



Picky losers and carefree winners prevail in collective risk dilemmas with partner selection

Fernando P. Santos¹ · Samuel Mascarenhas² · Francisco C. Santos² · Filipa Correia² · Samuel Gomes² · Ana Paiva²

Published online: 25 May 2020
© The Author(s) 2020

Abstract

Understanding how to design agents that sustain cooperation in multi-agent systems has been a long-lasting goal in distributed artificial intelligence. Proposed solutions rely on identifying free-riders and avoiding 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. The collective risk dilemma (CRD) is one of these games, and it is unclear whether the known mechanisms of cooperation remain effective in this case. Here we study the emergence of cooperation in CRD through partner-based selection. First, we discuss an experiment in which groups of humans and robots play a CRD. This experiment suggests that people only prefer cooperative partners when they lose a previous game (i.e., when collective success was not previously achieved). Secondly, we develop an evolutionary game theoretical model pointing out the evolutionary advantages of preferring cooperative partners only when a previous game was lost. We show that this strategy constitutes a favorable 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. We confirm these theoretical results through computer simulations considering a more complex strategy space. Third, resorting to online human-agent experiments, we observe that participants are more likely to accept playing in a group with one defector when they won in a previous CRD, when compared to participants that lost the game. These empirical results provide additional support to the human predisposition to use outcome-based partner selection strategies in human-agent interactions.

Keywords Cooperation · Collective risk dilemma · Public goods game · Game theory · Partner selection · Human-robot interaction · Complex systems

✉ Fernando P. Santos
fpsantos@princeton.edu

Extended author information available on the last page of the article

1 Introduction

Cooperation between self-interested agents has been a fundamental research topic in economics [20] and evolutionary biology [41]. Likewise, designing agents that sustain cooperation is a long-standing goal in multi-agent systems (MAS) [18, 32, 81]. Often agents take part in interactions that pose 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 [51, 65]; 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 [46, 57].

In several cooperation dilemmas, collective benefits are only distributed—or collective losses avoided—whenever a minimal number of agents cooperate [34, 44, 48, 66]. In natural settings, examples of this situation abound [44, 69], such as in the case of collective hunting in different species [5, 70]. Similar non-linear profiles have been found in whale hunting by humans [2] and in international relations [33]. This situation may also occur in voluntary vaccination [79]—where a certain fraction of individuals need to vaccinate for a population to achieve herd immunity—or simple daily tasks whose completion requires the effort of a minimum number of group members—such as college or company team projects. 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 benefits everyone. This situation inspired the so-called collective risk dilemma (CRD) [6, 37, 53, 71, 77]. 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, all players keep the money that was not invested. Otherwise, everyone loses the saved endowment with a given probability (therein lies the risk).

It is possible to apprehend the core of the dilemma at stake resorting to a simplified game. Following the model formalized in [53], 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 $0 < c < b$) 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 reached, everyone loses the saved endowment with a probability r (the risk). Assuming the most extreme scenario ($r = 1$) we can verify that, if all members cooperate, everyone earns $b - c$. If all members defect—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 optimum occurs when the threshold M is achieved without requiring one's contribution. In this situation, a defector earns all of the endowment b , while a cooperator just keeps $b - c$.

A possible way of incentivizing cooperation in CRD is to punish defectors or reward cooperators [7, 16, 21]. 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 gives rise to the so-called *second-order free riding* problems. A subtler way of eliciting cooperation in CRD—if opportunity is given—is to avoid agents known to have defected previously. In fact, mechanisms of such kind were previously

applied in the context of pairwise interactions, chiefly the prisoner's dilemma and the ultimatum game [14, 19, 42, 54]—or multiplayer interactions with linear payoffs [24, 28, 40]—that is, payoffs that increase linearly with the number of contributors, without the abrupt changes that occur when a threshold of contributions is attained, as in the CRD. It remains unclear, however,

- What are the individual advantages of selecting partners given own previous outcomes 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, studying the role of partner selection in CRD through an evolutionary game theoretical (EGT) model [80] and through human–agent experiments. We develop a 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 the CRD. Individuals that use this strategy cooperate, yet only accept to play with defectors when they previously achieved collective success. We compare **OC** 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 Cooperation, **SC**). We find that **OC** is the most prevailing strategy in a wide parameter region. In particular, **OC** wins against **SC** when 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 unconditional **C**s. This allows agents using **OC** to concede playing with **D** opponents when their representativeness in the population is low enough to still reach the collective threshold (M) in several interaction groups. Through simulations (Sect. 3.3), we confirm that **OC** is a highly prevalent strategy, even if agents are able to adopt additional conditional and defective strategies.

We complement this theoretical analysis with human–agent online experiments. These experiments allow us to confirm that, indeed, the outcome of a previous game affects the tolerance of playing with defectors. People are more likely to accept playing with defective agents if they won, previously, an instance of the CRD. In those experiments, participants play a CRD and are subsequently asked about their likelihood of joining a future team composed with some defectors. We observe that participants that lost the game are more likely to join a team with no defectors (compared with winners) and participants that won the game are more likely to join a team with one defector (compared with losers). This reveals that the outcome of a previous game indeed affects the tolerance of playing with defectors.

Our analysis is in part motivated by recent results revealing that humans accept interacting with defective partners when they previously faced collective success [9]. Those results were obtained after we conducted an experimental study with humans and robots. Using robots allowed us to fine-tune the strategies used and thus test explicitly a cooperative and defective partner. For those experiments, we frame CRD in the form of a band selection game, named *For The Record*, where agents are recruited to form a band and record albums. Cooperation means investing in mastering an instrument that contributes to 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. While in [9] we mainly discuss how humans rate robots' social attributes and the impact of game outcomes on group subjective measures, here we discuss the obtained results in terms of outcome-based partner choice in CRDs. For the

sake of completeness, we describe the human–robot experiments performed and the results obtained in Sect. 5.2. As we use the same game (*For the Record* [9, 10]) in the new online experiments, we fully describe this game in Sect. 5.1.

With these experiments and model, we open a new route to study strategies that efficiently incentivize cooperation in CRDs through partner selection conditioned on own success experiences. Next, in Sect. 2, we discuss several related approaches to elicit cooperation through defector avoidance, mainly in the context of pairwise interactions or multiplayer games with linear payoffs. In Sect. 3 we detail the theoretical model used to shed light on the role of outcome-based conditional strategies on CRDs. The analytical results are presented in Sect. 4, together with the result of agent-based simulations that allow confirming that outcome-based strategies have a high prevalence in CRDs even considering enlarged strategy spaces. The theoretical results allow understanding the advantages of outcome-based strategies in a long-term evolutionary context. We confirm, through human–agent online experiments, that humans are more likely to accept playing with defective agents if they previously had success playing a CRD game. In Sect. 5 we present that experimental setup, the game used, and discuss the obtained results. We end with Sect. 6, where we summarize our findings, point the limitations of our theoretical approach, and provide avenues for future work.¹

2 Related work

We focus on a social dilemma of cooperation, previously named collective risk dilemma (CRD), already alluded to above. This game was originally proposed in [37] 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) [53]. 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 (a conclusion also present in [48], where smaller groups lead to stable states with higher levels of cooperation). Here we follow the specification and notation in [53].

In the core of the 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. 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 Prisoner's Dilemma, Mor and Rosenschein found that allowing agents to opt out from a repeated interaction favors cooperative strategies [38]. In that work, individuals interact repeatedly with the same opponents. Similar mechanisms based on direct reciprocation have been also shown to elicit cooperation in N-person dilemmas [49, 76].

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,

¹ This manuscript constitutes an extended version of a work presented at AAMAS'19 [56].

provided that the average number of rounds per agents is high enough [19]. 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 [43]. Using reputations to adapt behaviors and punish unreasonably defective opponents is a principle that underlies indirect reciprocity [1]. In this context, Griffiths [22] showed that using reputations to discriminate and refusing to cooperate with defectors allows for cooperation to emerge. An alternative mechanism to avoid interacting with defectors is social network rewiring. Santos et al. [54] found that this mechanism allows for cooperation to emerge, a result also found in laboratory experiments [15] and public goods games [40]. Peleteiro et al. also used network rewiring, in combination with coalition formation, to promote the evolution of cooperative behavior [47]. Griffiths and Luck studied network rewiring as a way of improving tag-based cooperation [23]. Crawford 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 [12]. Following the same principle of avoiding interactions with defective opponents, Fernandez et al. studied anticipating mechanisms in the the context of Anticipation Games [14], an interaction paradigm proposed in [83]. In this case, agents refuse to play with agents if they were previously unfair. More recently, partner selection was pointed out as a cooperation mechanism also in the context of agents learning through reinforcement learning [3].

Most of previous works adopt strategies of defector avoidance in the context of 2-person games. In the context of multiplayer interactions, Hauert et al. [28] found that simply introducing the opportunity for agents to opt out from 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. 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 to commit to cooperate [24]. Commitments can also be seen as an elaborate form of pre-play signaling or quorum at the group level, also known to efficiently promote cooperation [45, 67]. In all cases, agents are allowed to opt out from the 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 may 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 the CRD theoretically (evolutionary game theory, EGT) were originally applied in the context of ecology and evolutionary biology [68]. 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 [4, 31, 36, 74, 75]. EGT was also recently applied to study social norms along different directions, namely the stability of normative systems [39], the emergence and time evolution of social norms [13], or the evolution of cooperation through norms and reputations [55, 59, 61, 82]. 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 [62]. In the context of human–robot interactions and cooperation dilemmas, it was also shown—as mentioned already—that cooperative partners are more often preferred whenever human players fail to previously achieve collective success; when winning, humans select the cooperative or defective opponents almost alike [9]. These experiments are fully detailed in Sect. 5 and the obtained results inspired us to develop a new evolutionary game theoretical model to interpret the advantages of selecting cooperative partners only when a previous game was lost.

3 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. We focus our attention on interactions following the previously introduced collective risk dilemma (CRD) [37, 53, 71]. Two baseline strategies are possible in this multi-player game: Cooperate and Defect. The Cooperators (**C**) pay a cost (c) in order to contribute to a collective endeavor (album quality, in the scenario of *For The Record* introduced above). 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 for an album 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 a successful group.

Using such a small strategy space allows us to build an analytic model which encompasses the main features of this dilemma. Nonetheless, in Sect. 3.3 we enlarge this strategy space, testing the validity of the analytic conclusions in more complex scenarios. Also, in the analytic model we assume that agents are able to uncover the strategy adopted by opponents in a group. In real scenarios, such possibility may depend on public reputations—whose availability may itself constitute a dilemma [59]—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 agents to prefer a cooperative partner only when they lose a previous game, assuming that information about previous interactions is available. Again, in Sect. 3.3 we explicitly consider, through simulations, agents that may fail to identify Defectors

in future interaction groups. In this section, and following the experimental results obtained [9], we develop simplified theoretical models to explore 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 an **SC** or as a **C**, depending on the probability of previously facing an unsuccessful interaction; if the collective goal was not achieved, as the experiments show, individuals significantly prefer to play with **C** partners, thus behaving as an **SC**. In the following analysis we will study 3-strategy dynamics, assuming that, at most, three different strategies can co-exist in the population. 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) **{OC, C, D}**, can be mapped onto scenario 1). Note that a scenario **{OC, C, SC}** is not interesting as, in this case, all agents play and cooperate, meaning that all strategies lead to the same payoff.

3.1 3-strategy game fitness

3.1.1 Scenario 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 playing in groups with size N , reads as

$$f_1^C(k, l) = \sum_{i=0}^{N-2} \frac{\binom{l-1}{i} \binom{Z-l-k}{N-1-i}}{\binom{Z-1}{N-1}} \Pi_C(i+1) + \frac{\binom{l+k-1}{N-1}}{\binom{Z-1}{N-1}} (b - c), \quad (1)$$

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 (at least one **D**, thus the summation limit $N - 2$) 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 $\binom{l}{i} \binom{Z-l-k}{N-1-i} / \binom{Z}{N}$ 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

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

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

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

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.

3.1.2 Scenario 2: {OC, SC, D}

Now we formalize the scenario in which l **OC**, k **SC** and $(Z - l - k)$ **D** strategies exist in a population. First, the probability that an agent **OC** loses a game (i.e., takes part in a group where collective success is not achieved) is given by

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

that is, the probability that the game occurs (no **SC** and **D** simultaneously in 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 $P(X = l') = \binom{l}{l'} (u_2)^{l'} (1 - u_2)^{l-l'}$. For the sake of simplicity, we will use the mean value of the distribution ($l' = u_2(k, l) \cdot l$) 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 $k + l'$ and the effective number of agents playing as **C** comes down to $l - l'$. The fitness of agent X (with strategy **C**, **SC** or **D**) can conveniently be written as

$$f_2^X(k, l) = f_1^X(k + l', l - l'). \quad (5)$$

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

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

3.1.3 Scenario 3: {OC, C, D}

Following the previous reasoning, in the **OC–C–D** scenario—where l adopt **OC**, k adopt **C** and $(Z - k - l)$ adopt **D**—the probability that an agent with strategy **OC** loses a game can be given by

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

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

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

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

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

3.2 3-strategy game dynamics

The previous fitness functions convey the average payoff pertaining to 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) [65]. 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 $p_{X,Y}$ [73] defined as

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

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 [52, 83]. We use $\beta = 1$ in our analysis: $\beta \ll 1$ corresponds to the so-called limit of weak selection, assuming that payoffs have almost no influence on the imitation process; $\beta \gg 1$ corresponds to strong selection, reducing imitation to a semi-deterministic process where any (small) difference in fitness leads to a sizable effect on imitation probabilities [73]. By using $\beta = 1$ we focus on an intermediate regime where fitness impacts imitation while allowing for stochastic effects associated with errors (e.g., in assessing others' fitness values).

The probability that one agent with strategy Y 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

$$T_{Y \rightarrow X}(k, l) = (1 - \mu) \frac{l}{Z} \frac{k}{Z - 1} p_{Y,X} + \mu \frac{l}{2Z}, \quad (11)$$

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 [58, 63, 72]. Likewise, the probability that one less agent adopts strategy X , by adopting strategy Y , is given by

$$T_{X \rightarrow Y}(k, l) = (1 - \mu) \frac{k}{Z} \frac{l}{Z - 1} p_{X,Y} + \mu \frac{k}{2Z}. \quad (12)$$

We are now able to define a Markov Chain where each state corresponds to a particular combination of three strategies (or two strategies, as in Fig. 1 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 independent of initial conditions), conveying the information about the long-term behavior of this chain (limiting and occupancy distribution) [35]. The stationary distribution represented in vector $\pi = [\pi_s]$ thus translates the long-run fraction of the time the system spends in each state $s = (k, l)$ —where k X agents, l Y agents and $(Z - l - k)$ W agents exist. This distribution is calculated as $\pi = \pi T$, where T is the transition matrix constructed resorting to the transition probabilities analogous to Eqs. (11) and (12), such that

$$\begin{aligned}
T_{(k,l) \rightarrow (k-1,l+1)} &= T_{X \rightarrow Y}(k, l) \\
T_{(k,l) \rightarrow (k+1,l-1)} &= T_{Y \rightarrow X}(k, l) \\
T_{(k,l) \rightarrow (k+1,l)} &= T_{W \rightarrow X}(k, l) \\
T_{(k,l) \rightarrow (k-1,l)} &= T_{X \rightarrow W}(k, l) \\
T_{(k,l) \rightarrow (k,l+1)} &= T_{W \rightarrow Y}(k, l) \\
T_{(k,l) \rightarrow (k,l-1)} &= T_{Y \rightarrow W}(k, l) \\
T_{(k,l) \rightarrow (k,l)} &= 1 - \sum_{i,j \in \{X,Y,W\}, i \neq j} T_{i \rightarrow j}(k, l) \\
T_{(x,y) \rightarrow (w,z)} &= 0, \text{ otherwise.}
\end{aligned} \tag{13}$$

A similar stationary distribution can be obtained through simulations, yet that would require intensive computational resources to obtain numerically precise results (as performed in Sect. 3.3 for an enlarged strategy space). Regarding the complexity of such simulations, see [30]. Our analytic approach (also recently used in [57] and, for example, [77]) 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 (e.g., see [58]). The gradient of selection portrays, for each configuration, the most likely evolutionary path. These gradients of selection read as

$$G(k, l) = (T_X^+(k, l) - T_X^-(k, l), T_Y^+(k, l) - T_Y^-(k, l)), \tag{14}$$

where, as an example, T_X^+ and T_X^- are the probabilities of having one more/less agent adopting strategy X , which are calculated as

$$T_X^+(k, l) = T_{Y \rightarrow X}(k, l) + T_{W \rightarrow X}(k, l) \tag{15}$$

and

$$T_X^-(k, l) = T_{X \rightarrow Y}(k, l) + T_{X \rightarrow W}(k, l) \tag{16}$$

Using these tools, in the top panels of Figs. 1, 2 and 3 we represent the gradient of selection (streamlines); in the bottom panel of Fig. 1 and the background of the simplexes in Figs. 2 and 3 we represent the stationary distribution(s).

3.3 Agent-based simulations

The analytic model previously presented provides important insights about the dynamics of outcome-based strategies in CRD. The analytic expressions for the average payoff obtained by each strategy (Eqs. 1–9) can be easily written as we limit the number of strategies to three. Similarly, we can effortlessly write the transition probabilities between strategic configurations (Eqs. 11–13) as we constrain our analysis to a 2-dimensional space. An analytical study of the population dynamics with more strategies would imply alternative methods [78]. In reality, however, the set of strategies studied so far (**OC**, **SC**, **C**, **D**) represents a sample from a larger strategy space. We can conceive a general conditional strategy in CRD as 3-bit tuple, based on all the combinations of actions and outcomes. As represented in Table 1, this results in eight possible strategies. Limiting the space of strategies considered can introduce non-trivial biases in the obtained results [17]. To test whether strategy

Table 1 Full strategy space studied through simulations

$\sigma[2]$: Play? D & Lost	$\sigma[1]$: Play? D & Won	$\sigma[0]$: Cooperate?	σ , Strategy
0 (No)	0	0 (Defect)	0
0	0	1 (Cooperate)	1 (SC)
0	1 (Yes)	0	2
0	1	1	3 (OC)
1	0	0	4
1	0	1	5
1	1	0	6 (D)
1	1	1	7 (C)

A strategy (σ) is here seen as a 3-bit string. The first (least significant, $\sigma[0]$) bit represents Cooperation (1) or Defection (0) when a game is accepted. The second bit ($\sigma[1]$) codifies group acceptance when a Defector is identified in the group and a game was previously won. As before, one means accepting to take part in the group while zero means rejection. The third (most significant, $\sigma[2]$) bit codifies acceptance/rejection when a Defector is identified in the group and a game was previously lost. We assume that a game is always accepted if no Defector is identified in the group. The last column of this table shows the strategy in a decimal format, where we also name the four strategies comprehensively analyzed in the analytic model (Sect. 3.1)

OC remains prevalent when the full repertoire of strategies is considered, we run agent-based simulations as fully described in Algorithm 1. As before, agents revise their behaviors through social learning [65], yet having now a wider spectrum of available strategies.

Algorithm 1: Pseudo-code of the main simulation cycle, exemplifying a single run, r . σ_i represents the strategy of agent i . The strategy can be seen as a length 3 bit string and, to simplify notation, codified in an integer between 0 and 7; $\sigma[0]$ represents the least significant bit and $\sigma[2]$ the most significant; s_i is a flag indicating whether agent i has had success in the last played game ($s_i = 1$ indicates success and $s_i = 0$ unsuccessful); g_r , c_r , a_r and m_r keep track of the i) total number of groups sampled to play a game in a run r , ii) total number of cooperative acts in run r , iii) total number of groups in which all agents accept to play and iv) total number of group that play and that achieve the threshold M (thereby having collective success). Finally, $X \sim \mathcal{U}(a, b)$ represents an integer randomly sampled, from the set $\{a, a + 1, \dots, b\}$, assuming a uniform probability distribution. $X \sim \mathcal{U}(0, 1)$ represents the standard uniform continuous distribution, from which a real number between 0 and 1 is sampled. For the reported results, we considered $Gens=25000$ and a number of samples to calculate fitness of $S = 5$ times the population size Z . Also, Fig. 6 represents the average over 100 runs.

```

Initialize:  $\sigma_i \leftarrow X \sim \mathcal{U}(0, 7)$ ,  $s_i = 1$ ,  $g_r = 0$ ,  $c_r = 0$ ,  $a_r = 0$  and  $m_r = 0$ 
for  $t \leftarrow 1$  to  $Gens$  do
  for  $j \leftarrow 1$  to  $Z$  do
    /* Sample two agents from the population */
     $A \leftarrow X \sim \mathcal{U}(1, Z)$  (agent to update)
     $B \leftarrow X \sim \mathcal{U}(1, Z)$ ,  $A \neq B$  (model agent)
    if  $X \sim \mathcal{U}(0, 1) < \mu$  then Mutation:
       $\sigma_A \leftarrow X \sim \mathcal{U}(0, 7)$ 
    else Imitation:
       $f_A \leftarrow fitness(A)$ 
       $f_B \leftarrow fitness(B)$ 
       $probab \leftarrow (1 + e^{-\beta(f_B - f_A)})^{-1}$ 
      if  $X \sim \mathcal{U}(0, 1) < probab$  then
         $\sigma_A \leftarrow \sigma_B$ 

Function fitness(i):
  fitness = 0
  for  $j \leftarrow 1$  to  $SZ$  do Calculate fitness as the average payoff in many games,  $S \times Z$ :
     $g_r = g_r + 1$ 
    1. Sample randomly  $N - 1$  agents to form group  $G$  and play with  $i$ .
    2. Calculate  $n_D$ , the number of defective agents.
    3. Check if everyone accepts playing in the group  $G$ , using Eq.(17); each of
       the  $n_D$  Defectors is identified with probability  $q$ ; agent  $k$  accepts to play in
       group  $G$  with probability  $p(n_D)$ . If the game is accepted  $a_r = a_r + 1$ ,
        $c_r = c_r + N - n_D$  and proceed to step 4. Otherwise, everyone gets 0 and the
       main cycle continues to iteration  $j + 1$ :
        $p(n_D) = (\sigma_i[2](1 - s_i) + \sigma_i[1](s_i))(1 - (1 - q)^{n_D - (1 - \sigma_i[0])}) + (1 - q)^{n_D - (1 - \sigma_i[0])}$ 
    4. Check if collective success is achieved; success occurs if the number of
       cooperators is equal or higher than  $M$ . Update  $fitness$  according to:

    if  $N - n_D \geq M$  then Success:
       $m_r = m_r + 1$ 
      foreach  $k$  in  $G$  do
         $s_k = 1$ 
        if  $\sigma_i[0] = 1$  then  $i$  is a Cooperator:
           $fitness = fitness + b - c$ 
        else
           $fitness = fitness + b$ 
    else Unsuccess:
      if  $\sigma_i[0] = 1$  then  $i$  is a Cooperator:
         $fitness = fitness - p - c$ 
      else
         $fitness = fitness - p$ 
    return  $\frac{fitness}{SZ}$ 
  
```

End Function

We also underline that, in the simulations, we account for the fact that a Defector may not always be promptly identified. We assume that a Defector is only identified with a probability q . This said, the probability that a game is accepted by an agent

Table 2 List of mathematical symbols used

Symbol	Meaning
N	Group size
b	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
q	Probability of identifying a defector

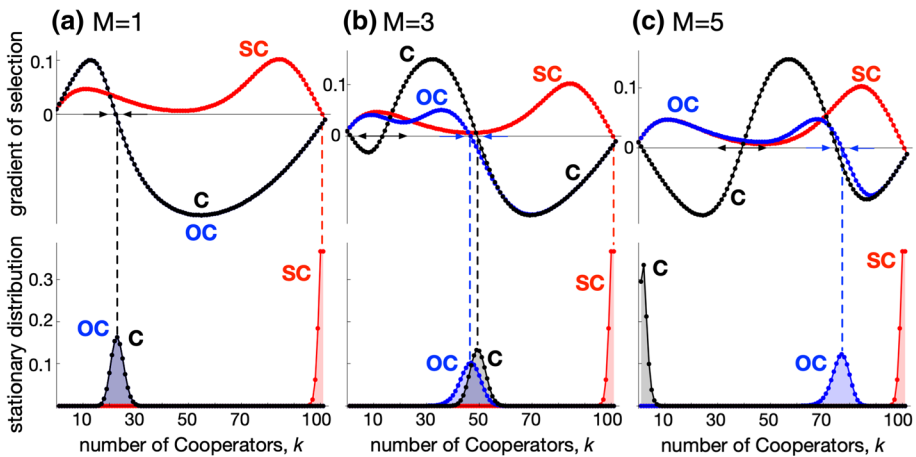


Fig. 1 2-strategy dynamics of OC, C or SC against D. The top panels represent the gradient of selection (the most plausible evolutionary path; when above the horizontal xx axis, it is more likely that cooperators—C, OC or SC—spread; below the horizontal axis, it is more likely that D spreads; this way, arrows on top of the xx axis represent the most likely direction of evolution). The bottom panels represent the stationary distribution, i.e., the long-run fraction of the time the system spends in each state. We test different dilemmas: $M = 1$ (a), $M = 3$ (b) and $M = 5$ (c). For $M = 1$ (a), OC and C are equivalent, as only one cooperator is needed to have collective success and thus an OC agent is always successful and always accepts to play with D—thus behaving as a C. The differences between OC and C become evident for higher M (e.g., $M = 3$, panel b). For a strict dilemma where a lot of cooperators are required for collective success ($M = 5$, panel c), we can observe that strategy C (black curves) is unable to invade a population of Ds; strategy SC (red curves) has, for all k , a higher probability of being adopted than Ds, leading to a high prevalence of SC (portrayed by the red distribution skewed to the right, in the bottom panels), which is supported by the positive gradient of selection. Finally, strategy OC (blue curves) is able to invade the population of Ds and stabilize a configuration in which OCs and Ds co-exist. Parameters used: $r = 1$, $\mu = 0.01$, $N = 8$, $b = 10$, $c = 2$, $Z = 100$, $p = 2$, $q = 1$ (Color figure online)

A that only accepts playing in groups where Defectors are not identified is given by $(1 - q)^{n_D - (1 - \sigma_i[0])}$, where n_D is the total number of Defectors in the group and $(1 - \sigma_i[0])$ accounts for the fact that agent A may itself be one of the Defectors ($\sigma_i[0] = 1$ if the agent is a Cooperator). All together, the probability that an agent i with strategy σ_i accepts to take part in a group is given by,

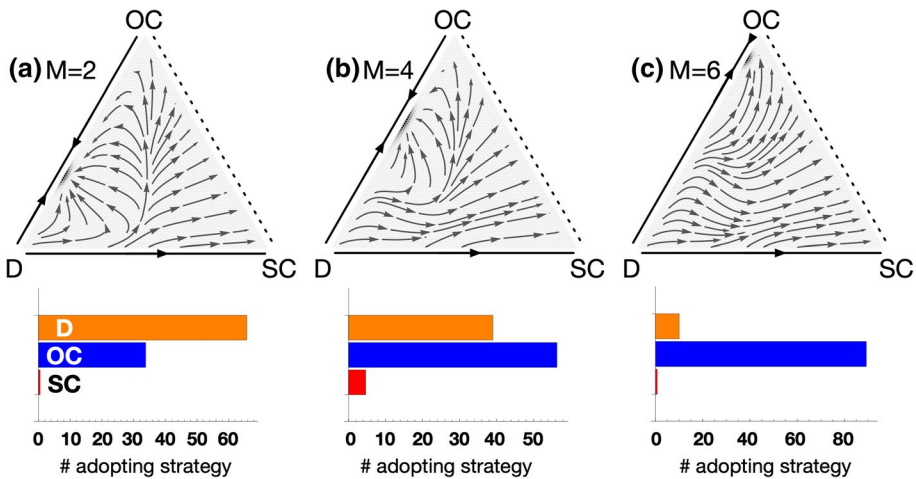


Fig. 2 3-strategy dynamics between D, SC and OC strategies. In the top panels we portray the gradient of selection (streamlines pointing out the most likely direction of evolution) and stationary distribution (background grayscale; the darker, the more time is spent in that state). The vertices of the 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 three strategies co-exist. 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 and M is low, the defective strategy prevails. For high $M = 4$ (panel b) and $M = 6$ (panel c), however, most of time is spent in states where OC is highly prevalent. Parameters: $r = 1$, $\mu = 0.01$, $N = 8$, $b = 10$, $c = 2$, $p = 2$, $Z = 100$, $q = 1$

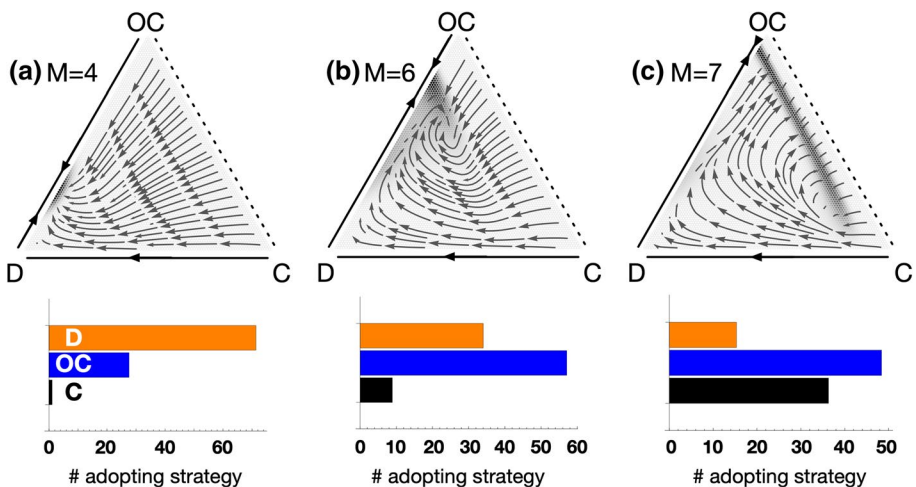


Fig. 3 3-strategy dynamics between D, C and OC strategies. The interpretation of these plots follows Fig. 2. We can observe that, when D, C and OC co-exist and M is high, a lot of time is spent in states with high prevalence of C and OC. Parameters: $r = 1$, $\mu = 0.01$, $N = 8$, $b = 10$, $c = 5$, $p = 2$, $Z = 100$, $q = 1$

$$(\sigma_i[2](1 - s_i) + \sigma_i[1](s_i))(1 - (1 - q)^{n_D - (1 - \sigma_i[0])}) + (1 - q)^{n_D - (1 - \sigma_i[0])} \quad (17)$$

where s_i is a flag indicating if agent i won ($s_i = 1$) or lost ($s_i = 0$) the previous played game, and $\sigma[i]$ is the bit i of strategy σ , as represented in Table 1.

Table 2 summarizes the notation used. As we will see next, the simulations performed confirm that, whenever defectors can be identified in a group (relatively high q , $q > 0.1$), cooperation can emerge and be stable through the high prevalence of strategies **OC** and **C**.

4 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 cooperative 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. 1) 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. 1 portray the gradient of selection (top) and the stationary distribution (bottom) when considering that only two strategies are present in the population. This analysis nicely characterizes the competition between cooperators (**C**, **SC** and **OC**) and unconditional Defectors (**D**), missing however the potentially important interplay among cooperative strategies. The 2-strategy dynamics—along the edges of a 3-dimensional simplex—can be obtained resorting to the 3-strategy models presented. Namely, (1) the dynamics of strategy **C** against **D** was obtained from Scenario 1 (Sect. 3.1.1) considering $k = 0$ (**SC** absent from the population), (2) the dynamics of strategy **SC** against **D** was obtained from Scenario 1 considering $l = 0$ (**C** absent from the population) and finally, (3) the dynamics of strategy **OC** against **D** was obtained from Scenario 2 (Sect. 3.1.2) considering $k = 0$ (**SC** absent from the population).

In Fig. 1 we show that **SC** is the strategy allowing the higher prevalence of cooperators (for the scenario $N = 8$, $b = 10$, $c = 2$, $p = 2$). This occurs as **SC** prevents the exploitation from **D** agents, by refusing to take part in groups with defectors. This way, 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 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 ($k = 50$). We also observe that **SC** always leads to full cooperation, even if that is not necessary to guarantee high levels of group achievement (i.e., obtaining M cooperators per group to obtain success). As Fig. 1c conveys, **OC**, on the other hand, allows for an efficient co-existence of **D** and **OC** that maintains high levels of group achievement at lower contribution cost.

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. 2. 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 **Ds**, while reaping the benefits of playing in groups that achieve collective success (even if they have some **D** partners, less than $N - M$). The streamlines in Fig. 2 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. 1. However, in the interior of the simplexes (when three strategies co-exist and for high M , panels b and c) the gradients point, in a large fraction of configurations, right and upwards, which indicates a higher probability that **OC** successively replaces **D** and **SC**.

When **C**, **OC** and **D** are considered, we also observe a high prevalence of **OC** (Fig. 3). For $M = 6$ (panel b, Fig. 3), we notice a cycling dynamics: **OCs** are needed to initially punish **Ds** and open space for the evolution of **Cs**; 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 **OCs**—which, with some probability, still refuse to play in groups with **Ds**. With the increased number of successful groups, **OCs** will increasingly play as **C**, making these two strategies almost neutral, i.e., receiving a very close fitness. Whenever **Cs** replace **OCs**, the barriers for the subsequent invasion of **Ds** 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.

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, we verify that **OC** profits from high M (Fig. 5).

The previous analysis and results assumed the co-existence of three strategies (maximum) and we focused our attention in the combinations **OC–C–D** and **OC–SC–D**. As pointed before and evidenced in Table 1, those triples of strategies correspond to a sample from a larger strategy space. However, through simulations following Algorithm 1, we confirm the high prevalence of strategy **OC**, even in an enlarged strategy space. In Fig. 6 we show that, whenever q (the probability of identifying a Defector) is high enough, cooperation emerges even in CRDs with high M . In this case, cooperation is supported by a co-existence of unconditional cooperators (**C**, black graph on Fig. 6a and b) and outcome-based cooperators (**OC**, blue). Naturally, cooperation requires higher q to emerge if c —the cost of cooperating—is also higher (Fig. 6 right panels, b and d). We shall highlight that, when the simulations reported in Fig. 6 are repeated only considering unconditional cooperation (**C**) and unconditional defection (**D**)—i.e., in the absence of partner-selection strategies—cooperation and group success vanish, similarly to what is observed for $q \approx 0$.

5 User studies

The previous theoretical results shed light on the evolutionary advantages of outcome-based partner selection on CRDs. Those results inform about the *ultimate causes* of adopting such strategies, also allowing to infer the long-term dynamics of cooperation and coordination in human and human–agent hybrid populations [11, 46, 64]. Notwithstanding, we confirm that humans employ outcome-based strategies when playing CRDs with agents for a short period of time. Here we describe the game used to perform such experiments (*For The Record*). For the sake of completeness, we recover the experimental analysis proposed in [9] and discuss the obtained results from the perspective of partner selection in CRD. These results served as inspiration to perform the theoretical analysis presented in the previous sections. In turn, a new set of online experiments were designed from scratch

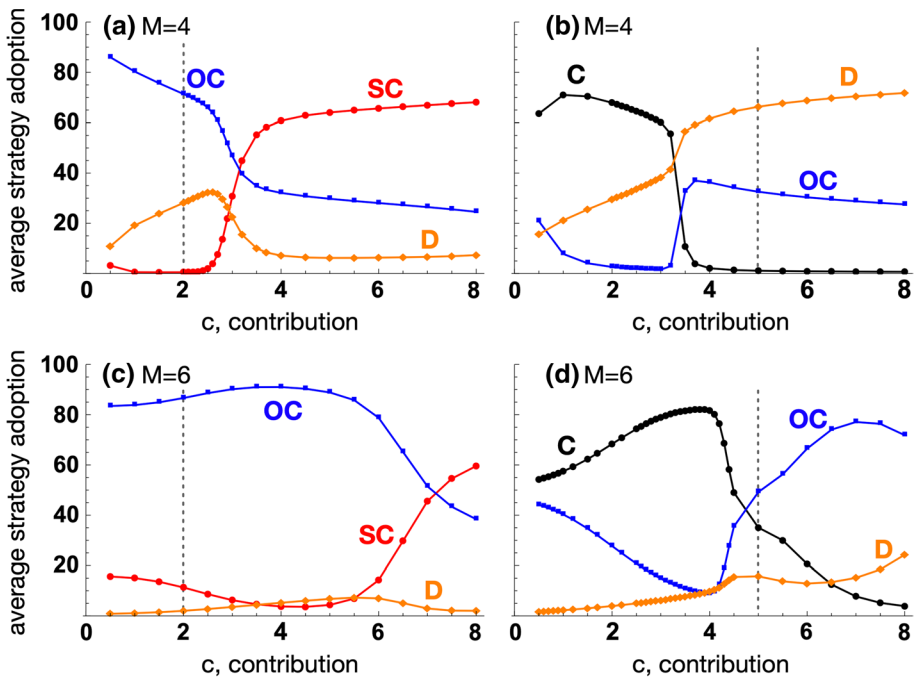


Fig. 4 The advantages of OC over SC are more evident for low c , that is, whenever cooperation requires paying less costs (panels a and c). Contrarily, the advantages of OC over C are more evident for high c (panels b and d). In both cases, higher M favors OC, which here can be observed by comparing $M = 4$ (panels a and b), with $M = 6$ (panels c and d). The vertical dashed lines represent the scenarios tested in Fig. 5, where we further analyze the effect of varying M ; we focus on the cost configurations that favor OC, namely low c for Scenario 1 (OC–SC–D) and high cost for Scenario 2 (C, SC, D). Other parameters: $r = 1$, $\mu = 0.01$, $N = 7$, $Z = 100$, $p = 2$, $b = 10$, $q = 1$

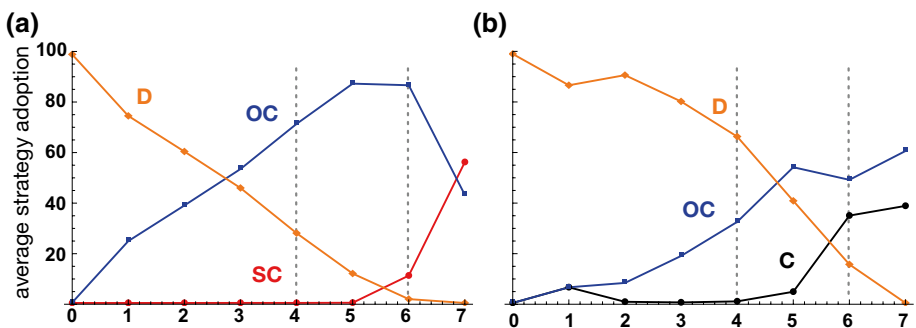


Fig. 5 The advantages of OC are (in general) more evident for high M , that is, the situations in which is the collective goal requires more cooperators. For the extreme case $N = M$, we find that SC is slightly more prevalent than OC. Here we depict the effect of M for values of cost c where OC prevails (following Fig. 4; dashed vertical lines show the values of M explored in Fig. 4). Other parameters: $r = 1$, $\mu = 0.01$, $N = 7$, $Z = 100$, $p = 2$, $b = 10$, $c = 2$ (left panel, a) and $c = 5$ (right panel, b), $q = 1$

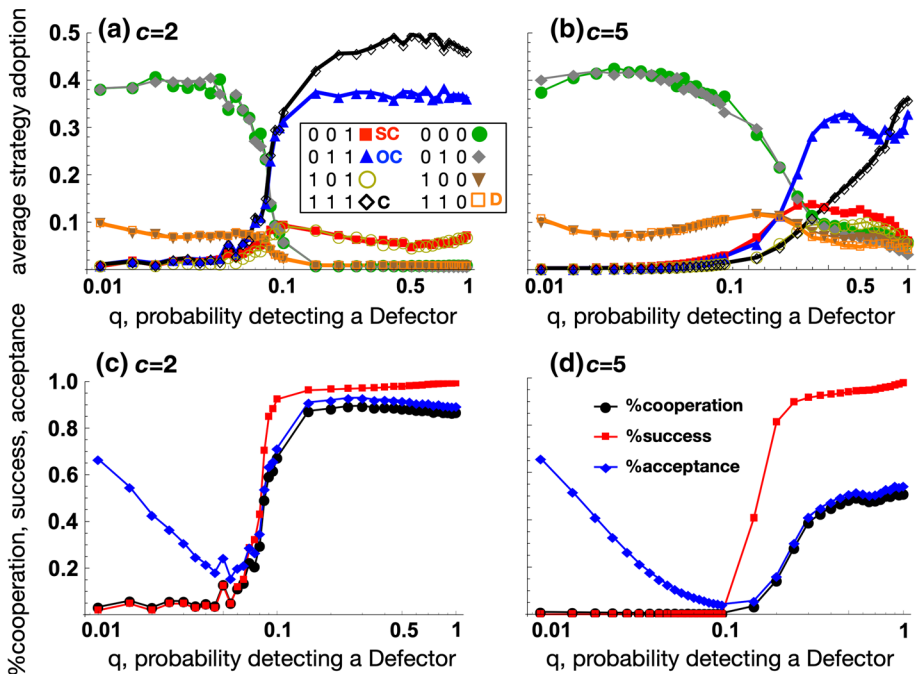


Fig. 6 Results of agent-based simulations, following Algorithm 1. Top panels (a and b) represent the prevalence of strategies as a function of q , the probability of identifying a defector. The strategies represented follow the configuration of bits detailed in Table 1, which, to ease readability, we reproduce in the inset legend. Bottom panels (c and d) represent the resulting cooperation (black), success (red) and game acceptance frequencies (blue). We can observe that defective strategies (represented on the right side of the legend) prevail when q is low. As we are focusing on the scenario of high $M = 6$, achieving cooperation requires a challenging coordination effort from the cooperators. This is naturally alleviated if Defectors can be identified in the population. As q increases, cooperation increases at the expense of a higher adoption of strategies C and OC. We confirm that OC is a highly prevalent strategy—being only slightly less adopted than C—even considering the possible co-existence of the full space of strategies represented in Table 1. These results represent the average over 100 runs of simulating the code in Algorithm 1 for 25,000 generations. Other parameters: $N = 7$, $\mu = 0.01$, $M = 6$, $Z = 100$, $p = 2$, $b = 10$ (Color figure online)

to validate the theoretical conclusions. Below we also present those experiments and the obtained results.

5.1 For the Record

For the Record is a game based on an iterated public goods dilemma, also containing some additional ludic elements added to its design such as a musical theme and having players roll dice to determine the outcomes of actions [9, 10]. Players assume the role of musicians that form a band together with the goal of recording and selling successful albums. Each player has two skills, instrument and marketing, that are used for distinct actions during the game and can be upgraded in the beginning of each new round. To maximize the chances of an album to be successful, players can spend skill points on their respective instrument—increasing the number of dice they role when determining their contribution to the band’s collective ability (i.e., the sum of all players’ contributions in one round). For

an album to be considered successful, its quality must surpass a threshold value, which is also determined by dice rolling. The threshold represents a collective risk for the band as albums that fail to surpass it are considered failures and players are unable to make any profit out of them. If an album is successful, the gains of each player are determined by throwing the accumulated number of dice in marketing skills. This way, the decision of a player involves a dilemma that opposes group success—more likely when investments in instrument skills occur—and individual success—that increases with investment in personal marketing skills.

In the first user study, players played a version of the *For the Record* game in the presence of two robots as teammates; in the follow-up study, participants played an online version of the game with two artificial agents. In both settings, the total amount of rounds/albums was set to 5. Players start the game with both skills, instrument and marketing, at level 1, which means in the first round they can roll only 1 die to improve the album or to determine their own profit through personal marketing (assuming the album achieved success), respectively. At the start of each round, each player first decides to invest 1 die (6 faces) to improve the level of either the instrument or the marketing skill. Therefore, in the last round, a player may reach up to level 5 in one of the two skills, according to her investment decision. An album achieves success if its final value surpasses a threshold, given by rolling 2 dice with 20 faces for the first three rounds and 3 dice with 20 faces in the last two rounds. This increase in difficulty is communicated to the player as the band deciding to transition to the international market. The value of the album is given by the sum of throwing all dice invested by the 3 players. If the album achieves success, each individual will earn a profit that is determined by multiplying the result of throwing the dice invested in marketing by 1000\$. In the last round, 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 3.5 in the last round (as the level of his marketing skill was never improved). 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). We shall also note that we need at least 2 cooperators (out of 3 players) in the group ($M = 2$), for the expected value of the album—given by throwing all dice invested in instrument skills by the cooperators—to surpass the expected value of the threshold in the last round—given by throwing three dice with twenty faces. The cost of cooperating is expected to be $c = 14$ ($b = 17.5$; $b - c = 3.5$ and thereby $17.5 - c = 3.5$). This way, we tested high expected values of c and M , relatively to b and N : $c/b = 0.8$ and $M/N = 0.66$.

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 [53].

5.2 Human–robot experiments

We experimentally tested *For the Record* using a 3-players setting, in which two robotic agents played with a human player. The goal was not only to compare how people perceive

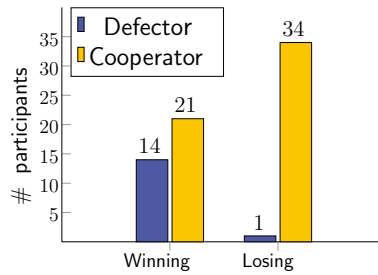


Fig. 7 Behavioral experiments on partner selection grouped by conditions, i.e., if collective goals were achieved in the last round (winning) or not (losing). The results suggest that cooperative partners (yellow bars) are only preferred whenever collective success is not achieved. In *winning* configurations, humans select the cooperative or defective opponents almost alike (Color figure online)

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 (adopting strategy **C**, see Table 1), whilst the other one (the defector) unconditionally opted to defect (adopting strategy **D**). 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 min 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 findings regarding partner selection revealed a significant association between the preferred robot and the game result ($\chi^2(1) = 14.339$, $p < 0.001$, $\phi_c = 0.453$). Further analysis of the same preferences across conditions (see Fig. 7) showed that 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 [9]. These findings inspired us to develop the aforementioned evolutionary game theoretical model to interpret the advantages of selecting cooperative partners only when a previous game was lost.

5.3 Online experiments

In order to check if the results obtained in the the evolutionary model (that a predominant strategy for our game is to pick team members depending on previous outcomes) match those of a scenario with real users, we conducted an online user study. The following hypotheses were considered:

- **H1:** The more defectors a team has, the less players are willing to join it.

- **H2:** Players who lost a previous CRD are more willing to join a team with no defectors—compared with those that previously achieved success in a CRD.
- **H3:** Players who won a previous CRD are more willing to join a team which has defectors—compared with those that were previously unsuccessful in a CRD.

To test these hypotheses, participants played *For the Record* alongside two AI agents. Similarly to the experiment using robots, one of the agents was a pure cooperator (played for the benefit of the band in all rounds) and the other was a pure defector (played for his own benefit in all rounds). We again scripted the result of the dice rolls so that the outcome of the game played (i.e., win or loss) by each participant was manipulated. In the end, we checked the participants' tolerance to defectors by measuring, on a 7-point Likert scale, their willingness to join three different hypothetical teams of two players (with zero, one and two defectors). When describing these hypothetical teams, we only showed how the team's players acted in a previous game session. The questions approaching this measure were modeled as:

Just by knowing how the other two players in each team invested their skill points in a previous session of *For the Record*, please indicate (from 1 to 7) how much you would like to join the following hypothetical teams:

- Player 1: Invested 4 points in the [Marketing/Instrument] skill.
- Player 2: Invested 4 points in the [Marketing/Instrument] skill.

The procedure used throughout the online experiments is detailed in the following section.

5.3.1 Experimental procedure

In the beginning, participants read a tutorial on how to play *For the Record*. The tutorial included a detailed description of the rules presented in Sect. 5.1 along with the actions which could be taken throughout the game. Illustrative examples displaying the game's interface elements were also added to ease comprehension.

After going through the tutorial, participants completed a quiz which aimed to check their comprehension of how to play. They could not advance until they answered the quiz questions correctly. The quiz included questions such as "How many players will play the game *For the Record*, including yourself?". Next to completing the quiz, they played the game with two AI agents as described in the beginning of this session. When participants finished playing the game, they were asked to complete a questionnaire.

In the end, participants received a monetary reward (\$4). This reward was given regardless of the obtained outcome while playing. The mean time spent by each participant while doing our experiment was 36.8 min. This time included ≈ 15 min for the tutorial and quiz, ≈ 10 min for completing the game and ≈ 15 min for the final survey.

5.3.2 Sample

Our experiment was executed resorting to the crowdsourcing website Mechanical Turk (MTurk) [50].² From MTurk, we extracted data from 90 participants. The validity of the

² <https://www.mturk.com/>.

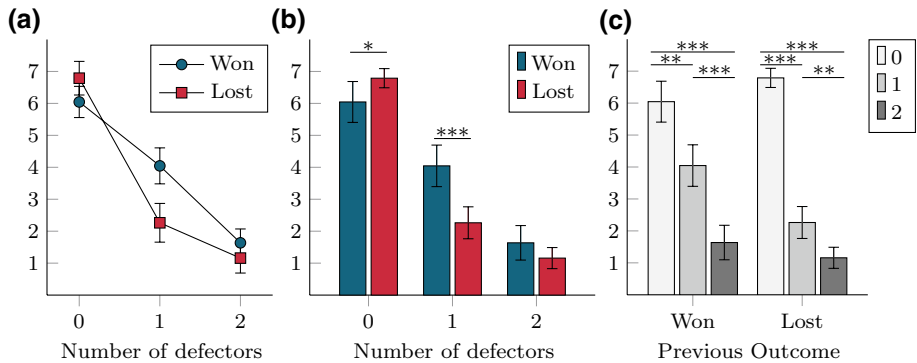


Fig. 8 Willingness to join a team according to the number of defectors on the team (within-subjects factor) and the previous outcome (between-subject factor) obtained by the participant. Error bars are the 95% confidence intervals of the means. The marks *, **, and *** represent $p < 0.05$, $p < 0.01$, and $p < 0.001$, respectively. Different charts group the information differently. **a** Estimated marginal means computed by Mixed ANOVA with a significant interaction effect between the within-subjects and the between-subjects factors, $p < 0.001$. **b** Comparisons between the previous outcomes across different number of defectors on the team. **c** Comparisons between different number of defectors on the team across the levels of the previous outcome

data was approached by asking participants two questions aimed to test their attention (“How many albums the band must fail in order to collapse?” and “How many dice are rolled for the international market?”), and four questions aimed to check if our manipulation of the agents’ strategies was perceived (such as “How many times did [one of the AI Players] upgrade the [instrument/marketing] skill?”). After removing non-valid entries (participants who did not answer correctly to either the questions which checked their attention or their knowledge about the game presented in the beginning of the section), only 41 participants remained. 26 of the participants were male (63.41%). The participants’ ages ranged from 22 to 60 ($M = 38.195$, $SD = 9.357$).

5.3.3 Results

In order to assert our hypotheses, we run a Mixed analysis of variance (ANOVA) on the willingness to join a team with the number of defectors as the within-subjects factor and the previous outcome of the participant as the between-subjects factor.

There was a statistically significant main effect of the previous outcome on the willingness to join a hypothetical team ($F(1, 39) = 9.994$, $p = 0.003$, $r = 0.452$). Participants that won a previous game reported higher willingness to join another team ($M = 3.909$, $SD = 0.511$), compared to participants that lost a previous game ($M = 3.404$, $SD = 0.510$). There was also a significant main effect of the number of defectors ($F(2, 78) = 164.020$, $p < 0.001$). As the number of defectors on the hypothetical teams increases (i.e., 0, 1, and 2), the reported willingness to join those teams decreases ($M = 6.147$, $SD = 1.140$; $M = 3.154$, $SD = 1.313$; $M = 1.397$, $SD = 1.012$, respectively). Furthermore, we found a significant interaction effect between the previous outcome of the participant and the number of defectors (Fig. 8a, $F(2, 78) = 10.085$, $p < 0.001$).

To break down this interaction, two contrasts were performed: (1) comparing each level of the previous outcome across teams with different number of defectors, using

Mann-Whitney U tests; and (2) comparing the number of defectors across each level of previous outcome, using Friedman's ANOVA tests.

In the first contrast analysis (Fig. 8b), there was a significant difference of the previous outcome of the participant for teams with 0 and 1 defectors ($U = 138.000, Z = -2.314, p = 0.021, r = 0.361$ and $U = 68.500, Z = -3.765, p < 0.001, r = 0.588$, respectively). When evaluating the willingness to join a team with zero defectors, participants that won a previous game reported lower levels ($M = 6.045, SD = 1.135$) compared to participants that lost a previous game ($M = 6.789, SD = 1.133$). However, for the willingness to join a team with one defector, participants that won a previous game reported higher levels ($M = 4.045, SD = 1.309$) compared to participants that lost a previous game ($M = 2.263, SD = 1.308$). Regarding teams with two defectors, we did not find a significant difference between participants that previously won ($M = 1.636, SD = 1.008$) and lost ($M = 1.158, SD = 1.007$).

In the second contrast analysis (Fig. 8c), we found a significant difference between the three levels of defectors on the team (i.e., 0, 1, and 2) in the group of participants that previously won a game ($\chi^2(2) = 34.400, p < 0.001$). Post-hoc analyses using the Wilcoxon Ranks tests allowed for further pairwise comparisons. The willingness to join a team with 1 defector was lower than with 0 defectors ($Z = -2.752, p = 0.006, r = 0.587$), as well as the willingness to join a team with 2 defectors was lower than both with 1 defector ($Z = -3.876, p < 0.001, r = 0.826$) and with 0 defectors ($Z = -3.982, p < 0.001, r = 0.849$). In the group of participants that previously lost the game, we also found a significant difference between the three levels of defectors on the team ($\chi^2(2) = 32.380, p < 0.001$). Similarly to the previous group, the willingness to join a team with 1 defector was lower than with 0 defectors ($Z = -3.848, p < 0.001, r = 0.883$), as well as the willingness to join a team with 2 defectors was lower than both with 1 defector ($Z = -2.980, p = 0.003, r = 0.684$) and with 0 defectors ($Z = -4.184, p < 0.001, r = 0.960$).

Overall, these results mostly support our three hypotheses. Firstly, the significant main effect of the number of defectors on the willingness to join a team validated H1. Indeed, as the number of defectors on a team increased, the willingness to join it decreased. This result also highlights that the number of defectors is an important consideration when joining a new team. Secondly, H2 was also validated based on the significant difference of the previous outcome on the willingness to join a team without defectors, which was indeed higher for participants that lost a previous game compared to participants that won. Finally, H3 was only partially verified. Although participants that won a previous game were more willing to join a team with one defector compared to participants that lost, no similar significant difference was found for joining teams with two defectors. On the one hand, a positive outcome can indeed positively influence the tolerance to have one defector on the team. On the other hand, however, this effect may disappear when the number of defectors is too high.

Interestingly, when analyzing the same results on a different perspective, by grouping participants according to their previous outcome, additional conclusions can also be drawn. The difference between the willingness to join a team with zero defectors and a team with one defector is much more salient in the group that lost the previous game than within the group that won (i.e., effect sizes of $r = 0.883$ and $r = 0.587$, respectively). This difference emphasizes how previous outcomes influence the predisposition to tolerate defectors. In particular, how losers assert a preference for a team without defectors, while winners become more lenient in the same choices.

An important remark of our user study, for the purpose of generalizing the current findings, is the fact that participants have partnered with exactly one defector in the previous game. Further studies should analyze if similar results also occur after playing a CRD with zero or more than one defector on the team.

6 Conclusion and discussion

Here we explore partner selection in collective risk dilemmas (CRDs). In the context of prisoner's dilemmas [19] or public goods games [24, 40], previous studies found that introducing strategies that refuse playing with defectors opens space for cooperative strategies to invade the previously stable defective equilibria. In CRDs, 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. 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. To further confirm the success and usage of **OC**-like strategies, we resort to human–robot and human–agent experiments, which allows controlling agents behavior and explicitly test a cooperative and defective artificial partner. After the game, we ask the human subjects whether they would prefer teams only with cooperators or with some defectors, to play with in the future. Humans reveal significantly more tolerance for groups with one defector when winning a previous CRD. 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 [53]) 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 [76], or by stressing all combinations of strategy avoidance and actions played [49]. We further resorted to agent-based simulations in order to confirm that these analytic insights remain valid when assuming an enlarged space of strategies, considering, e.g., strategies that only accept playing with cooperators and yet decide to defect—a malicious version of **OC**. We confirmed that **OC** becomes widely adopted—at *par* with **C**—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 at least one defector, 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 resorting to reputations [59], commitments [24, 26], intention recognition [25, 27] or even more complex agent architectures [14]. Notwithstanding, through simulations, we also investigate the role of failing to identify Defectors, by considering that those agents are only identified with a probability q . We confirmed that q plays an important role in eliciting cooperation. Nonetheless, our results indicate that even for relatively low identification probabilities (e.g., one in ten possible detections of defectors) there is a high prevalence of outcome-based cooperative strategies as also shown in the experimental results.

The strategy space considered in our base model is constrained to focus on (*memory-1*) outcome-based partner selection. It is also possible to conceive, however, an enlarged strategy space allowing both (1) group participation and (2) action selection to be dependent on (i) the history of previous outcomes and (ii) anticipated strategies in the group. Previous works focus on action discrimination [49] (also with larger memory spans [29]) in the context of linear public goods games. We focus on memory-1 strategies, with conditional group participation, to highlight the (already non-trivial) dynamics associated with CRD and outcome-based partner selection.

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 [57, 60], Multiplayer Trust Games [8] or N-Person Stag-Hunt Games [44]. Also, the conclusions that we derive suggest, mainly, that **OC** generally benefits from high M and that **OC** becomes more prevalent than **SC** when the cost of cooperating (c) is low. These predictions may be tested through new experiments with human subjects, thus opening new avenues for a symbiosis between theoretical and experimental analysis of collective action problems.

Acknowledgements This work was supported by national funds through Fundação para a Ciência e Tecnologia under Projects UIDB/50021/2020, SFRH/BD/118031/2016, PTDC/EEL-SII/5081/2014, PTDC/MAT/STA/3358/2014 and PTDC/EEISII/7174/2014. F.P.S. acknowledges support from the James S. McDonnell Foundation 21st Century Science Initiative in Understanding Dynamic and Multi-scale Systems—Postdoctoral Fellowship Award. We thank Ramona Merhej and anonymous reviewers for suggesting improvements to early versions of this manuscript.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Alexander, R. D. (1987). *The biology of moral systems*. Piscataway: Transaction Publishers.
- Alvard, M. S., Nolin, D. A., Brosius, P., Kuznar, L. A., Rousseau, J., Ruttan, L. M., et al. (2002). Rousseau's whale hunt? Coordination among big-game hunters. *Current Anthropology*, 43(4), 533–559.
- Anastassacos, N., Hailes, S., & Musolesi, M. (2020). Partner selection for the emergence of cooperation in multi-agent systems using reinforcement learning. In *Proceedings of AAAI'20*. AAAI Press
- Bloembergen, D., Tuyls, K., Hennes, D., & Kaisers, M. (2015). Evolutionary dynamics of multi-agent learning: A survey. *Journal of Artificial Intelligence Research*, 53, 659–697.
- Boesch, C. (2002). Cooperative hunting roles among tai chimpanzees. *Human Nature*, 13(1), 27–46.
- Chakra, M. A., & Traulsen, A. (2012). Evolutionary dynamics of strategic behavior in a collective-risk dilemma. *PLoS Computational Biology*, <https://doi.org/10.1371/journal.pcbi.1002652>.
- Chen, X., Sasaki, T., Brännström, Å., & Dieckmann, U. (2015). First carrot, then stick: How the adaptive hybridization of incentives promotes cooperation. *Journal of the Royal Society Interface*, 12(102), 20140935.
- Chica, M., Chiong, R., Kirley, M., & Ishibuchi, H. (2017). A networked n-player trust game and its evolutionary dynamics. *IEEE Transactions on Evolutionary Computation*, 22, 866–878.
- Correia, F., Mascarenhas, S., Gomes, S., Arriaga, P., Leite, I., Prada, R., Melo, F. S., & Paiva, A. (2019). Exploring prosociality in human–robot teams. In *Proceedings of HRI'19*. IEEE Press
- Correia, F., Mascarenhas, S., Gomes, S., Tulli, S., Santos, F. P., Santos, F. C., et al. (2019). For the record—A public goods game for exploring human–robot collaboration. In *Proceedings of AAMAS'19* (pp. 2351–2353). IFAAMAS
- Crandall, J. W., Oudah, M., Ishowo-Oloko, F., Abdallah, S., Bonnefon, J. F., Cebrian, M., et al. (2018). Cooperating with machines. *Nature Communications*, 9(1), 1–12.
- Crawford, C., Nanda Kumar, R., & Sen, S. (2018). Resisting exploitation through rewiring in social networks: Social welfare increase using parity, sympathy and reciprocity. In *Proceedings of AAMAS'18* (pp. 1915–1917). IFAAMAS.
- De, S., Nau, D. S., & Gelfand, M. J. (2017). Understanding norm change: An evolutionary game-theoretic approach. In *Proceedings of AAMAS'17* (pp. 1433–1441). IFAAMAS.
- Domingos, E. F., Burguillo, J. C., & Lenaerts, T. (2017). Reactive versus anticipative decision making in a novel gift-giving game. In *Proceedings of AAAI'17* (Vol. 17, pp. 4399–4405). AAAI Press.
- Fehl, K., van der Post, D. J., & Semmann, D. (2011). Co-evolution of behaviour and social network structure promotes human cooperation. *Ecology Letters*, 14(6), 546–551.
- Fehr, E., & Gächter, S. (2002). Altruistic punishment in humans. *Nature*, 415(6868), 137.
- García, J., & Traulsen, A. (2019). Evolution of coordinated punishment to enforce cooperation from an unbiased strategy space. *Journal of the Royal Society Interface*, 16(156), 20190127.
- Genesereth, M. R., Ginsberg, M. L., & Rosenschein, J. S. (1986). Cooperation without communication. In *Proceedings of AAAI'86* (pp. 51–57). Elsevier.
- Ghang, W., & Nowak, M. A. (2015). Indirect reciprocity with optional interactions. *Journal of Theoretical Biology*, 365, 1–11.
- Gintis, H., Bowles, S., Boyd, R. T., Fehr, E., et al. (2005). *Moral sentiments and material interests: The foundations of cooperation in economic life* (Vol. 6). Cambridge: MIT Press.
- Góis, A. R., Santos, F. P., Pacheco, J. M., & Santos, F. C. (2019). Reward and punishment in climate change dilemmas. *Scientific Reports*, 9(1), 1–9.
- Griffiths, N. (2008). Tags and image scoring for robust cooperation. In *Proceedings of AAMAS'08* (pp. 575–582). IFAAMAS.
- Griffiths, N., & Luck, M. (2010). Changing neighbours: Improving tag-based cooperation. In *Proceedings of AAMAS'10* (pp. 249–256). IFAAMAS.
- Han, T. A., Pereira, L. M., & Lenaerts, T. (2017). Evolution of commitment and level of participation in public goods games. *Autonomous Agents and Multi-Agent Systems*, 31(3), 561–583.
- Han, T. A., Pereira, L. M., & Santos, F. C. (2012). Corpus-based intention recognition in cooperation dilemmas. *Artificial Life*, 18(4), 365–383.
- Han, T. A., Pereira, L. M., & Santos, F. C. (2012). The emergence of commitments and cooperation. In *Proceedings of AAMAS'13, AAMAS '12* (pp. 559–566). IFAAMAS, Richland, SC.
- Han, T. A., Pereira, L. M., Santos, F. C., et al. (2011). The role of intention recognition in the evolution of cooperative behavior. In *Proceedings of IJCAI'11*.
- Hauert, C., De Monte, S., Hofbauer, J., & Sigmund, K. (2002). Volunteering as red queen mechanism for cooperation in public goods games. *Science*, 296(5570), 1129–1132.
- Hilbe, C., Martinez-Vaquero, L. A., Chatterjee, K., & Nowak, M. A. (2017). Memory-n strategies of direct reciprocity. *Proceedings of the National Academy of Sciences*, 114(18), 4715–4720.

30. Hindersin, L., Wu, B., Traulsen, A., & García, J. (2019). Computation and simulation of evolutionary game dynamics in finite populations. *Scientific Reports*, 9(1), 1–21.
31. Hogg, T. (1995). Social dilemmas in computational ecosystems. In *Proceedings of IJCAI'95* (pp. 711–718).
32. Jennings, N. R., Sycara, K., & Wooldridge, M. (1998). A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1(1), 7–38.
33. Jervis, R. (1978). Cooperation under the security dilemma. *World Politics*, 30(2), 167–214.
34. Kollock, P. (1998). Social dilemmas: The anatomy of cooperation. *Annual Review of Sociology*, 24(1), 183–214.
35. Kulkarni, V. G. (2016). *Modeling and analysis of stochastic systems*. Boca Raton: Chapman and Hall/CRC.
36. Matarić, M. J. (1995). Issues and approaches in the design of collective autonomous agents. *Robotics and Autonomous Systems*, 16(2–4), 321–331.
37. Milinski, M., Sommerfeld, R. D., Krambeck, H. J., Reed, F. A., & Marotzke, J. (2008). The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proceedings of the National Academy of Sciences*, 105(7), 2291–2294.
38. Mor, Y., & Rosenschein, J. S. (1995). Time and the prisoner's dilemma. In *Proceedings of ICMAS'95* (pp. 276–282).
39. Morales, J., Wooldridge, M., Rodríguez-Aguilar, J. A., & López-Sánchez, M. (2018). Off-line synthesis of evolutionarily stable normative systems. *Autonomous Agents and Multi-Agent Systems*, 32, 635–671.
40. Moreira, J. A., Pacheco, J. M., & Santos, F. C. (2013). Evolution of collective action in adaptive social structures. *Scientific Reports*, 3, 1521.
41. Nowak, M. A. (2006). *Evolutionary dynamics*. Cambridge: Harvard University Press.
42. Nowak, M. A., Page, K. M., & Sigmund, K. (2000). Fairness versus reason in the ultimatum game. *Science*, 289(5485), 1773–1775.
43. Olejarz, J., Ghang, W., & Nowak, M. A. (2015). Indirect reciprocity with optional interactions and private information. *Games*, 6(4), 438–457.
44. Pacheco, J. M., Santos, F. C., Souza, M. O., & Skyrms, B. (2009). Evolutionary dynamics of collective action in n-person stag hunt dilemmas. *Proceedings of the Royal Society of London B*, 276(1655), 315–321.
45. Pacheco, J. M., Vasconcelos, V. V., Santos, F. C., & Skyrms, B. (2015). Co-evolutionary dynamics of collective action with signaling for a quorum. *PLoS Computational Biology*, 11(2), e1004101.
46. Paiva, A., Santos, F. P., & Santos, F. C. (2018). Engineering pro-sociality with autonomous agents. In *AAAI'18* (pp. 7994–7999).
47. Peleteiro, A., Burguillo, J. C., & Chong, S. Y. (2014). Exploring indirect reciprocity in complex networks using coalitions and rewiring. In *Proceedings of AAMAS'14* (pp. 669–676). IFAAMAS.
48. Peña, J., & Nöldeke, G. (2018). Group size effects in social evolution. *Journal of Theoretical Biology*, 457, 211–220.
49. Pinheiro, F. L., Vasconcelos, V. V., Santos, F. C., & Pacheco, J. M. (2014). Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Computational Biology*, 10(11), e1003945.
50. Rand, D. G. (2012). The promise of mechanical turk: How online labor markets can help theorists run behavioral experiments. *Journal of Theoretical Biology*, 299, 172–179.
51. Rand, D. G., & Nowak, M. A. (2013). Human cooperation. *Trends in Cognitive Sciences*, 17(8), 413–425.
52. Rand, D. G., Tarnita, C. E., Ohtsuki, H., & Nowak, M. A. (2013). Evolution of fairness in the one-shot anonymous ultimatum game. *Proceedings of the National Academy of Sciences*, 110(7), 2581–2586.
53. Santos, F. C., & Pacheco, J. M. (2011). Risk of collective failure provides an escape from the tragedy of the commons. *Proceedings of the National Academy of Sciences*, 108(26), 10421–10425.
54. Santos, F. C., Pacheco, J. M., & Lenaerts, T. (2006). Cooperation prevails when individuals adjust their social ties. *PLoS Computational Biology*, 2(10), e140.
55. Santos, F. P. (2017). Social norms of cooperation in multiagent systems. In *Proceedings of AAMAS'17* (pp. 1859–1860). IFAAMAS.
56. Santos, F. P., Mascarenhas, S. F., Santos, F. C., Correia, F., Gomes, S., & Paiva, A. (2019). Outcome-based partner selection in collective risk dilemmas. In *Proceedings of AAMAS'19* (pp. 1556–1564). IFAAMAS.
57. Santos, F. P., Pacheco, J. M., Paiva, A., & Santos, F. C. (2019). Evolution of collective fairness in hybrid populations of humans and agents. In *Proceedings of AAAI'19*. AAAI Press.

58. Santos, F. P., Pacheco, J. M., & Santos, F. C. (2016). Evolution of cooperation under indirect reciprocity and arbitrary exploration rates. *Scientific Reports*, 6, 37517.
59. Santos, F. P., Pacheco, J. M., & Santos, F. C. (2018). Social norms of cooperation with costly reputation building. In *Proceedings of AAAI'18* (pp. 4727–4734). AAAI Press.
60. Santos, F. P., Santos, F. C., Melo, F. S., Paiva, A., & Pacheco, J. M. (2016). Dynamics of fairness in groups of autonomous learning agents. In *International conference on autonomous agents and multiagent systems (workshops' best papers book)* (pp. 107–126). Springer.
61. Santos, F. P., Santos, F. C., & Pacheco, J. M. (2018). Social norm complexity and past reputations in the evolution of cooperation. *Nature*, 555(7695), 242.
62. Sen, S., et al. (2018). The effects of past experience on trust in repeated human-agent teamwork. In *Proceedings of AAMAS'18* (pp. 514–522). IFAAMAS.
63. Sequeira, P., Melo, F. S., & Paiva, A. (2011). Emotion-based intrinsic motivation for reinforcement learning agents. In *International conference on affective computing and intelligent interaction* (pp. 326–336). Springer.
64. Shirado, H., & Christakis, N. A. (2017). Locally noisy autonomous agents improve global human coordination in network experiments. *Nature*, 545(7654), 370–374.
65. Sigmund, K. (2010). *The calculus of selfishness*. Princeton: Princeton University Press.
66. Skyrms, B. (2004). *The stag hunt and the evolution of social structure*. Cambridge: Cambridge University Press.
67. Skyrms, B. (2010). *Signals: Evolution, learning, and information*. Oxford: Oxford University Press.
68. Smith, J. M., & Price, G. R. (1973). The logic of animal conflict. *Nature*, 246(5427), 15.
69. Souza, M. O., Pacheco, J. M., & Santos, F. C. (2009). Evolution of cooperation under n-person snow-drift games. *Journal of Theoretical Biology*, 260(4), 581–588.
70. Stander, P. E. (1992). Cooperative hunting in lions: The role of the individual. *Behavioral Ecology and Sociobiology*, 29(6), 445–454.
71. Tavoni, A., Dannenberg, A., Kallis, G., & Löschel, A. (2011). Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proceedings of the National Academy of Sciences*, 108(29), 11825–11829.
72. Traulsen, A., Hauert, C., De Silva, H., Nowak, M. A., & Sigmund, K. (2009). Exploration dynamics in evolutionary games. *Proceedings of the National Academy of Sciences*, 106(3), 709–712.
73. Traulsen, A., Nowak, M. A., & Pacheco, J. M. (2006). Stochastic dynamics of invasion and fixation. *Physical Review E*, 74(1), 011909.
74. Tuyls, K., & Nowé, A. (2005). Evolutionary game theory and multi-agent reinforcement learning. *The Knowledge Engineering Review*, 20(1), 63–90.
75. Tuyls, K., & Parsons, S. (2007). What evolutionary game theory tells us about multiagent learning. *Artificial Intelligence*, 171(7), 406–416.
76. Van Segbroeck, S., Pacheco, J. M., Lenaerts, T., & Santos, F. C. (2012). Emergence of fairness in repeated group interactions. *Physical Review Letters*, 108(15), 158104.
77. Vasconcelos, V. V., Santos, F. C., & Pacheco, J. M. (2015). Cooperation dynamics of polycentric climate governance. *Mathematical Models and Methods in Applied Sciences*, 25(13), 2503–2517.
78. Vasconcelos, V. V., Santos, F. P., Santos, F. C., & Pacheco, J. M. (2017). Stochastic dynamics through hierarchically embedded Markov chains. *Physical Review Letters*, 118(5), 058301.
79. Wang, Z., Bauch, C. T., Bhattacharyya, S., d'Onofrio, A., Manfredi, P., Perc, M., et al. (2016). Statistical physics of vaccination. *Physics Reports*, 664, 1–113.
80. Weibull, J. W. (1997). *Evolutionary game theory*. Cambridge: MIT Press.
81. Weiss, G. (1999). *Multiagent systems: A modern approach to distributed artificial intelligence*. Cambridge: MIT Press.
82. Xu, J., García, J., & Handfield, T. (2019). Cooperation with bottom-up reputation dynamics. In *Proceedings of AAMAS'19* (pp. 269–276). IFAAMAS.
83. Zisis, I., Di Guida, S., Han, T., Kirchsteiger, G., & Lenaerts, T. (2015). Generosity motivated by acceptance-evolutionary analysis of an anticipation game. *Scientific Reports*, 5, 18076.

Affiliations

Fernando P. Santos¹  · **Samuel Mascarenhas²** · **Francisco C. Santos²**  ·
Filipa Correia²  · **Samuel Gomes²**  · **Ana Paiva²**

Samuel Mascarenhas
samuel.mascarenhas@gaips.inesc-id.pt

Francisco C. Santos
franciscoesantos@tecnico.ulisboa.pt

Filipa Correia
filipacorreia@tecnico.ulisboa.pt

Samuel Gomes
samuel.gomes@tecnico.ulisboa.pt

Ana Paiva
ana.paiva@inesc-id.pt

¹ Department of Ecology and Evolutionary Biology, Princeton University, Princeton, NJ 08544, USA

² INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, Lisbon, Portugal