Outcome-based Partner Selection in Collective Risk Dilemmas

Fernando P. Santos[*]
Princeton University, Department of
Ecology and Evolutionary Biology
08544 NJ, USA

Samuel F. Mascarenhas
INESC-ID & Instituto Superior
Técnico, Universidade de Lisboa
Lisbon, Portugal

Francisco C. Santos
INESC-ID & Instituto Superior
Técnico, Universidade de Lisboa
Lisbon, Portugal

Filipa Correia
INESC-ID & Instituto Superior
Técnico, Universidade de Lisboa
Lisbon, Portugal

Samuel Gomes
INESC-ID & Instituto Superior
Técnico, Universidade de Lisboa
Lisbon, Portugal

Ana Paiva
INESC-ID & Instituto Superior
Técnico, Universidade de Lisboa
Lisbon, Portugal

ABSTRACT

Understanding how to design agents that sustain cooperation in multi-agent systems has been a long lasting goal in distributed Artificial Intelligence. Solutions proposed rely on identifying defective agents and avoid cooperating or interacting with them. These mechanisms of social control are traditionally studied in games with linear and deterministic payoffs, such as the Prisoner's Dilemma or the Public Goods Game. In reality, however, agents often face dilemmas in which payoffs are uncertain and non-linear, as collective success requires a minimum number of cooperators. These games are called Collective Risk Dilemmas (**CRD**), and it is unclear whether the previous mechanisms of cooperation remain effective in this case. Here we study cooperation in **CRD** through partner-based selection. First, we discuss an experiment in which groups of humans and robots play a **CRD**. We find that people only prefer cooperative partners when they lose a previous game (*i.e.*, when collective success was not previously achieved). Secondly, we develop a simplified evolutionary game theoretical model that sheds light on these results, pointing the evolutionary advantages of selecting cooperative partners only when a previous game was lost. We show that this strategy constitutes a convenient balance between strictness (only interact with cooperators) and softness (cooperate and interact with everyone), thus suggesting a new way of designing agents that promote cooperation in **CRD**.

KEYWORDS

Cooperation; Collective Risk Dilemma; Game Theory; Partner selection; Human-Robot Interaction; Complex systems.

ACM Reference Format:
Fernando P. Santos, Samuel F. Mascarenhas, Francisco C. Santos, Filipa Correia, Samuel Gomes, and Ana Paiva. 2019. Outcome-based Partner Selection in Collective Risk Dilemmas. In *Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), Montreal, Canada, May 13–17, 2019,* IFAAMAS, 9 pages.

1 INTRODUCTION

Cooperation between self-interested agents has been a fundamental research topic in economics [12] and evolutionary biology [27]. Likewise, designing agents that sustain cooperation is a long-standing goal in multi-agent systems (MAS) [10, 20, 60]. Often agents take part in interaction paradigms that pose them the dilemma of choosing between maximizing individual gains or cooperating for the sake of social good. Studying cooperation is thereby significant for two reasons: on the one hand, to understand the biological and cultural mechanisms developed by humans (and other species) that allow altruism to evolve [35]; on the other hand, to learn how to engineer agents and incentive schemes that enable cooperation to emerge through decentralized interactions, thus allowing for social desirable outcomes that benefit all [31].

In some cooperative interactions, collective benefits are only distributed – or collective losses avoided – whenever a minimal number of agents decides to cooperate. Crucially, this interaction paradigm sits at the heart of climate negotiations, needed to prevent the hazardous consequences of climate change. Reducing greenhouse gas emissions stands as a costly action that, if done by a sufficient number of countries, allows preventing catastrophic outcomes and benefit everyone. This situation inspired the so-called Collective Risk Dilemma (**CRD**) [23, 38, 51]. In this game, a group of agents interact during a few rounds; in each round, subjects decide how much to invest, from their personal endowment, in order to prevent dangerous climate change (*i.e.*, the collective goal). The goal is achieved if the sum of all contributions amounts to a certain threshold. If the collective goal is achieved, every player keeps the money that was not invested. Otherwise, everyone looses the saved endowment with a given probability (therein lies the risk).

This situation is common to many collective action problems. It is possible to apprehend the core of the dilemma at stake, resorting to a simplified game. Following the model formalized in [38], we can assume that, in a group of size N, each agent starts with an endowment b and the actions available are Cooperate (**C**) or Defect (**D**). Cooperating means contributing with c (where $c < b$) towards the collective goal; Defecting means contributing 0. The collective goals are achieved if at least M agents cooperate. If the required number of contributions is not verified, everyone looses the saved endowment with a probability r (the risk). Assuming the most extreme scenario ($r = 1$) we can verify that, if everyone cooperates everyone earns $b - c$. If everyone defects – or, in general, if the threshold M of cooperators is not attained – everyone earns 0. Cooperation is noticeably desirable, however it may be hard to implement: the individual optimal occurs when the threshold M is achieved without requiring one's contribution. In this situation, a defector earns b, while a cooperator just keeps $b - c$.

[*] fpsantos@princeton.edu

Proc. of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019), N. Agmon, M. E. Taylor, E. Elkind, M. Veloso (eds.), May 13–17, 2019, Montreal, Canada.

A possible way of incentivizing cooperation in **CRD** is to punish defectors or reward cooperators [2, 9, 57]. These mechanisms, however, require the existence of costly monitoring institutions or the willingness of individuals to spend an amount to directly reward or punish an opponent. This origins the so-called *second-order free riding* problems. A more subtle way of eliciting cooperation in **CRD** is the avoidance of agents known to have previously behaved as defectors. In fact, mechanisms of such kind were previously applied in the context of pairwise interactions – chiefly, the Prisoner's Dilemma or the Ultimatum Game [7, 11, 28] – or multiplayer interactions with linear payoffs [16, 18, 26] – that is, payoffs that increase linearly with the number of contributors, without the abrupt changes that occur when a threshold of contributors is attained, as in the **CRD**. It remains unclear, however,

- How will individuals decide to select or avoid defectors given the previous success in **CRD** encounters? or, in general
- Whether defector avoidance constitutes an effective mechanism to elicit cooperation in **CRD**.

In this paper we explore these questions, focusing on partner selection in **CRD** through a human-robot experiment and an Evolutionary Game Theoretical (**EGT**) model [59].

First, we conduct an experimental study with humans and robots. Using robots allows us to fine-tune the strategies used and thus test explicitly a cooperative and defective opponent. We frame **CRD** in the form of a band selection game: agents are recruited to form a band and record albums. Cooperation means investing in mastering an instrument that contributes for the success of the band's album; defect means investing in self-marketing. An album is successful if a threshold album value is achieved – which is positively impacted by the instrument skill of each player. After interacting in a group with a cooperative and defective robot, we ask each human participant which robot would be selected for a future game. Surprisingly, we find that humans tend to select with significantly more frequency the cooperative opponent only when they faced a previous collective failure. If collective success is achieved, humans select cooperative or defective opponents alike.

Inspired by this result, we develop an **EGT** model that allows studying, in the context of an evolving population, the scenarios in which an Outcome-based Cooperative strategy (**OC**) prevails in the context of **CRD**. Individuals that use this strategy cooperate, yet only accept to play with defectors when they previously achieved collective success. We compare this strategy with a strategy that always Cooperates and always plays (**C**), a strategy that always Defects (**D**) and a strategy that cooperates but always refuses playing with defectors (coined Strict Cooperators, **SC**). We find that **OC** is the most prevailing strategy in a wide parameter region. In particular, **OC** wins against **SC** when threshold M is high and the cost of cooperating, c, is low. We find that **OC** conveniently combines the strict component of **SC** (refusing playing with **D**s) with the softness of **C**s. This allows agents using **OC** to concede playing with **D** opponents when their representativeness in the population is low enough to still guarantee reaching the collective threshold M within the majority of interaction groups.

With this model, we open a new route to study strategies that efficiently incentivize cooperation in **CRD**s through partner selection conditioned on own success experiences. In the next section we discuss several approaches to elicit cooperation through defector avoidance, mainly in the context of pairwise interactions or multiplayer games with linear payoffs. In section 3 we detail our experimental framework and provide the obtained results. Next, in section 4, we detail the theoretical model used to shed light on the results obtained experimentally. The theoretical results are further presented in section 5. We end with section 6, where we summarize our findings, point the limitations of our theoretical approach, and provide avenues for future work.

2 RELATED WORK

In this paper, we focus on a multiplayer social dilemma of cooperation previously named Collective Risk Dilemma (**CRD**), already alluded to in the previous section. This game was originally proposed in [23] with the goal of investigating decision-making in the context of greenhouse gas emission reduction and the avoidance of dangerous climate change. Later on, **CRD** was analyzed theoretically, resorting to Evolutionary Game Theory, **EGT** [38]. The authors found that, similarly to what was verified in the experiments, high-risk leads to higher contributions. Additionally, small group sizes were found to be particularly suitable to sustain cooperation. Here we follow the specification and notation in [38].

In the core of **CRD** lies a dilemma of cooperation, in which contributing to the collective target is at odds with individual interest. Even if missing the collective threshold has a huge impact in everyone's payoff, the decision to Defect – expecting that others contribute enough to achieve the collective goal – is the strategy that maximizes the individual payoffs of agents. As we explore in the present paper, several approaches to solve the dilemma of cooperation are based on mechanisms of defector identification and interaction avoidance. In the context of the iterated Prisoners' Dilemma, Mor and Rosenschein found that allowing agents to opt out from a repeated interaction opponent eases the dominance of cooperative strategies [24]. In that work, individuals interact repeatedly with the same opponents. Avoiding defectors can, alternatively, be accomplished through reputations or social network rewiring. In this context, Ghang and Nowak found that reputations and optional interactions can be combined such that cooperation evolves among self-interested agents, provided that the average number of rounds per agents is high enough [11]. In that work, a cooperator only accepts a game when the reputation of the opponent does not indicate her to be a defector. Also, for a game to take place, both agents must accept to play the game. An extension to private interactions was later suggested in [29]. Using reputations to adapt behaviors and punish unreasonably defective opponents is a principle that underlies indirect reciprocity. In this context, Griffiths showed that using reputations to discriminate and refusing to cooperate with defectors allows for cooperation to emerge [13]. An alternative mechanism to avoid interacting with defectors is social network rewiring. Santos et al. [39] found that this mechanism allows for cooperation to emerge even when the average degree of networks – a factor known to reduce cooperation in static networks – is constant, a result also found in lab experiments [8], and Public Goods Games [26]. Peleteiro et al. also used network rewiring, in combination with coalition formation, to promote the evolution of cooperative behavior [32]. Griffiths and Luck studied network rewiring as a way of improving tag-based cooperation [14]. More recently, Kumar et al. studied cooperation and network rewiring, focusing on how human-like motivations – such as sympathy, equality preferences and reciprocity – affect the resulting social network topology [5]. Following the same principle of avoiding interactions with defective opponents, Fernandez et al. studied anticipating mechanisms in the the context of Anticipation Games [7], an interaction paradigm proposed in [61]. In this case, agents refuse to play with agents if they were previously defective/unfair.

The previous works adopt strategies of defector avoidance in the context of 2-person games. In the context of multiplayer interactions, Hauert et al. found that simply introducing the opportunity for agents to opt out form a Public Goods Game (a strategy called Loner) creates a cycling dynamics that prevents the stability of defection. Interestingly, this strategy does not rely on knowledge about the strategies of others [18]. More recently, Han et al. studied Public Goods Games and commitments, assuming that agents may only accept to take part in an interaction group provided that a minimum number of group members decided committed to cooperate [16]. In all cases, agents are allowed to opt out from interaction groups, providing a possibility to dissuade defection.

So far, defector avoidance mechanisms were implemented in pairwise cooperation dilemmas (Prisoner's Dilemma) or multiplayer cooperation games with linear and deterministic payoffs (Public Goods Game). Here we address – both experimentally and theoretically – a new type of conditional strategies in the **CRD**, based on the overall group success. As mentioned, in the **CRD** the payoffs depend, ultimately, on a threshold value of contributions that must be achieved to guarantee group success. This said, the decision of agents to take part in groups with defective opponents might be based, not only on opponents' strategies, but also on the previous success/failure experienced. Strategies of this kind were seldom studied. Our work attempts to provide a first step in filling this gap.

The methods that we use to study **CRD** theoretically (Evolutionary Game Theory, **EGT**) were originally applied in the context of ecology and evolutionary biology [50]. Notwithstanding, previous works within **AI** (and particularly the **MAS** community) revealed that adopting a population dynamics perspective provides important insights regarding multi-agent learning and co-evolving dynamics [1, 19, 22, 54, 55]. **EGT** was also recently applied, for example, to study social norms along different directions, namely the stability of normative systems [25], the emergence and time evolution of social norms [6], or the evolution of cooperation through norms and reputations [43, 45, 46]. Finally, recent results suggest that partner selection can be a mechanism to coordinate actions of humans and agents, showing that past interactions with virtual agents shape the subsequent levels of human trust in virtual teammates [47]. In the next section we present, precisely, a Human-Robot experiment in the context of a cooperation dilemma.

3 EXPERIMENTAL RESULTS

In the field of Human-Robot Interaction (**HRI**) there has been a rising interest in exploring how people interact with robots in social dilemmas [37]. The motivation for this line of research is in part driven by the potential for robots to become true collaborative partners that have autonomy and become more than mere tools that obey our commands [15]. People will then need to build trust relationships with the robots they collaborate with. Using well established social dilemmas, such as a Public Goods Game, can provide important insights into the dynamics of these relationships.

With this motivation in mind, we developed a digital collaborative game – named *For the Record* – that was designed to be played by mixed human-robot teams. Thematically, the game consists of players assuming the role of musicians that form a band together with the goal of recording and selling successful albums. The game is composed by several rounds, each corresponding to the publication of an album on the market. Individuals may invest on instrumental skills (cooperate, with a direct impact on the success of the band), or on marketing skills (defect, with an impact on individual profit). In order for an album to succeed, its musical quality must surpass a threshold. Otherwise, the album is considered a failure and no one receives any profit. The uncertainty in the game is caused by the fact that the contributions to the album's quality, the value for the market's threshold, and the amount of profit made by each player are all determined by rolling a set of dice.

Specifically, from round to round (in a total of 5 rounds), each player decides to invest one die (6 faces) to improve the instrument skills or to improve self-marketing. In the last round, an album achieves success if its value surpasses a threshold, given by rolling 3 dice with 20 faces. The value of the album is given by the sum of trowing all dice invested by the 3 players along the 5 rounds. If the album achieves success, each individual will either earn 3 points or the sum of points given by the result of throwing the dice invested in marketing. The expected payoff of an individual that always Defects (and assuming that, nonetheless, the album achieved success) is $3.5 \times 5 = 17.5$. A player that always cooperates receives 3 and the dilemma lies in the difference between these payoffs: those that chose the first option (defect) will make the most profit but will hurt the band's capability of making successful albums consistently. This way, while payoff is only realized when an album's quality reaches a minimum threshold, the pressure to free-ride – defecting and relying on others' contributions to increase the album's quality – is high (as in the **CRD**).

Finally, the band has a fixed upper limit on the amount of albums that can fail. If such limit is reached, then the band collapses, causing the game to end prematurely and all the players lose their accumulated profits. This catastrophe condition reinforces the need for collaboration. Even if framed within a specific context, the nature of this dilemma is general enough to capture the non-linear (and uncertain) nature of many Human collective endeavors [38].

We experimentally tested *For the Record* using a 3-players setting, in which 2 robotic agents played with a human player. The goal was, not only to compare how people perceive robotic partners which apply different strategies to play this collaborative game, but also to evaluate which of such partners would people select for future partnerships. In particular, one of the robots (the collaborator) unconditionally opted to cooperate, whilst the other one (the defector) unconditionally opted to defect. Although we have hypothesized that different outcomes would lead to different perceptions of the team and its members, we expected individuals to reveal a significant preference for the cooperator robot .

The user study was conducted at a company facility where 70 participants with ages ranging from 22 to 63 ($M = 34.6, SD =$

11.557) were recruited. The task lasted for 30 minutes and consisted of 1) a briefing, 2) the game with the robotic players and 3) a survey. The dice rolls were scripted to manipulate the outcome of the game using a between-subjects design, which could either result in a *winning* or *losing* outcome. To assess how participants perceived the team and the robotic partners, several measures were applied (*e.g.*, trust, attribution of responsibility, social attributes). Moreover, participants were asked to select one of the two robotic partners, the cooperator or the defector, for a hypothetical future game.

The particular findings regarding the partner selection revealed a significant association between the preferred robot and the game result ($\chi^2(1) = 14.339, p < 0.001, \phi_c = 0.453$). A further analysis of the same preferences across conditions (see Fig. 1) showed the cooperator is significantly preferred over the defector after losing the game ($\chi^2(1) = 31.114, p < 0.01, r = 0.889$). However, no significant difference was found in the partner selection after winning the game ($\chi^2(1) = 1.400, p = 0.237, r = 0.040$). A detailed description and discussion of the remaining measures is presented in [4]. These findings inspired us to develop the following evolutionary game theoretical model to interpret the advantages of selecting cooperative partners only when a previous game was lost.

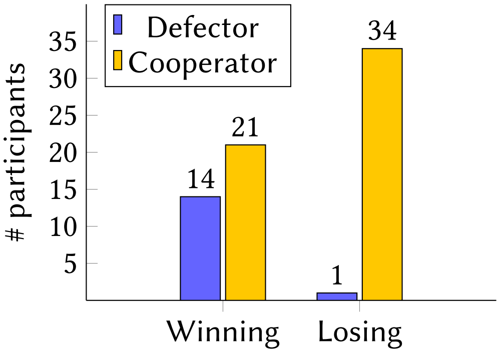

Figure 1: Behavioral experiments on partner selection grouped by conditions, *i.e.*, if collective goals were achieved in the last round (winning) or not (loosing). The results suggest that cooperative partners (yellow bars) are only preferred whenever collective success is not achieved. In *winning* **configurations, humans select the cooperative or defective opponents almost alike.**

4 THEORETICAL MODEL

In order to shed light on the advantages and disadvantages of such a strategy, we build a theoretical model based on evolutionary game theory. Let us assume a population with Z agents. Maintaining the barebones of the dilemma at stake in *For The Record*, we focus our attention on the previously introduced Collective Risk Dilemma (**CRD**) [23, 38]. Two baseline strategies are possible in this multiplayer game: Cooperate and Defect. The Cooperators (**C**) pay a cost (c) in order to contribute to a collective endeavor (album quality, in the previous scenario). The Defectors (**D**) refuse contributing and retain the cost, which contributes to increase their relative individual payoff compared to the cooperators (investing in individual marketing skills). Agents are assembled in groups with size N. Success in the group is achieved if at least M agents cooperate towards the collective goal – a threshold that, in *For The Record*, corresponds to the minimum market value that an album must accomplish, in order to be successful. In case of success, each agent in the group receives a benefit b (*e.g.* sell a lot of albums, fame). In case of failure, each agent in the group has a penalty p (failure and mocking as a band and as individual musicians; or, as in the *For The Record*, risking that the game ends prematurely and all the players lose their accumulated profits). To capture the role of partner selection and, in particular, to intuit the reason for this selection to depend on a previous failure, we consider three types of cooperators:

- Unconditional Cooperator (**C**): Always cooperates and always plays with any agent;
- Strict Cooperator (**SC**): Always cooperates yet only plays with those perceived as cooperators.
- Outcome-based Cooperator (**OC**): Always cooperates; only plays with those known to be cooperators when was previously in an unsuccessful group; plays with any agent when was previously in an successful group.

While in the experiments with *For The Record* we considered an iterated game repeated over several rounds, in the simplified model (which we now study theoretically), interactions are assumed to be one-shot. However, agents are assumed to be able to uncover the strategy adopted by opponents in a group – in real scenarios, such phenomena may depend on the availability of public reputations or previous direct interactions. We abstain from addressing the role of repeated interactions, reputation or other strategy anticipation mechanisms in order to focus on the reasons for an agent to prefer a cooperative partner only when she looses a previous game, assuming that information about previous interactions is available. Following the experimental results obtained, we aim at exploring the potential advantages, from an evolutionary point of view, of using strategy **OC**, when compared with **SC**.

We shall first notice that, by using **OC**, an agent will either behave as a **SC** or as a **C**, depending on the probability of being in a previous unsuccessful interaction; if the collective goal was not achieved, as the experiments show, individuals significantly prefer to play with **C** partners, thus behaving as a **SC**. In our analysis we will study the 3-strategy dynamics, assuming that, at most, three different strategies can co-exist in the population. This is more likely to occur when the exploration rate of agents is low [58]. We start by formalizing the scenario 1) {**C**, **SC**, **D**}; then we show how the other two scenarios of interest, 2) {**OC**, **SC**, **D**} or 3) {**C**, **SC**, **D**}, can be mapped onto scenario 1).

4.1 3-strategy game fitness

*4.1.1 {**C**, **SC**, **D**}:* When there are k agents adopting strategy **SC**, l agents adopting **C** and $Z - k - l$ agents adopting **D**, the fitness (or average payoff) of an agent adopting **C**, resulting from plays in groups with size N, reads as,

$$f_1^C(k,l) = \sum_{i=1}^{N-1} \left(\frac{\binom{l-1}{i}\binom{Z-l-k}{N-1-i}}{\binom{Z-1}{N-1}} \Pi_C(i+1) \right) + \frac{\binom{l+k-1}{N-1}}{\binom{Z-1}{N-1}}(b-c), \quad (1)$$

where $\Pi_C(i) = \Theta(i-M)b - c - [1 - \Theta(i-M)]p$ is the payoff of **C** obtained in a group with i **C**s and $N - i$ **D**s and $\Theta(x)$ is the

Heaviside step function: $\Theta(x) = 1$ if $x \geq 0$ and $\Theta(x) = 0$ otherwise. Note that collective success requires at least M cooperators. The first term of the right hand side of Eq (1) represents the payoff earned in groups where only **C**s and **D**s take part; the second term adds the payoff in groups where **C**s and **SC**s take part, where the threshold M is always achieved. Also, note that $\binom{l}{i}\binom{Z-l-k}{N-i}/\binom{Z}{N}$ is the probability (hypergeometric) of sampling a group with size N with i Cooperators (**C**) and $N-i$ Defectors (**D**), from a population with l **C**s, k **SC**s and $Z-k-l$ **D**s. The fitness of agent **D** stands as,

$$f_1^D(k,l) = \sum_{i=0}^{N-1} \left(\frac{\binom{l}{i}\binom{Z-k-l-1}{N-1-i}}{\binom{Z-1}{N-1}} \Pi_D(i) \right). \tag{2}$$

where $\Pi_D(i) = \Theta(i-M)b - [1-\Theta(i+1-M)]p$ is the payoff of a defector in a group with i cooperators. The fitness of **SC** reads

$$f_1^{SC}(k,l) = \frac{\binom{k+l-1}{N-1}}{\binom{Z-1}{N-1}}(b-c). \tag{3}$$

as **SC**s always prefer **C** – refusing to play with **D** – and, so, the only groups they concede to play in are those composed by **SC**s and **C**s.

4.1.2 $\{\boldsymbol{OC}, \boldsymbol{SC}, \boldsymbol{D}\}$*:* Now we formalize the scenario in which strategies **OC**, **SC** and **D** can co-exist in a population. First, the probability that an agent **OC** looses a game (*i.e.*, takes part in a group where collective success is not achieved) is given by,

$$u_2(k,l) = \sum_{i=0}^{M-2} \left(\frac{\binom{l-1}{i}\binom{Z-k-l}{N-1-i}}{\binom{Z-1}{N-1}} \right), \tag{4}$$

that is, the probability that the game occurs (no **SC** and **D** simultaneously the group) and less than M individuals with strategy **OC** take part in the group. We can now realize that, with probability $u_2(k,l)$, an individual with strategy **OC** will play as **SC**; with probability $(1-u_2(k,l))$ an agent will play with strategy **C**. This said, we may use the fitness functions detailed in the previous section to describe the evolutionary dynamics in the present **OC**-**SC**-**D** scenario. If each **OC** individual becomes **SC** with probability $u_2(k,l)$, the probability that, out of l **OC** agents, l' become **SC** ($P(X = l')$) is given by the binomial distribution $P(X = l') = \binom{Z}{l'}(u_2)^{l'} + (1-u_2)^{l-l'}$. For the sake of simplicity, we will use the mean value of the distribution ($l' = u_2(k,l).l$) as the average number of **OC** agents that will play as **SC**. This way, the effective number of agents playing as **SC** will be given by $k' = k + l'$ and the effective number of agents playing as **C** comes down to $l - k'$. The fitness of agent X (with strategy **C**, **SC** or **D**) can conveniently be written as

$$f_2^X(k,l) = f_1^X(k', l-k'). \tag{5}$$

The fitness of an agent playing **OC** can be written as

$$f_2^{OC}(k,l) = u_2(k,l)f_2^{SC}(k,l) + (1-u_2(k,l))f_2^C(k,l). \tag{6}$$

4.1.3 $\{\boldsymbol{OC}, \boldsymbol{D}, \boldsymbol{C}\}$*:* Following the previous reasoning, in the 3rd scenario, the probability that an agent with strategy **OC** looses a game can be given by,

$$u_3(k,l) = \sum_{i=0}^{M-2} \left(\frac{\binom{l+k-1}{i}\binom{Z-k-l}{N-1-i}}{\binom{Z-1}{N-1}} \right). \tag{7}$$

In this case, using $l' = u_3(k,l).l$, the effective number of agents playing as **SC** will be given by l' and the effective number of agents playing as **C** is $k' = k + l - l'$. Thus, we have,

$$f_3^X(k,l) = f_1^X(l', k+l-l'). \tag{8}$$

The fitness of an agent playing **OC** can be written as

$$f_3^{OC}(k,l) = u_3(k,l)f_3^{SC}(k,l) + (1-u_3(k,l))f_3^C(k,l). \tag{9}$$

4.2 3-strategy game dynamics

The previous fitness functions convey the average payoff pertaining each strategy. With those quantities we are able to analyze the evolutionary dynamics of strategy adoption, assuming that, at each moment in time, the most successful strategies have a higher probability of being adopted through social learning (*e.g.*, imitation) [49]. In general, we assume that an agent with strategy X will imitate an agent with strategy Y with a probability given by the sigmoid function $p_{X,Y}$ [53] defined as,

$$p_{X,Y} = (1 + e^{\beta(f_X - f_Y)})^{-1}, \tag{10}$$

where β is the selection intensity, controlling how dependent is the imitation process on the fitness differences and often used to better fit experimental data with theoretical predictions [36, 61]. We use $\beta = 1$ in our analysis. The probability that one more agent adopts strategy X, from a configuration in which k agents adopt X, l adopt strategy Y and $Z-k-l$ adopt W is given by,

$$T_X^+(k,l) = (1-\mu)\frac{k}{Z}\left(\frac{l}{Z-1}p_{Y,X} + \frac{Z-k-l}{Z-1}p_{W,X}\right) + \mu\frac{Z-l}{2Z}, \tag{11}$$

where we add a mutation term μ. This setup assumes that with probability $(1-\mu)$ agents resort to social learning and with probability (μ) to exploration – *i.e.*, randomly adopting any strategy [42, 48, 52]. Likewise, the probability that one less agent adopts strategy X from a configuration in which k agents adopt X, l adopt Y and $Z-k-l$ adopt W is given by,

$$T_X^-(k,l) = (1-\mu)\frac{k}{Z}\left(\frac{l}{Z-1}p_{X,Y} + \frac{Z-k-l}{Z-1}p_{X,W}\right) + \mu\frac{l}{Z}. \tag{12}$$

We are now able to define a Markov Chain where each state corresponds to a particular combination of 3 strategies (or 2, see Fig. 2 below) and where transition probabilities between adjacent states are given by Eqs (11) and (12). As the corresponding Markov Chain is irreducible (whenever $\mu > 0$), its stationary distribution is unique and conveys the information about the long-term behavior of this chain (limiting and occupancy distribution) [21]. The stationary distribution represented in vector $\boldsymbol{\pi} = [\pi_k]$ thus translates the long-run fraction of the time the system spends in each state $s = (k,l)$ – where $k = s_k$ X agents and $l = s_l$ Y agents exist. This distribution is calculated as $\boldsymbol{\pi} = \boldsymbol{\pi}T$, where T is the transition matrix constructed resorting to Eqs (11) and (12) such that

$$\begin{aligned}
T_{(k,l)\to(k+1,l)} &= T_X^+(k,l) \\
T_{(k,l)\to(k-1,l)} &= T_X^-(k,l) \\
T_{(k,l)\to(k,l+1)} &= T_Y^+(k,l) \\
T_{(k,l)\to(k,l-1)} &= T_Y^-(k,l) \\
T_{(k,l)\to(k,l)} &= 1 - T_X^+(k,l) - T_X^- - T_Y^+ - T_Y^- \\
T_{(x,y)\to(w,z)} &= 0, \text{otherwise.}
\end{aligned} \tag{13}$$

A similar stationary distribution would be obtained through simulations, yet requiring intensive computational resources to obtain numerically precise results. In particular, our approach (also recently used in [41]) has the advantage of providing an expedite intuition on the origins of such distributions through the so-called gradients of selection, whose numerical calculation would also require extensive simulations covering all possible population states. The gradient of selection portrays, for each configuration, the most likely evolutionary path. These gradients of selection read as,

$$\Delta T(k,l) = (T_X^+(k,l) - T_X^-(k,l), T_Y^+(k,l) - T_Y^-(k,l)). \quad (14)$$

Using these tools, in panel a) of Fig. 2 and Fig. 3 we represent the gradient of selection (streamlines) whereas in panel b) of Fig. 2 and the background of the simplexes in Fig. 3 we represent the stationary distribution(s). Table 1 summarizes the notation used:

Symbol	Meaning
N	group size
b	initial endowment
c	contribution of cooperators
M	min number of cooperators for collective success
r	risk
μ	mutation / exploration probability
β	selection intensity
Z	population size
p	penalty incurred with collective failure

Table 1: List of mathematical symbols used.

5 THEORETICAL RESULTS

Here we show the results of studying the model previously introduced, with the goal of clarifying the advantages of strategy **OC** over **SC** (or **C**) in the long-run. Intuitively, a strategy **SC** – only preferring to play with **C** partners – would have all the ingredients to constitute a desirable behavior, from the individual point of view. In fact, by comparing the 2-strategy dynamics of strategies **C**, **OC** and **SC** against **D**, we can evince (Fig. 2) that **SC** is the strategy in a better position to invade and fixate in a population composed by the selfish agents **D**. The results in Fig. 2 portray the gradient of selection (panel a) and the stationary distribution (panel b) when considering that only two strategies are present in the population. This analysis nicely charaterizes the competition between cooperators (**C**, **SC** and **OC**) and unconditional Defectors (**D**s), missing however the potentially important interplay among cooperative strategies. The 2-strategy dynamics can be obtained by resorting to the 3-strategy models presented previously. Namely, i) the dynamics of strategy **C** against **D** was obtained from scenario 1 (section 4.1.1) considering $k = 0$ (**SC** absent from the population), ii) the dynamics of strategy **SC** against **D** was obtained from scenario 1 considering $l = 0$ (**C** absent from the population) and finally, iii) the dynamics of strategy **OC** against **D** was obtained from scenario 2 (section 4.1.2) considering $k = 0$ (**SC** absent from the population). In Fig. 2 we show that **SC** is the strategy allowing the higher prevalence of cooperators (for the scenario M=4, N=7, b=10, c=2, p=2). This occurs as **SC** prevents the exploitation from **D** agents, by refusing to take part in groups with defectors. This way,

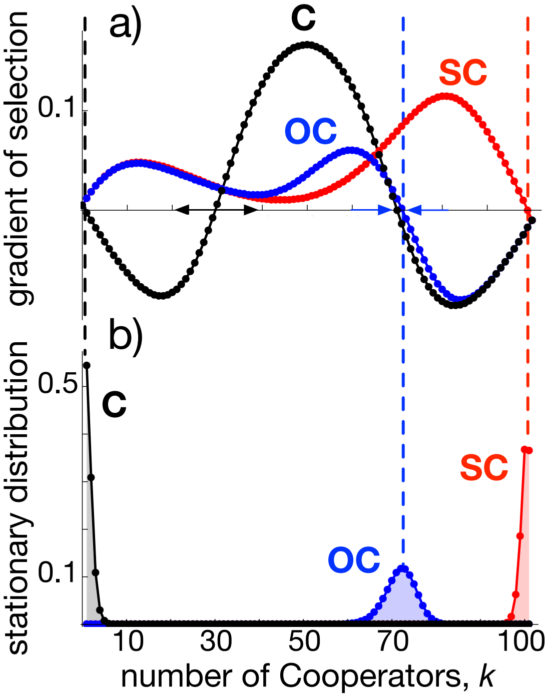

Figure 2: 2-strategy dynamics of OC, C and SC against D. In a) we represent the gradient of selection (the more plausible evolutionary path; when above the horizontal xx axis, it is more likely that cooperators – C, OC or SC – spread; below the horizontal axis, it is more likely that D spreads; this way, arrows on top of the xx axis represent the most likely direction of evolution). In b) we represent the stationary distribution, *i.e.*, the long-run fraction of the time the system spends in each state. We can observe that strategy C (black curves) is unable to invade a population of D, for this combination of parameters; the strategy SC (red curves) invades a population of Ds and the system ends up in a state where most of the population adopts SC (portrayed by the red distribution skewed to the right, in panel b), which is supported by the positive gradient of selection in panel a). Finally, strategy OC (blue curve) is able to invade the population of Ds and stabilze a configuration in which OCs and Ds co-exist. Parameters used: μ=0.01, M=4, N=7, b=10, c=2, Z=100.

defectors are unable to achieve the benefits of collective success in any possible group. The Unconditional Cooperators (**C**) obtain less payoff than defectors when taking part in successful groups in which a defector also has the benefit of collective success, yet without contributing to that endeavor. **OC** constitutes a middle point between the two strategies: whenever few cooperators exist, **OC** is unable to take part in successful groups and thus behaves as **SC**. When success is easier to be achieved – given the increased number of cooperators – **OC** is willing to play with **D** partners, thus recovering from the strictness of **SC** that condemns this strategy

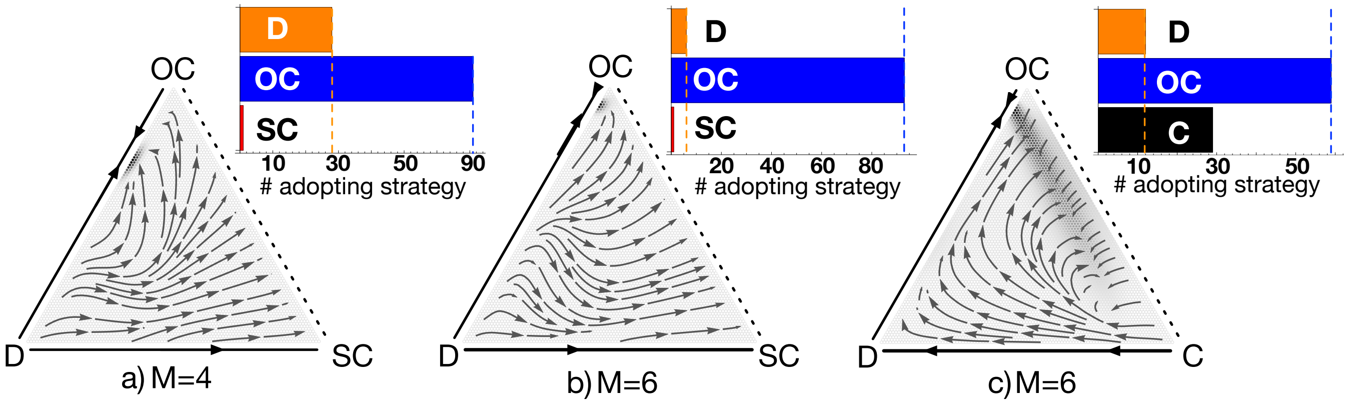

Figure 3: 3-strategy dynamics between D, C, SC, OC strategies. In the top panels we portray the gradient of selection (streamlines pointing the most likely direction of evolution, starting in each possible state) and stationary distribution (background grayscale; the darker, the more time is spent in that state). The vertices of the simplexes represent configurations in which only one strategy exists in the population (label close to the corresponding vertex). The edges correspond to configurations in which two strategies co-exist and the interior of the simplexes comprises the configurations where 3 strategies co-exit. The information regarding the stationary distribution is summarized in the bottom panels, where we represent the average usage of all strategies (*i.e.*, the frequency of strategies in each population configuration weighted by the probability of being on that state). a) When D, SC and OC co-exist, most of time is spent in states where OC is highly prevalent; b) this is even more evident if we consider high M. c) when D, C and OC co-exist, a lot of time is spent in states with high prevalence of C and OC, yet with an higher fraction of individuals using OC. Parameters: μ=0.01, N=7, M=4, b=10, c=2 (panels a and b), c=5 (panel c), p=2, Z=100.

to a very low fitness (and gradient $T^+_{SC} - T^-_{SC}$ close to 0) when the population is composed by half of cooperators and half of defectors.

The point is now to know how does **SC** behave when a third strategy (**OC**) is introduced in the **SC**-**D** dynamics. The effect of considering an **OC**-**SC**-**D** dynamics can be apprehended in Fig. 3a and b. We can realize that, by introducing strategy **OC** in a population of **SC**s and **D**s, most of the time will be spent in states with a high prevalence of **OC**. In fact, **OC**s are able to constitute a stable strategy that concedes the existence of a small fraction of **D**, while reaping the benefits of playing in groups that achieve collective success (even if they have a very small number of **D** partners). The streamlines in Fig. 3a and b show that **SC** dominates **D** (vectors in the bottom edge of the simplexes) and there is a co-existence between **D** and **OC** (left edges of the simplexes). This was precisely the conclusion in Fig. 2. However, in the interior of the simplexes (when 3 strategies co-exist) the gradients point, in large fraction of configurations, upwards, which indicates a higher probability that **OC** successively replaces **D** and **SC**. Interestingly, as observed in Fig. 3c, **OC** can also be advantageous relative to **C** when c is high. In this case, we observe a cyclical dynamics: **OC**s are needed to initially punish **D**s and open space for the evolution of **C**s; when strategy **D** vanishes, **C** becomes advantageous compared with **OC**, as the adopters of this strategy manage to take part in more successful groups than **OC**s – which, with some probability, still refuse to play in groups with **D**s. With the increased number of successful groups, **OC**s will increasingly play as **C**, making these two strategies almost neutral, *i.e.*, receiving a very close fitness. Whenever **C**s replace **OC**s, the barriers for the subsequent invasion of **D**s are alleviated. This way, the fraction of **D** agents increases, which, again, evidences the advantages of **OC** over **C** and opens space for the re-invasion of **OC** players. Finally, in Fig. 4, we observe that the advantages of **OC** over **SC** are augmented (or exist) for low c. Contrarily, **OC** tends to be more prevalent than **C** when c is high. In general, however, we verify that **OC** profits from high M (Fig. 5).

Recovering the experiments performed, we may note that, in *For The Record*, the expected payoff of a defector is 17.5 (expected value of rolling 5 dice with 6 faces). A player that always cooperates receives a payoff of 3. Assuming that individuals will always cooperate or defect, we need at least 2 cooperators (out of 3 players) in the group ($M = 2$), for the expected value of the album to surpass the expected value of the threshold in the last round. The cost of cooperating is expected to be $c = 14.5$ ($b = 17.5$; $b - c = 3$ and thereby $17.5 - c = 3$). This way, we tested high expected values of c and M, relatively to b and N: $c/b = 0.83$ and $M/N = 0.66$. Future experiments shall test different expected values of M and c.

6 CONCLUSION AND DISCUSSION

Here we explore partner selection in Collective Risk Dilemmas (**CRD**). In the context of Prisoner's Dilemmas [11] or Public Goods Games [16, 26], previous studies found that introducing strategies that refuse playing with defectors opens space for cooperative strategies to invade the previously stable defective equilibria. In **CRD**, a new component is introduced: group success or failure in achieving the collective goals. It is thereby unclear which strategies are more efficient in promoting cooperation, given that they can be conditioned on 1) the strategies of opponents in the group or 2) previous success or failure experience. Here we resort to a human-robot experiment, which allows controling the robot behavior and explicitly test a cooperative and defective artificial partner. After the game, we ask the human subjects whether they would prefer the

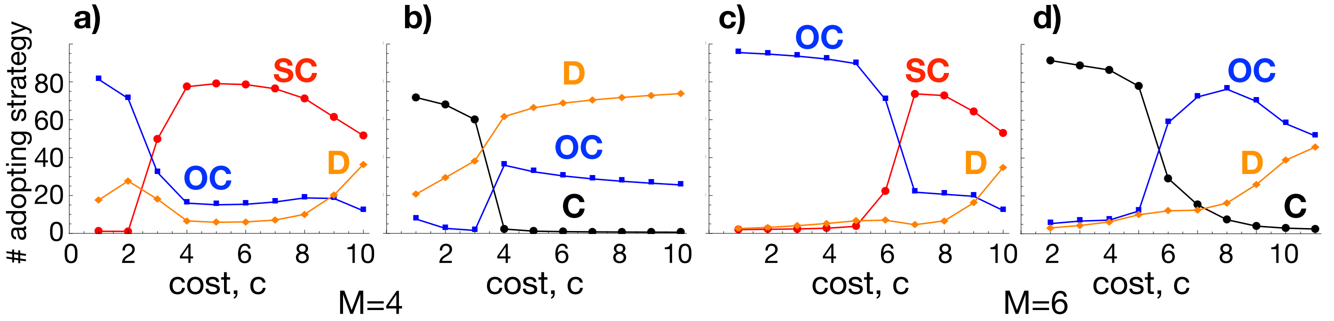

Figure 4: The advantages of OC over SC are more evident for low *c*, that is, whenever cooperation requires paying less costs. Contrarily, the advantages of OC over C are more evident for high *c*. Other parameters: *N*=7, *Z*=100, *p*=2, *b*=10

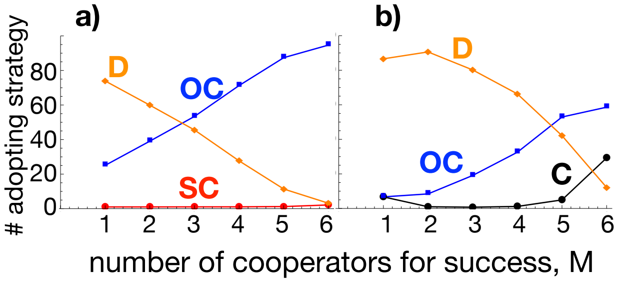

Figure 5: The advantages of OC are more evident for high *M*, that is, the situations in which is the collective goal requires more cooperators. Other parameters: *N*=7, *Z*=100, *p*=2, *b*=10, $c = 2$ (left panel, a) and $c = 5$ (right panel, b)

defective or cooperative partner to play with, in the future. Humans select the cooperative partner significantly more often when they previously took part in a group that failed to achieve the collective goals. Next, resorting to an evolutionary game theoretical model, we test a strategy (that we called **OC**, Outcome-based Cooperator, cooperating and only accepting to play with defectors when group success was achieved previously) in comparison with the unconditional Cooperator strategy (**C**), the unconditional Defector strategy (**D**) and the Strict Cooperator strategy (**SC**) – that cooperates but only accepts playing with other cooperators, regardless previous game outcomes. We find that **OC** can be more prevalent than **C** and **SC**, preventing the invasion of defectors and, at the same time, conceding to play in group configurations that, despite having a few defectors, can nonetheless manage to achieve group success. In summary, answering to the initial posed questions, outcome-based cooperation in **CRD** seems to be both efficient in promoting cooperation and likely to be used by human subjects.

The theoretical model proposed allows studying three co-existing strategies in the population. We focus on studying **OC** in comparison with **C** and **D** (the traditional strategies studied in the context of **CRD** [38]) and **SC** (the strategy only accepting to play with cooperators that, intuitively, should have had the highest prevalence). Notwithstanding, even keeping binary actions (**C** and **D**), strategies can become increasingly complex by discriminating based on the number of cooperators in the group [33, 56], or by stressing all combinations of strategy avoidance and action played. One could think about a strategy that only accepts playing with cooperators and yet decides to defect – a malicious version of **OC**. In the context of Public Goods Games it was found that introducing extra punitive strategies – as anti-social punishment – may prevent cooperation [34]. Thereby, our future plans include extending the current theoretical framework to access the robustness of cooperation in **CRD**, when the full repertoire of strategies is considered.

We shall underline that, in the present work, we are mainly concerned with analyzing the advantages of an outcome-based strategy like **OC** against strategies **C**, **D** or **SC**. We do this comparison assuming that both discriminatory strategies (**SC** and **OC**) have access to the same level of information. This way, we assume, as a baseline, that all agents are able to anticipate accurately the action used by an opponent, using this information to decide taking part – or not – in a group. Future approaches may combine **CRD** with models of reputation that allow anticipating the strategies of opponents [40], commitments [16, 17], or even consider more complex agent architectures that are capable of anticipating [7].

The theoretical model proposed can be, in the future, extended to study outcome-based strategies in other multiplayer games, particularly those with non-linear payoffs such as Multiplayer Ultimatum Games [41, 44], Multiplayer Trust Games [3] or N-Person Stag-HuntGames [30]. Also, the conclusions that we derive now, mainly the evidence that **OC** becomes more prevalent than **SC** when the cost of cooperating (*c*) is low and the group success threshold (*M*) is high, can inform new experiments with human subjects, thus opening new avenues for a symbiosis between theoretical and experimental analysis in collective action problem.

ACKNOWLEDGEMENTS

Supported by FCT Portugal through grants SFRH/BD/118031/2016, PTDC/EEI-SII/5081/2014, PTDC/MAT/STA/3358/2014, and UID/CEC/50021/2013. F.P.S. acknowledges support from the James S. McDonnell Foundation.

REFERENCES

[1] Daan Bloembergen, Karl Tuyls, Daniel Hennes, and Michael Kaisers. 2015. Evolutionary dynamics of multi-agent learning: a survey. *Journal of Artificial Intelligence Research* 53 (2015), 659–697.

[2] Xiaojie Chen, Tatsuya Sasaki, Åke Brännström, and Ulf Dieckmann. 2015. First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation. *Journal of the Royal Society Interface* 12, 102 (2015), 20140935.

[3] Manuel Chica, Raymond Chiong, Michael Kirley, and Hisao Ishibuchi. 2017. A Networked N-player Trust Game and its Evolutionary Dynamics. *IEEE Transactions on Evolutionary Computation* (2017).

[4] Filipa Correia, Samuel Mascarenhas, Samuel Gomes, Patricia Arriaga, Iolanda Leite, Rui Prada, Francisco S. Melo, and Ana Paiva. 2019. Exploring Prosociality in Human-Robot Teams. In *Proceedings of the 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI'19)*. IEEE Press.

[5] Chad Crawford, Rachna Nanda Kumar, and Sandip Sen. 2018. Resisting Exploitation Through Rewiring in Social Networks: Social Welfare Increase using Parity, Sympathy and Reciprocity. In *Proceedings of AAMAS'18*. IFAAMAS, 1915–1917.

[6] Soham De, Dana S Nau, and Michele J Gelfand. 2017. Understanding norm change: An evolutionary game-theoretic approach. In *Proceedings of AAMAS'17*. IFAAMAS, 1433–1441.

[7] Elias Fernández Domingos, Juan-Carlos Burguillo, and Tom Lenaerts. 2017. Reactive Versus Anticipative Decision Making in a Novel Gift-Giving Game.. In *Proceedings of AAAI'17*, Vol. 17. AAAI Press, 4399–4405.

[8] Katrin Fehl, Daniel J van der Post, and Dirk Semmann. 2011. Co-evolution of behaviour and social network structure promotes human cooperation. *Ecology letters* 14, 6 (2011), 546–551.

[9] Ernst Fehr and Simon Gächter. 2002. Altruistic punishment in humans. *Nature* 415, 6868 (2002), 137.

[10] Michael R Genesereth, Matthew L Ginsberg, and Jeffrey S Rosenschein. 1986. Cooperation without communication. In *Proceedings of AAAI'86*. Elsevier, 51–57.

[11] Whan Ghang and Martin A Nowak. 2015. Indirect reciprocity with optional interactions. *Journal of Theoretical Biology* 365 (2015), 1–11.

[12] Herbert Gintis, Samuel Bowles, Robert T Boyd, Ernst Fehr, et al. 2005. *Moral sentiments and material interests: The foundations of cooperation in economic life*. Vol. 6. MIT press.

[13] Nathan Griffiths. 2008. Tags and image scoring for robust cooperation. In *Proceedings of AAMAS'08*. IFAAMAS, 575–582.

[14] Nathan Griffiths and Michael Luck. 2010. Changing neighbours: improving tag-based cooperation. In *Proceedings of AAMAS'10*. IFAAMAS, 249–256.

[15] Victoria Groom and Clifford Nass. 2007. Can robots be teammates?: Benchmarks in human–robot teams. *Interaction Studies* 8, 3 (2007), 483–500.

[16] The Anh Han, Luís Moniz Pereira, and Tom Lenaerts. 2017. Evolution of commitment and level of participation in public goods games. *Autonomous Agents and Multi-Agent Systems* 31, 3 (2017), 561–583.

[17] The Anh Han, Luís Moniz Pereira, and Francisco C. Santos. 2012. The Emergence of Commitments and Cooperation. In *Proceedings of AAMAS'13 (AAMAS '12)*. IFAAMAS, Richland, SC, 559–566.

[18] Christoph Hauert, Silvia De Monte, Josef Hofbauer, and Karl Sigmund. 2002. Volunteering as red queen mechanism for cooperation in public goods games. *Science* 296, 5570 (2002), 1129–1132.

[19] Tad Hogg. 1995. Social dilemmas in computational ecosystems. In *Proceedings of IJCAI'95*. 711–718.

[20] Nicholas R Jennings, Katia Sycara, and Michael Wooldridge. 1998. A roadmap of agent research and development. *Autonomous Agents and Multi-agent Systems* 1, 1 (1998), 7–38.

[21] Vidyadhar G Kulkarni. 2016. *Modeling and analysis of stochastic systems*. Chapman and Hall/CRC.

[22] Maja J Matarić. 1995. Issues and approaches in the design of collective autonomous agents. *Robotics and Autonomous Systems* 16, 2-4 (1995), 321–331.

[23] Manfred Milinski, Ralf D Sommerfeld, Hans-Jürgen Krambeck, Floyd A Reed, and Jochem Marotzke. 2008. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences* 105, 7 (2008), 2291–2294.

[24] Yishay Mor and Jeffrey S Rosenschein. 1995. Time and the Prisoner's Dilemma.. In *Proceedings of ICMAS'95*. 276–282.

[25] Javier Morales, Michael Wooldridge, Juan A Rodríguez-Aguilar, and Maite López-Sánchez. 2018. Off-line synthesis of evolutionarily stable normative systems. *Autonomous Agents and Multi-Agent Systems* (2018), 1–37.

[26] Joao A Moreira, Jorge M Pacheco, and Francisco C Santos. 2013. Evolution of collective action in adaptive social structures. *Scientific Reports* 3 (2013), 1521.

[27] Martin A Nowak. 2006. *Evolutionary dynamics*. Harvard University Press.

[28] Martin A Nowak, Karen M Page, and Karl Sigmund. 2000. Fairness versus reason in the ultimatum game. *Science* 289, 5485 (2000), 1773–1775.

[29] Jason Olejarz, Whan Ghang, and Martin A Nowak. 2015. Indirect reciprocity with optional interactions and private information. *Games* 6, 4 (2015), 438–457.

[30] Jorge M Pacheco, Francisco C Santos, Max O Souza, and Brian Skyrms. 2009. Evolutionary dynamics of collective action in N-person stag hunt dilemmas. *Proceedings of the Royal Society of London B* 276, 1655 (2009), 315–321.

[31] Ana Paiva, Fernando P Santos, and Francisco C Santos. 2018. Engineering Pro-Sociality with Autonomous Agents. *AAAI'18* (2018), 7994–7999.

[32] Ana Peleteiro, Juan C Burguillo, and Siang Yew Chong. 2014. Exploring indirect reciprocity in complex networks using coalitions and rewiring. In *Proceedings of AAMAS'14*. IFAAMAS, 669–676.

[33] Flavio L Pinheiro, Vitor V Vasconcelos, Francisco C Santos, and Jorge M Pacheco. 2014. Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Computational Biology* 10, 11 (2014), e1003945.

[34] David G Rand and Martin A Nowak. 2011. The evolution of antisocial punishment in optional public goods games. *Nature Communications* 2 (2011), 434.

[35] David G Rand and Martin A Nowak. 2013. Human cooperation. *Trends in Cognitive Sciences* 17, 8 (2013), 413–425.

[36] David G Rand, Corina E Tarnita, Hisashi Ohtsuki, and Martin A Nowak. 2013. Evolution of fairness in the one-shot anonymous Ultimatum Game. *Proceedings of the National Academy of Sciences* 110, 7 (2013), 2581–2586.

[37] Eduardo Benítez Sandoval, Jürgen Brandstetter, Mohammad Obaid, and Christoph Bartneck. 2016. Reciprocity in human-robot interaction: a quantitative approach through the prisoner's dilemma and the ultimatum game. *International Journal of Social Robotics* 8, 2 (2016), 303–317.

[38] Francisco C Santos and Jorge M Pacheco. 2011. Risk of collective failure provides an escape from the tragedy of the commons. *Proceedings of the National Academy of Sciences* 108, 26 (2011), 10421–10425.

[39] Francisco C Santos, Jorge M Pacheco, and Tom Lenaerts. 2006. Cooperation prevails when individuals adjust their social ties. *PLoS Computational Biology* 2, 10 (2006), e140.

[40] Fernando P Santos. 2017. Social Norms of Cooperation in Multiagent Systems. In *Proceedings of AAMAS'17*. IFAAMAS, 1859–1860.

[41] Fernando P Santos, Jorge M Pacheco, Ana Paiva, and Francisco C Santos. 2019. Evolution of collective fairness in hybrid populations of humans and agents. In *Proceedings of AAAI'19*. AAAI Press.

[42] 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.

[43] Fernando P Santos, Jorge M Pacheco, and Francisco C Santos. 2018. Social Norms of Cooperation with Costly Reputation Building. In *Proceedings of AAAI'18*. AAAI Press, 4727–4734.

[44] Fernando P Santos, Francisco C Santos, Francisco S Melo, Ana Paiva, and Jorge M Pacheco. 2016. Dynamics of fairness in groups of autonomous learning agents. In *International Conference on Autonomous Agents and Multiagent Systems (workshops' best papers book)*. Springer, 107–126.

[45] 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), e1004709.

[46] 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 (2018), 242.

[47] Sandip Sen et al. 2018. The Effects of Past Experience on Trust in Repeated Human-Agent Teamwork. In *Proceedings of AAMAS'18*. IFAAMAS, 514–522.

[48] Pedro Sequeira, Francisco S Melo, and Ana Paiva. 2011. Emotion-based intrinsic motivation for reinforcement learning agents. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, 326–336.

[49] Karl Sigmund. 2010. *The calculus of selfishness*. Princeton University Press.

[50] J Maynard Smith and George R Price. 1973. The logic of animal conflict. *Nature* 246, 5427 (1973), 15.

[51] Alessandro Tavoni, Astrid Dannenberg, Giorgos Kallis, and Andreas Löschel. 2011. Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences* 108, 29 (2011), 11825–11829.

[52] Arne Traulsen, Christoph Hauert, Hannelore De Silva, Martin A Nowak, and Karl Sigmund. 2009. Exploration dynamics in evolutionary games. *Proceedings of the National Academy of Sciences* 106, 3 (2009), 709–712.

[53] Arne Traulsen, Martin A Nowak, and Jorge M Pacheco. 2006. Stochastic dynamics of invasion and fixation. *Physical Review E* 74, 1 (2006), 011909.

[54] Karl Tuyls and Ann Nowé. 2005. Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review* 20, 1 (2005), 63–90.

[55] Karl Tuyls and Simon Parsons. 2007. What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence* 171, 7 (2007), 406–416.

[56] Sven Van Segbroeck, Jorge M Pacheco, Tom Lenaerts, and Francisco C Santos. 2012. Emergence of fairness in repeated group interactions. *Physical Review Letters* 108, 15 (2012), 158104.

[57] Vitor V Vasconcelos, Francisco C Santos, and Jorge M Pacheco. 2013. A bottom-up institutional approach to cooperative governance of risky commons. *Nature Climate Change* 3, 9 (2013), 797.

[58] Vítor V Vasconcelos, Fernando P Santos, Francisco C Santos, and Jorge M Pacheco. 2017. Stochastic dynamics through hierarchically embedded Markov chains. *Physical Review Letters* 118, 5 (2017), 058301.

[59] Jörgen W Weibull. 1997. *Evolutionary game theory*. MIT press.

[60] Gerhard Weiss. 1999. *Multiagent systems: a modern approach to distributed artificial intelligence*. MIT press.

[61] Ioannis Zisis, Sibilla Di Guida, TA Han, Georg Kirchsteiger, and Tom Lenaerts. 2015. Generosity motivated by acceptance-evolutionary analysis of an anticipation game. *Scientific Reports* 5 (2015), 18076.