

# Cooperation with Bottom-up Reputation Dynamics

Jason Xu

Faculty of Information Technology,  
Monash University  
Melbourne, Australia  
jxu139@student.monash.edu

Julian García

Faculty of Information Technology,  
Monash University  
Melbourne, Australia  
julian.garcia@monash.edu

Toby Handfield

SOPHIS, Monash University  
Melbourne, Australia  
toby.handfield@monash.edu

## ABSTRACT

Cooperation among selfish agents can be promoted by allowing agents to condition behavior on reputation. Social norms – dictating how agents update the reputations of others – are central in determining whether this mechanism is effective. In particular, norms that reward justified defection have been shown to promote cooperation. A major limitation of existing models is that they assume all agents adopt a uniform norm, in a top down fashion. Here we show that when agents can spontaneously adopt novel norms, a learning process will see them drift towards socially undesirable outcomes. We present a model where agents can choose both how to react to reputations and how to assign the reputations of others – making social norms emergent. In this scenario cooperation can only be achieved when the space of norms is severely restricted. In the real world, reputation systems have a mixed record. This is often attributed to the costly nature of assigning reputations, and the ability of agents to easily whitewash their reputations. Our result suggests that even if these issues are overcome, enabling cooperation via reputation is likely to require additional mechanisms or restrictions upon the norms of the agents in the system.

## CCS CONCEPTS

• **Computing methodologies** → **Multi-agent systems; Cooperation and coordination; Modeling and simulation;**

## KEYWORDS

Reputation; Cooperation; Dynamics; Evolutionary Game Theory

### ACM Reference Format:

Jason Xu, Julian García, and Toby Handfield. 2019. Cooperation with Bottom-up Reputation Dynamics. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019*, IFAAMAS, 8 pages.

## 1 INTRODUCTION

Fostering cooperation among self-interested agents is a challenge in natural as well as artificial systems [1]. If cooperation is costly, but the benefits of cooperation can be enjoyed by all parties, the temptation to pay no cost is a dominant strategy and cooperation collapses [27]. To allow cooperation to flourish, some mechanism that enables correlated interactions – so that cooperative types meet each other more frequently, or so they can treat uncooperative types differently – is required. One such mechanism is direct reciprocity, which allows agents that interact repeatedly to condition their

strategies on past interactions. In this setting, strategies such as Tit-for-tat can support cooperative equilibria [2, 7, 13, 38].

When agents are anonymous or do not interact repeatedly, they can instead rely on indirect reciprocity [23]. This mechanism depends on the existence of public reputations, which allow agents to favour those that enjoy a good reputation by cooperating more with them. The cost of cooperation can then be offset by the benefits of having a good reputation [16]. This idea of reputation is ubiquitous across computing applications, including distributed and multiagent systems [8, 41].

Reputation systems are used across many domains [10], but their application is not without issues [11]. These include white-washing reputations [3], eliciting specious feedback [14], or low rates of participation [28]. Models of indirect reciprocity can be used as important tools to design reputation-based systems [30, 31, 33]. Such models provide a dynamic account of how agents solve the challenge of cooperation in simple but illustrative scenarios. These dynamic features are crucial because these settings often lead to multiple equilibria where static models do not necessarily single out a prediction [16].

The framework of indirect reciprocity uses evolutionary game theory (EGT). In EGT agents are boundedly rational and, instead of solving for equilibrium, follow simple rules that are updated by a dynamic learning process [36]. Agents that perform well in a population are more likely to be imitated or copied in subsequent generations. Originally conceived in evolutionary biology, EGT is also relevant for cultural learning by processes of imitation [17] and as a device for predicting human behaviour [12]. Although these processes may seem idiosyncratic to biological modelling, its long-term convergence is qualitatively similar to related processes such as reinforcement learning, or other learning mechanisms that rely exclusively on individual information [5].

In models of indirect reciprocity agents learn how to react to the reputations of others from payoffs derived from a series of interactions with random partners [23]. A strategy therefore determines whether an agent with a particular reputation value is worth cooperating with. A separate process is required to update the reputations of the agents. In these models, a social norm [42] determines how agents will assign the reputations of others after each encounter. Thus, a norm is simply a function that produces a reputation value given a combination of factors. Accordingly, these models can be used to study and compare alternative social norms for reputation assignment.

Early work has shown that very simple norms that assign a good reputation for cooperating and bad reputation for defecting are unable to sustain high levels of cooperation, because they cannot distinguish ‘justified punishment’ from uncooperative behavior [22]. Later work has shown that, in any space of possible norms,

the set of norms that sustain high levels of cooperation share a small number of intuitive features: they discriminate against those who defect against those with good reputation, and they allow agents with good reputation to retain a good reputation if they defect against a ‘bad’ agent [24, 25]. These results have since been generalized, with some qualification, to a broad range of cases [e.g. 26, 29, 33].

Traditionally, models of indirect reciprocity follow a top-down approach [40]. The social norm to assign reputations is imposed on the system offline [34]. This assumption limits the scope of applications to situations where the behaviour of the agents can be artificially constrained in a centralised fashion. The point of departure for this paper is to extend the models of indirect reciprocity to analyse the prospects of cooperation if agents are free to adopt idiosyncratic methods of assigning reputation, thus modelling a system whereby norms emerge ‘bottom up’ [35].

Our paper addresses the problem of modelling selfish agents in a cooperative setting, who can choose whether to cooperate or not on the basis of public reputation information while also having freedom to choose how to report the reputations of those with whom they have interacted. We examine the dynamic process whereby agents learn to react to and assign the reputations of others. This investigation gives a more thorough assessment of the prospects for maintaining cooperation using a mechanism of indirect reciprocity in environments where agents are unconstrained and social norms are emergent. Our main result is that cooperation can only be sustained when the set of norms is restricted. The dynamic process of learning always allows for defection to take over cooperative norms and actions.

The rest of this paper is organised as follows: Section 2 describes the basic setup of indirect reciprocity. Section 3 describes the technique used to analyze bottom-up reputation dynamics. We present our results in Section 4 and discuss the implications in Section 5.

## 2 PRELIMINARIES

### 2.1 Prisoner’s dilemma

We use the prisoner’s dilemma (PD) game as a fundamental model of cooperation. For simplicity, we focus on a subset of PDs determined by two parameters,  $b$  and  $c$  [21], where the gain from defection is constant, regardless of the other player’s behavior. Agents can choose to cooperate at a cost  $c$ , bestowing a benefit  $b$  on the opponent. Alternatively they may defect, at no cost, possibly reaping benefits from the opponent if she cooperates. The payoff matrix is given by:  $\begin{pmatrix} 0 & b \\ -c & b-c \end{pmatrix}$ , with actions defect (0) and cooperate (1) respectively. For  $b > c > 0$  defection is dominant, and mutual defection is the only Nash equilibrium.

The techniques we describe will apply to any symmetric  $2 \times 2$  game, but we focus on cooperation in this setting as a benchmark application.

### 2.2 Strategies: Actions and Norms

Following [32], [33] and [31], we consider agents with binary reputations 0 and 1. For convenience we will sometimes refer to these labels as “bad” or “good” (respectively), without ascribing any particular meaning to them. Since agents condition their actions on

reputations, the full action set for a game with discrete binary reputations comprises 4 strategies as follows:

- *Defectors*: Always defect regardless of opponent reputation
- *Paradox*: Cooperates with “bad” opponents, and defects towards “good” opponents.
- *Reciprocators*: Defect towards “bad” opponents, and cooperates with “good” opponents.
- *Cooperators*: Always cooperate regardless of opponent reputation

We can denote the strategy set  $S = \{0, 1, 2, 3\}$ ; with the corresponding binary encoding of each strategy, determining the action towards good, and bad individuals respectively. For example, strategy 1 (Paradox) is  $(01)_2$ , which implies defecting against individuals with 1 reputation, and cooperating against individuals with 0 reputation. Strategy 2 (Reciprocator) is  $(10)_2$ , which implies cooperating against individuals with 1 reputation, and defecting against individuals with reputation 0.

Now we define the space of social norms, i.e., functions to assign reputations to others based on the outcome of the game and the reputations of those involved in each interaction. For a PD game, a second order norm would transform the action of the donor, and the reputation of the recipient, into a new reputation for the donor [32]. Since there are 4 combinations of actions and reputations, an assignment norm can be encoded as a bit-string of length 4; thus we obtain a space of 16 possible norms.

The norms are naturally encoded with integers,  $0, 1, \dots, 15$ . The interpretation is given by the binary representation as follows: The first bit determines what reputation to assign to an agent that cooperates against a good agent; the second bit determines what reputation to assign to agents defecting against good agents; the third bit determines the reputation of someone that cooperates against a bad agent, and the last bit determines what reputation to assign to someone who defects against a bad agent. Hence

$$d = (d_{G,C}, d_{G,D}, d_{B,C}, d_{B,D})_2.$$

For instance, norm 9 is encoded  $(1, 0, 0, 1)_2$  and only assigns a good reputation to agents that either defect with a bad agent (bit 0), or cooperate with a good agent. Norm 10,  $(1, 0, 1, 0)_2$  assigns a good reputation to those that cooperate and bad to those that defect, without regarding the reputation of the opponent. And so on.

### 2.3 Top-down social norms

We first describe the standard top-down approach, where agents all use the same norm to update reputations as decided by a central planner. Following [32], we consider a group of  $N$  agents playing a PD game with binary reputations as described in Section 2.2. We use the small mutation assumption, namely that new behaviors arise in the population at a slow enough rate that there are never more than two types present in the population at a given time [4]. At each time-step, random pairs are formed and the game is played. After each interaction, a randomly chosen agent in each interaction pair gets her reputation updated according to the specified norm. As is standard, we assume that agents are subject to execution errors, defecting when they intended to cooperate with small probability  $\epsilon$ . They may also commit errors while judging the reputations of others (with probability  $\chi$ ), or errors may occur when reputations

are updated at the community level (with probability  $\alpha$ ). To incorporate execution errors, we transform the strategy  $p \rightarrow (1 - \epsilon)p$ , and to incorporate public reputation error, we transform the norm  $d \rightarrow (1 - 2\alpha)d + \alpha$ .

An agent's new reputation is based on a report from the interaction partner – what we call “second-party reporting”. This assumption is of potential interest, given concerns about reputation systems based on imperfect observations [9]. The second-party case is a natural best-case scenario, justifying relatively low error rates in our model. We have also derived results for third-party reporting, but under the small mutation assumption which we use for our analytic derivations, because there are only two types in the population at any given time, there is no interesting difference between the approaches.

The agents repeat this process for a number of rounds. Due to errors, the process is an ergodic discrete Markov Chain whose stationary distribution can be completely characterized. For a given population, this probability distribution specifies what is the probability of each reputation configuration. Given that the distribution of reputations and strategies determines the distribution of payoffs, we can thus compute the expected payoff in the long run for each type of player in a given population.

To model the dynamic behaviour of agents, [32] use an imitation process [37]. This is also a discrete Markov chain, whose stationary distribution determines which strategies are prevalent in the population, and therefore, the level of cooperation in the system.

The conclusion from [32] is that only 2 out of 16 possible norms enable cooperation: “Stern Judging” (SJ) and “Simple Standing” (SS). SJ assigns a good reputation to those cooperating with good partners as well as those defecting with bad partners, and assigns bad reputation otherwise; it is encoded  $9 = (1001)_2$ . SS assigns a good reputation to those cooperating – regardless of the partner's reputation – as well as those defecting against bad individuals; it is encoded  $11 = (1011)_2$ . The important properties shared by these norms are that they (i) reward cooperation with other individuals with a “good” reputation, and (ii) they are discriminating with regard to defection: they regard defection against bad individuals as good, and defection against good individuals as bad. In models investigating larger spaces of norms, these same properties are observed to be present in all and only the norms which promote cooperation at high levels [25]. In a top-down set-up, accordingly, the only way to obtain cooperation is for a central planner to enforce a cooperative norm such as SJ [32].

### 3 BOTTOM-UP SOCIAL NORMS

We now describe the dynamical processes for reputation and learning. These processes are assumed to happen at different timescales, whereby agents adopt a fixed behavioral strategy for a sufficiently long time that the reputation distribution stabilizes before agents exploring other strategies. Elsewhere – [32], [33] – it has been shown that this numerical technique closely aligns with agent-based simulations.

#### 3.1 Reputation dynamics

In the case of bottom-up social norms, each agent is characterized by the pair  $(p, d)$ , where  $p$  is one of the behavioral strategies  $S$ , and

$d$  is the norm used by the agent to report reputation. A strategy is therefore a pair  $(p, d)$  and the strategy space now comprises  $16 \times 4 = 64$  strategies, instead of 4 in the top-down setting. We will refer to the space of all possible pairs  $(p, d)$  as  $\hat{S}$ .

$$\hat{S} = \{(p, d) | p \in \{0, 1, 2, 3\}, d \in \{0, 1, 2, \dots, 15\}\}$$

After a pair  $[(p, d), (p', d')]$  interacts, each agent employs its norm to report a reputation to the broader community. Taking into account the various possible errors, and where  $C_x^p$  is the probability that an agent with strategy  $p$  cooperates with a partner of reputation  $x$ , the probability the agent being assessed by norm  $d$  will acquire a good reputation is given by:

$$G_x^{p,d} = (1 - \chi)(C_x^p d_{x,C} + (1 - C_x^p) d_{x,D}) + \chi(C_x^p d_{1-x,C} + (1 - C_x^p) d_{1-x,D})$$

The probability of being assigned a bad reputation is  $1 - G_x^{p,d}$ .

With the above expressions, we can now track the changes in the number of good and bad individuals within the population by developing a transition matrix for the corresponding finite Markov chain. For a population of size  $Z$ , we assume there are  $k$  individuals using strategy  $(p, d)$  and  $Z - k$  individuals using strategy  $(p', d')$ . We denote the number of individuals with good reputation of each type as  $h, h'$ , respectively.

The probability of a transition to a state in which one additional individual using strategy  $p, d$  acquires a good reputation can be given by:

$$H_{p,d}^+(h, h') = \frac{k - h}{Z} \left( \frac{h}{Z - 1} G_G^{p,d} + \frac{h'}{Z - 1} G_G^{p',d'} + \frac{k - h - 1}{Z - 1} G_B^{p,d} + \frac{Z - k - h'}{Z - 1} G_B^{p',d'} \right)$$

And the probability of a transition to one additional agent who uses  $p, d$  acquiring a bad reputation can be given by:

$$H_{p,d}^-(h, h') = \frac{h}{Z} \left( \frac{h - 1}{Z - 1} (1 - G_G^{p,d}) + \frac{h'}{Z - 1} (1 - G_G^{p',d'}) + \frac{k - h}{Z - 1} (1 - G_B^{p,d}) + \frac{Z - k - h'}{Z - 1} (1 - G_B^{p',d'}) \right)$$

Equivalent expressions for  $p', d'$  are straightforward extensions of the above.

The matrix  $H$  is then given by:

$$H_{i,j} = \begin{cases} H_{p,d}^+(h_i, h'_i), & h_j = h_i + 1 \wedge h'_j = h'_i \\ H_{p,d}^-(h_i, h'_i), & h_j = h_i - 1 \wedge h'_j = h'_i \\ H_{p',d'}^+(h_i, h'_i), & h_j = h_i \wedge h'_j = h'_i + 1 \\ H_{p',d'}^-(h_i, h'_i), & h_j = h_i \wedge h'_j = h'_i - 1 \\ H^-(h_i, h'_i), & i = j \\ 0, & \text{otherwise} \end{cases}$$

where

$$H^-(h_i, h'_i) = 1 - H_{p,d}^+(h_i, h'_i) - H_{p,d}^-(h_i, h'_i) - H_{p',d'}^+(h_i, h'_i) - H_{p',d'}^-(h_i, h'_i)$$

Entries in the matrix  $(i, j)$  specify the transition probability from state  $(h_i, h'_i)$  to state  $(h_j, h'_j)$ , where  $i$  and  $j$  are indices that range over the full space of reputation configurations, for given  $Z, k$ . The resulting Markov process is ergodic and accordingly, we derive the stationary distribution  $\sigma$ , representing the long-run probability distribution over reputation states. This is obtained from the eigenvector of  $\mathbf{H}$  associated with eigenvalue 1.

### 3.2 Learning dynamics

To calculate payoffs, and therefore the learning dynamics, we first calculate the probabilities that a given agent makes a donation and receives a donation, give the configuration of reputations in the population  $(h, h')$ .

$$D_p(h, h') = \frac{h}{k} \left( \frac{h-1+h'}{Z-1} C_G^p + \frac{Z-h-h'}{Z-1} C_B^p \right) + \frac{k-h}{k} \left( \frac{h+h'}{Z-1} C_G^p + \frac{Z-h-h'-1}{Z-1} C_B^p \right)$$

The probability that an individual using strategy  $p$  receives a donation is  $R_p(h, h')$ .

$$R_p(h, h') = \frac{h}{k} \left( \frac{k-1}{Z-1} C_G^p + \frac{Z-k}{Z-1} C_G^{p'} \right) + \frac{k-h}{k} \left( \frac{k-1}{Z-1} C_B^p + \frac{Z-k}{Z-1} C_B^{p'} \right)$$

The payoff of individuals using strategy  $p$  is calculated by:

$$f_p(k, h, h') = bR_p(h, h') - cD_p(h, h').$$

Payoffs for individuals using strategy  $p'$  can be calculated in an analogous way. We can now calculate the average fitness of a strategy by summing over the stationary distribution of every possible reputation configuration.

$$\bar{f}_p(k) = \sum_{h>0} \sum_{h'>0} \sigma_{h,h'} f_p(k, h, h')$$

From these payoff values we now specify the learning process. Agents imitate those that are more successful as follows. At each time period, we choose two agents – a focal agent and a role model. The focal agent imitates the strategy of the role agent with probability  $I_{pr}$  defined as follows:

$$I_{pr} = [1 + \exp(\Delta f)]^{-1}$$

where  $\Delta f$  is the payoff difference between the focal and the role agent. With a small exploration probability, an imitator will switch to another random strategy in the strategy space. This allows for agents to occasionally try out new strategies. The dynamics can be specified by inspecting fixation probabilities. This is the probability that a single individual with strategy  $j$ , will take over a population of agents using strategy  $i$ , and is defined as follows:

$$\rho_{i \rightarrow j} = \left[ 1 + \sum_{l=1}^{Z-1} \prod_{k=1}^l \frac{T^-(k)}{T^+(k)} \right]^{-1}$$

where,  $T^\pm(k)$  is the probability that a population with  $k$  type  $i$  agents, and  $Z-k$  agents of type  $j$ , will have  $k \pm 1$  agents of type  $i$  in the next time-step. These are straightforward to compute on the basis of  $I_{pr}$  [37]. The Markov chain with transition matrix  $M =$



Figure 1: Cooperation top-down vs bottom up.

$[\rho_{i \rightarrow j}]_{ij}$  approximates the imitation dynamics [4]. The long term evolutionary dynamics can be summarized with  $\phi$ , the stationary distribution associated with  $M$ . This distribution is defined over the strategy space, and we denote the abundance of strategy  $i$  in the long run,  $\phi_i$ .

To measure the degree of cooperativeness in the population, we compute a cooperation index (CI), namely the weighted average of cooperative acts that takes place in each state of the dynamics [32]. We use the fixation probabilities we have just derived as weights.  $D$  is as described above, the probability of an arbitrary agent cooperating.  $\sigma^d(p, k)$  is the probability having  $k$  good individuals in a population under norm  $d$ .

Thus,

$$CI = \sum_{s \in \hat{S}} \phi_s \sum_{k=0}^Z D_{p_s}(k, 0) \sigma^d(p_s, k)$$

where, for any  $s \in \hat{S}$ ,  $s = (p_s, d_s)$ .

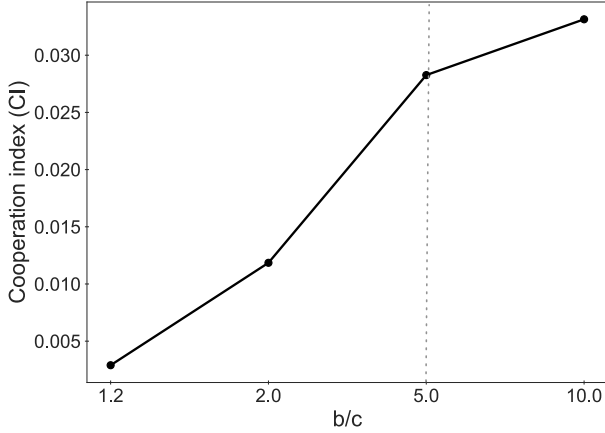
## 4 RESULTS

To simplify comparisons and for demonstration purposes we follow [32] in setting default parameters, with  $\epsilon = 0.08$ ,  $\alpha = \chi = 0.01$ ,  $b/c = 5$  and  $Z = 50$ . Results are qualitatively similar across a wide range of parameters. We first study the effect of bottom-up norms on overall cooperation; next, we look at the typical dynamics of imitation, and finally we inspect how different restrictions on norms affect cooperation.

### 4.1 Cooperation: Top-down vs Bottom-up

We start by measuring how the cooperation index changes when allowing agents to use emergent social norms. These results are depicted in Figure 1 as a function of the number of agents. For a benchmark comparison, we also plot cooperation achieved using the most effective cooperation-promoting norms as found in [32].

To quantify the loss of efficiency due to decentralization, in similar fashion to the price of anarchy [18], we measure the ratio of bottom-up cooperation over highest cooperation achieved top-down. The bottom-up approach can only reach up to 7% of the most efficient top-down social norm in small populations, and a meagre



**Figure 2: Bottom up CI for increasingly cheap cooperation.**

3% for more than 50 agents. The collapse of cooperation is evident across all population sizes.

Interestingly, for bottom-up reputation  $CI$  is monotonically decreasing with the number of agents in the range studied. This stands in contrast to our top-down approach, where larger groups lead to more cooperation. Nonetheless, a top-down approach is vastly more effective in fostering efficient social outcomes. This holds even in the case of very cheap cooperation with large  $b/c$  ratios, as shown in Figure 2.

## 4.2 Dynamics of Bottom-up reputation

We now turn to understanding the collapse of cooperation. To do this, we step back from the measures of cooperation and inspect  $\phi$ . This is depicted in Figure 3 – the inserts show strategies as grouped by action ( $p$ ) and social norm ( $d$ ). Bottom-up reputation dynamics overwhelmingly favor defecting strategies who take up  $> 80\%$  of the distribution.

No particular norm is strongly favored in the long term. This implies the ability of successful top-down norms, such as SJ and SS, to withstand defection is lost when different norms are allowed to co-exist. To further understand how social norms lose ground in the population we can inspect the transition matrix of the imitation process  $M$ .

The matrix  $M$  contains the transition probabilities  $\rho_{i \rightarrow j}$ , between any two strategies  $i$  and  $j$ . If two strategies always get the same payoff, the chance that one of them will take over from one single mutant is  $\frac{1}{Z}$  [19]. We can classify each transition probability using this benchmark of neutrality as follows [20]:

- A population of type  $i$  agents can **repel invasions** by type  $j$  agents if  $\rho_{i \rightarrow j} < \frac{1}{Z}$ .
- A population of type  $i$  agents is **invaded** by type  $j$  agents if  $\rho_{i \rightarrow j} > \frac{1}{Z}$ .
- A population of type  $i$  agents will **drift** to type  $j$  agents if  $\rho_{i \rightarrow j} = \frac{1}{Z}$ .

Using default parameters, Figure 4 shows this classification for all the entries in  $M$ . We highlight (with arrows) residents that are

able to repel invasions from defectors. These correspond to social norms that are potential catalysts of cooperation.

We find 8 strategies that can repel defection. Half of them use the paradox action strategy (i.e. defect against ‘good’ reputations and cooperate with ‘bad’ reputation opponents). Given that the reputation labels are devoid of meaning, these paradox actions foster cooperation by coordinating on the ‘bad’ = 0 label. The other half of the norms are reciprocators who coordinate to cooperate on the 1 label. The two types use symmetric norms that simply flip the meaning of the label leading to the same behaviour, so the mirror of a given norm  $d$  is  $(1 - d_{B,C}, 1 - d_{B,D}, 1 - d_{G,C}, 1 - d_{G,D})$ . This symmetry has been discussed elsewhere [24, 32].

Ignoring the symmetry that leads to equivalent behaviour, and focusing on reciprocator types that coordinate to cooperate on the 1 label, the pairs  $(p, d)$  that can repel defection are: (2, 8), (2, 9), (2, 10) and (2, 11). The corresponding social norms are: Shunning (SH), which only assigns a good reputation to those that cooperate with good opponents; Stern Judging (SJ) which in addition rewards justified defection; Image scoring (IS) which disregards reputations rewarding only cooperative acts; and Simple Standing (SS) which behaves like IS but also rewards justified defection.

While all of these norms are able to withstand the direct invasion of defectors, Figure 4 shows none of them are immune to invasion by neutral mutants. These mutants typically have the same behavioral strategy but alternative norms. Because an agent’s norm makes negligible difference to its payoff, two agents with the same behavior and different norms  $[(p, d), (p, d’)]$  will frequently be neutral with respect to each other. Through random processes of drift, therefore, it is always possible for a neutral mutant to take over the whole population, given enough time [19].

Neutral mutants who employ less effective norms are therefore harmful to cooperation because they create an indirect path for defectors to invade. Starting from a population entirely composed of reciprocators using an effective norm like SJ, a neutral mutant can always drift in with a bad norm, and then the population can be invaded directly by non-cooperative strategies. This stepping stone dynamic has been observed in models of direct reciprocity as well [6, 38]. Figure 5 depicts a typical transition from a good social norm into defection.

## 4.3 Restricting the space of norms

We now discuss how cooperation can be restored. While allowing for all norms destroys the prospects of cooperation, restricting the space of norms that can operate in the system can potentially bring some efficiency back. This exercise will also help us understand how certain norms may be more disruptive than others when working in tandem with norms that are known to foster cooperation.

To tackle this question, we analyze the dynamics arising from only allowing a small subsets of norms. Given a subset  $E \subset \hat{S}$ , we build the matrix  $M_E$  that describes all the transitions between strategies of the form  $(p, d)$  with  $d \in E$ . We then compute the stationary distribution, and the corresponding index  $CI$ . We further measure the ratio of bottom-up cooperation in subset  $E$  over highest cooperation achieved top-down. We do this for all subsets of size 2 and 3. Table 1 lists the subsets with a  $CI$  of 0.2 or higher – for

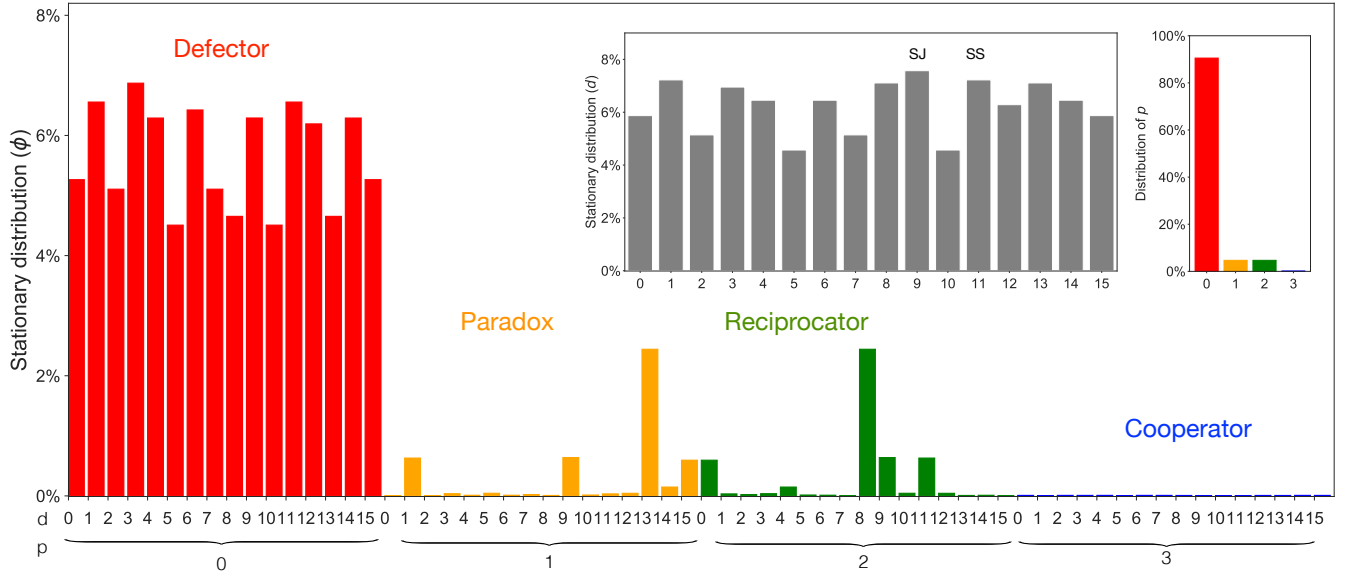


Figure 3: Long-term dynamics  $\phi$  with a bottom-up approach. Inset figures group strategies by norms and by actions

simplicity we have skipped the mirror combinations that lead to the same behaviour and achieve identical scores.

The top-performing subsets all include the SJ social norm. Most norms coincide with the set of norms that withstand defection as described above (SJ, SS, SH and IS). These norms are also very similar to each other. We can use the edit distance as a way to measure how consistent the norms are compared to each other –

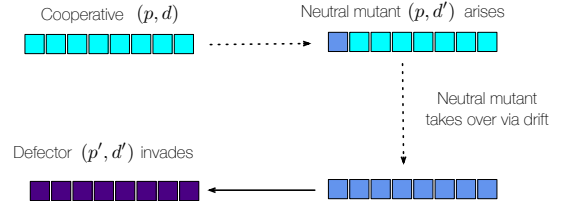


Figure 5: Typical dynamics from cooperation into defection

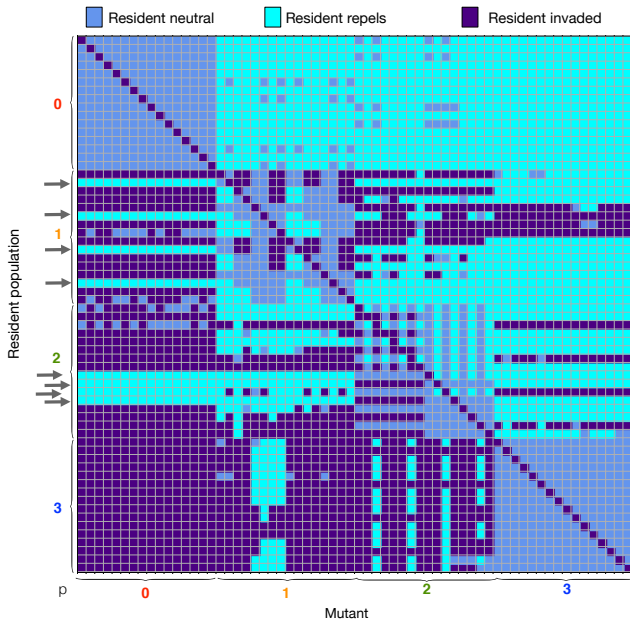


Figure 4: Transition matrix

1 or 2 on Table 1. Similar norms play along well with each other since reputations and actions employed are now aligned.

Using norms that are aligned can bring a substantial amount of cooperation back – up to 85% with SJ and SS. However, as we increase the number of norms cooperation collapses as inconsistencies between norms unavoidably arise. We can thus conclude that only a very restricted flavor of bottom-up dynamics is effective at keeping defection at bay.

Table 1: Cooperation with a subset of social norms

Subset of Norms	CI	% of optimal cooperation
(9, 11)	0.71	85
(8, 9, 11)	0.38	46
(9, 10, 11)	0.30	36
(9, 10)	0.30	36
(9, 13)	0.29	34
(8, 11)	0.27	33
(8, 9, 13)	0.25	30
(8, 9, 10)	0.23	27
(9, 11, 13)	0.2	23

## 5 DISCUSSION

We have formulated the dynamics of imitation for a game of cooperation based on indirect reciprocity. Unlike previous approaches, our formulation allows for emergent social norms, in which agents themselves determine how to update the reputations of others. In other words, we inspect the reputation dynamics with bottom-up social norms, and show why it is hard to sustain cooperation in this setting.

Our main result demonstrates that indirect reciprocity is, at best, a radically incomplete mechanism for the maintenance of cooperation in a decentralized fashion. Absent an additional factor such as population structure, direct reciprocity, or policing of reputation reporting, cooperation collapses when agents are free to innovate and adopt alternative social norms.

The exact results predicting the collapse of cooperation rely on two simplifying assumptions: the exploration rate used by the agents is low, and reputations are assumed to change faster than strategies. Relaxing these assumptions is unlikely to restore cooperation. In top-down models, it has been shown that higher exploration rates tend to harm cooperation [29]; and that the timescale separation between learning and reputation assignment does not have a strong effect in the top-down approach [32]. There is no reason to suspect these assumptions would have the opposite effect in the model presented here. Our model also represents a “best case” for cooperation scenario by assuming reporting to be costless [31]. Notwithstanding that, we still find cooperation is significantly impaired.

In a broader context, our result can be illuminated by observing that a reputation system is itself a variety of public good. A system of communication that allows effective discrimination between friend and foe is fragile, and requires ongoing maintenance to be valuable. Incorrect reports, relative to the prevailing system, are a form of informational pollution: degrading the cooperative environment. In this model, the challenge might appear to be much easier to surmount, because reporting reputation is costless. But because there is no penalty for harming the reputational environment (by, e.g. adopting a mutant norm), mere processes of neutral drift suffice to make cooperative equilibria unattainable.

Public goods games are a particularly difficult variety of cooperative problem to solve, because they distribute responsibility for maintaining a public benefit over numerous agents. Without a coordinating mechanism to hold individual agents to account, defection is likely to be an optimal strategy, and cooperation collapses. This challenge affects not just reputational systems, but any system of costless communication where there are potential conflicts of interest [15], and it remains a pressing puzzle to understand how to replicate the successes of human communication in artificial systems. With the increasing presence of artificial agents in the human infosphere [39], understanding how to use indirect reciprocity to preserve cooperation is an urgent challenge.

## ACKNOWLEDGMENTS

This work was supported by resources provided by the Pawsey Supercomputing Centre with funding from the Australian Government and the Government of Western Australia.

The authors would particularly like to thank Dr. Alexis Espinosa-Gayosso at Pawsey Supercomputing Centre.

Toby Handfield’s research on this project was supported by the Australian Research Council (DP150100242).

## REFERENCES

- [1] R. Axelrod. 1984. *The evolution of cooperation*. Basic Books, New York, NY.
- [2] Jakob Foerster, Richard Y Chen, Maruan Al-Shedivat, Shimon Whiteson, Pieter Abbeel, and Igor Mordatch. 2018. Learning with opponent-learning awareness. In *Proceedings of the 17th International Conference on Autonomous Agents and Multi-Agent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 122–130.
- [3] Eric J Friedman\* and Paul Resnick. 2001. The social cost of cheap pseudonyms. *Journal of Economics & Management Strategy* 10, 2 (2001), 173–199.
- [4] D. Fudenberg and L. A. Imhof. 2006. Imitation processes with small mutations. *Journal of Economic Theory* 131 (2006), 251–262.
- [5] D. Fudenberg and D. Levine. 1998. *The Theory of Learning in Games*. MIT Press, Cambridge MA.
- [6] Julián García and Matthijs van Veelen. 2016. In and out of equilibrium I: Evolution of strategies in repeated games with discounting. *Journal of Economic Theory* 161 (2016), 161–189.
- [7] Julian Garcia, Matthijs van Veelen, et al. 2018. No strategy can win in the repeated prisoner’s dilemma: Linking game theory and computer simulations. *Frontiers in Robotics and AI* 5, 102 (2018).
- [8] Jones Granatyr, Vanderson Botelho, Otto Robert Lessing, Edson Emilio Scalabrín, Jean-Paul Barthès, and Fabricio Enembreck. 2015. Trust and reputation models for multiagent systems. *ACM Computing Surveys (CSUR)* 48, 2 (2015), 27.
- [9] Chris Haynes, Michael Luck, Peter McBurney, Samhar Mahmoud, Tomáš Vitek, and Simon Miles. 2017. Engineering the emergence of norms: a review. *The Knowledge Engineering Review* 32 (2017), 1–31.
- [10] Ferry Hendrikx, Kris Bubendorfer, and Ryan Chard. 2015. Reputation systems: A survey and taxonomy. *J. Parallel and Distrib. Comput.* 75 (2015), 184–197. <https://doi.org/10.1016/j.jpdc.2014.08.004>
- [11] Kevin Hoffman, David Zage, and Cristina Nita-Rotaru. 2009. A survey of attack and defense techniques for reputation systems. *ACM Computing Surveys (CSUR)* 42, 1 (2009), 1.
- [12] Moshe Hoffman, Sigrid Suetens, Uri Gneezy, and Martin A Nowak. 2015. An experimental investigation of evolutionary dynamics in the Rock-Paper-Scissors game. *Scientific reports* 5 (2015), 8817.
- [13] L. A. Imhof and M. A. Nowak. 2010. Stochastic evolutionary dynamics of direct reciprocity. *Proceedings of the Royal Society B* 277 (2010), 463–468.
- [14] Radu Jurca and Boi Faltings. 2003. An incentive compatible reputation mechanism. In *E-Commerce, 2003. CEC 2003. IEEE International Conference on*. IEEE, IEEE, 285–292.
- [15] M. Lachmann and C. T. Bergstrom. 2004. The disadvantage of combinatorial communication. *Proceedings of the Royal Society of London B: Biological Sciences* 271, 1555 (Nov. 2004), 2337–2343. <https://doi.org/10.1098/rspb.2004.2844>
- [16] George J Mailath and Larry Samuelson. 2006. *Repeated games and reputations: long-run relationships*. Oxford University Press.
- [17] Richard McElreath and Robert Boyd. 2008. *Mathematical models of social evolution: A guide for the perplexed*. University of Chicago Press.
- [18] Noam Nisan, Tim Roughgarden, Eva Tardos, and Vijay V Vazirani. 2007. *Algorithmic game theory*. Cambridge University Press.
- [19] M. A. Nowak. 2006. *Evolutionary dynamics*. Harvard University Press, Cambridge MA.
- [20] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg. 2004. Emergence of cooperation and evolutionary stability in finite populations. *Nature* 428 (2004), 646–650.
- [21] M. A. Nowak and K. Sigmund. 1990. The evolution of stochastic strategies in the prisoner’s dilemma. *Acta Applicandae Mathematicae* 20 (1990), 247–265.
- [22] M. A. Nowak and K. Sigmund. 1998. The Dynamics of Indirect Reciprocity. *Journal of Theoretical Biology* 194, 4 (Oct. 1998), 561–574.
- [23] M. A. Nowak and K. Sigmund. 2005. Evolution of indirect reciprocity. *Nature* 437 (2005), 1291–1298.
- [24] H. Ohtsuki and Y. Iwasa. 2004. How should we define goodness?—Reputation dynamics in indirect reciprocity. *Journal of Theoretical Biology* 231 (2004), 107–20.
- [25] H. Ohtsuki and Y. Iwasa. 2006. The leading eight: Social norms that can maintain cooperation by indirect reciprocity. *Journal of Theoretical Biology* 239 (2006), 435–444.
- [26] J. M. Pacheco, F. C. Santos, and F. A. C. C. Chalub. 2006. Stern-judging: A simple, successful norm which promotes cooperation under indirect reciprocity. *PLoS Computational Biology* 2 (2006), 1634–1638.
- [27] A. Rapoport and A. M. Chammah. 1965. *Prisoner’s Dilemma*. University of Michigan Press, Ann Arbor.

- [28] Paul Resnick, Richard Zeckhauser, Eric Friedman, and Ko Kuwabara. 2000. Reputation systems. *Commun. ACM* 43, 12 (2000), 45–45.
- [29] Fernando P. Santos, Jorge M. Pacheco, and Francisco C. Santos. 2016. Evolution of cooperation under indirect reciprocity and arbitrary exploration rates. *Scientific Reports* 6 (2016), 37517.
- [30] Fernando P. Santos, Jorge M. Pacheco, and Francisco C. Santos. 2017. Indirect reciprocity in finite populations of explorative agents. In *Adaptive Learning Agents Workshop at AAMAS: ALA 2017*. International Foundation for Autonomous Agents and Multiagent Systems.
- [31] Fernando P. Santos, Jorge M. Pacheco, and Francisco C. Santos. 2018. Indirect Reciprocity and Costly Assessment in Multiagent Systems. In *Proceedings of The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18)*. Association for the Advancement of Artificial Intelligence, AAAI.
- [32] Fernando P. Santos, Francisco C. Santos, and Jorge M. Pacheco. 2016. Social Norms of Cooperation in Small-Scale Societies. *PLoS Computational Biology* 12, 1 (2016), 1–13. <https://doi.org/10.1371/journal.pcbi.1004709>
- [33] Fernando P. Santos, Francisco C. Santos, and Jorge M. Pacheco. 2018. Social norm complexity and past reputations in the evolution of cooperation. *Nature* 555, 7695 (March 2018), 242–245.
- [34] Yoav Shoham and Moshe Tennenholtz. 1995. On social laws for artificial agent societies: off-line design. *Artificial intelligence* 73, 1-2 (1995), 231–252.
- [35] Yoav Shoham and Moshe Tennenholtz. 1997. On the emergence of social conventions: modeling, analysis, and simulations. *Artificial Intelligence* 94, 1-2 (1997), 139–166.
- [36] K. Sigmund. 2010. *The Calculus of Selfishness*. Princeton University Press.
- [37] A. Traulsen and C. Hauert. 2009. Stochastic evolutionary game dynamics. In *Reviews of Nonlinear Dynamics and Complexity*, Heinz Georg Schuster (Ed.). Vol. II. Wiley-VCH, Weinheim, 25–61.
- [38] M. van Veelen, J. García, D. G. Rand, and M. A. Nowak. 2012. Direct reciprocity in structured populations. *Proceedings of the National Academy of Sciences USA* 109 (2012), 9929–9934.
- [39] Onur Varol, Emilio Ferrara, Clayton A. Davis, Filippo Menczer, and Alessandro Flammini. 2017. Online human-bot interactions: Detection, estimation, and characterization. *arXiv preprint arXiv:1703.03107* (2017).
- [40] Daniel Villatoro, Sandip Sen, and Jordi Sabater-Mir. 2010. Of social norms and sanctioning: A game theoretical overview. *International Journal of Agent Technologies and Systems (IJATS)* 2, 1 (2010), 1–15.
- [41] George Vogiatzis, Ian MacGillivray, and Maria Chli. 2010. A probabilistic model for trust and reputation. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 225–232.
- [42] Michael Wooldridge. 2009. *An Introduction to Multiagent Systems*. John Wiley & Sons.