On the emergence of social conventions: Modeling, analysis, and simulations. *Artificial Intelligence*, 94, July 1997.

[Smit and Santos, 2024] Jacobus Smit and Fernando P. Santos. Fairness and cooperation between independent reinforcement learners through indirect reciprocity. In *Proc. of the 23rd International Conference on Autonomous Agents and Multiagent Systems*, 2024.

[Stewart and Raihani, 2023] Alexander J. Stewart and Nichola Raihani. Group reciprocity and the evolution of stereotyping. *Proc. of the Royal Society B: Biological Sciences*, 290, 2023.

[Villatoro *et al.*, 2010] Daniel Villatoro, Sandip Sen, and Jordi Sabater-Mir. Of social norms and sanctioning: A game theoretical overview. *International Journal of Agent Technologies and Systems (IJATS)*, 2, January 2010.

[Vinitsky *et al.*, 2023] Eugene Vinitsky, Raphael Köster, John P Agapiou, Edgar A Duéñez-Guzmán, Alexander S Vezhnevets, and Joel Z Leibo. A learning agent that acquires social norms from public sanctions in decentralized multi-agent settings. *Collective Intelligence*, 2, April 2023.

[Xu *et al.*, 2019] Jason Xu, Julian García, and Toby Handfield. Cooperation with bottom-up reputation dynamics. In *Proc. of the 18th International Conference on Autonomous Agents and MultiAgent Systems*, AAMAS '19. IFAAMAS, May 2019.