OPEN

Reward and punishment in climate change dilemmas

António R. Góis[1,2,3], Fernando P. Santos[4,1,2], Jorge M. Pacheco[5,6,2] & Francisco C. Santos [1,2,7*]

Mitigating climate change effects involves strategic decisions by individuals that may choose to limit their emissions at a cost. Everyone shares the ensuing benefits and thereby individuals can free ride on the effort of others, which may lead to the tragedy of the commons. For this reason, climate action can be conveniently formulated in terms of Public Goods Dilemmas often assuming that a minimum collective effort is required to ensure any benefit, and that decision-making may be contingent on the risk associated with future losses. Here we investigate the impact of reward and punishment in this type of collective endeavors — coined as collective-risk dilemmas — by means of a dynamic, evolutionary approach. We show that rewards (positive incentives) are essential to initiate cooperation, mostly when the perception of risk is low. On the other hand, we find that sanctions (negative incentives) are instrumental to maintain cooperation. Altogether, our results are gratifying, given the a-priori limitations of effectively implementing sanctions in international agreements. Finally, we show that whenever collective action is most challenging to succeed, the best results are obtained when both rewards and sanctions are synergistically combined into a single policy.

Climate change stands as one of our biggest challenges in what concerns the emergence and sustainability of cooperation[1,2]. Indeed, world citizens build up high expectations every time a new International Environmental Summit is settled, unfortunately with few resulting solutions implemented so far. This calls for the development of more effective incentives, agreements and binding mechanisms. The problem can be conveniently framed resorting to the mathematics of game theory, being a paradigmatic example of a Public Goods Game[3]: at stake there is a global good from which every single individual can profit, irrespectively of contributing to maintain it. Parties may free ride on the efforts of others, avoiding any effort themselves, while driving the population into the tragedy of the commons[4]. Moreover, since here cooperation aims at averting collective losses, this type of dilemmas is often referred as public *bad* games, in which achieving collective goals often depends on reaching a threshold number of cooperative group members[5–8].

One of the multiple obstacles attributed to such agreements is misperceiving the actual risk of future losses, which significantly affects the ensuing dynamics of cooperation[5,9]. Another problem relates to both the incapacity to sanction those who do not contribute to the welfare of the planet, and/or to reward those who subscribe to green policies[10]. Previous cooperation studies show that reward (positive incentives), punishment (negative incentives) and the combination of both[11–23] have a different impact depending on the dilemma in place. Assessing the impact of reward and punishment (isolated or combined) in the context of N-person threshold games — and in the particular case of climate change dilemmas — remains, however, an open problem.

Here we study, theoretically, the role of both institutional reward and punishment in the context of climate change agreements. Previous works consider the public good as a linear function of the number of contributors[12,17,21,22] and conclude that punishment is more effective than reward (for an optimal combination of punishment and reward see ref.[12]). We depart from this linear regime by modeling the returns on the public good as a threshold problem, combined with an uncertain outcome, represented by a risk of failure. As a result – and as detailed below – the dynamical portrait of our model reveals new internal equilibria[9], allowing to identify the dynamics of coordination and coexistence typifying collective action problems. As discussed below, the reward and punishment mechanisms will impact, in a non-trivial way, those equilibria.

[1]INESC-ID and Instituto Superior Técnico, Universidade de Lisboa, IST-Taguspark, 2744-016, Porto, Salvo, Portugal. [2]ATP-group, P-2744-016, Porto, Salvo, Portugal. [3]Unbabel, R. Visc. de Santarém 67B, 1000-286, Lisboa, Portugal. [4]Department of Ecology and Evolutionary Biology, Princeton University, Princeton, USA. [5]Centro de Biologia Molecular e Ambiental, Universidade do Minho, 4710 - 057, Braga, Portugal. [6]Departamento de Matemática e Aplicações, Universidade do Minho, 4710 - 057, Braga, Portugal. [7]Machine Learning Group, Université Libre de Bruxelles, Boulevard du Triomphe CP212, 1050, Bruxelles, Belgium. *email: franciscocsantos@tecnico.ulisboa.pt

We consider a population of size Z, where each individual can be either a Cooperator (**C**) or a Defector (**D**), when participating in a N-player Collective-Risk dilemma (**CRD**)[5,9,10,24–30]. In this game, each participant starts with an initial endowment B (viewed as the asset value at stake) that may be used to contribute to the mitigation of the effects of climate change. A cooperator incurs a cost corresponding to a fraction c of her initial endowment B, in order to help prevent a collective failure. On the other hand, a defector refuses to have any cost, hoping to free ride on the contributions of others. We require a minimum number of $0 < M \leq N$ cooperators in a group of size N before collective action is realized; if a group of size N does not contain at least M Cs, all members lose their remaining endowments with a probability r, where r ($0 \leq r \leq 1$) stands as the risk of collective failure. Otherwise, everyone will keep whatever she has. This **CRD** formulation has been shown to capture some of the key features discovered in recent experiments[5,24,31–33], while highlighting the importance of risk. In addition, it allows one to test model parameters in a systematic way that is not possible in human experiments. Moreover, the adoption of non-linear returns mimics situations common to many human and non-human endeavors[6,34–41], where a minimum joint effort is required to achieve a collective goal. Thus, the applicability of this framework extends well beyond environmental governance, given the ubiquity of such type of social dilemmas in nature and societies.

Following Chen *et al.*[12], we include both reward and punishment mechanisms in this model. A fixed group budget $N\delta$ (where $\delta \geq 0$ stands for a per-capita incentive) is assumed to be available, of which a fraction w is applied to a reward policy and the remaining 1-w to a punishment policy. We assume the effective impact of both policies to be equivalent, meaning that each unit spent will directly increase/decrease the payoff of a cooperator/defector by the same amount. For details on policies with different efficiencies, see Methods.

Instead of considering a collection of rational agents engaging in one-shot Public Goods Games[32,42], here we adopt an evolutionary description of the behavioral dynamics[9], in which individuals tend to copy those appearing to be more successful. Success (or fitness) of individuals is here associated with their average payoff. All individuals are equally likely to interact with each other, causing all cooperators and defectors to be equivalent, on average, and only distinguishable by the strategy they adopt. Therefore, and considering that only two strategies are available, the number of cooperators is sufficient to describe any configuration of the population. The number of individuals adopting a given strategy (either C or D) evolves in time according to a stochastic birth–death process[43,44], which describes the time evolution of the social learning dynamics (with exploration): At each time-step each individual (X, with fitness f_X) is given the opportunity to change strategy; with probability μ, X randomly explores the strategy space[45] (a process similar to mutations in a biological context that precludes the existence of absorbing states). With probability $(1$-$\mu)$, X may adopt the strategy of a randomly selected individual (Y, with fitness f_Y), with a probability that increases with the fitness difference (f_Y-f_X)[44]. This renders the stationary distribution (see Methods) an extremely useful tool to rank the most visited states given the ensuing evolutionary dynamics of the population. Indeed, the stationary distribution provides the prevalence of each of the population's possible configuration, in terms of the number of Cs (k) and Ds (Z-k). Combined with the probability of success characterizing each configuration, the stationary distribution can be used to compute the overall success probability of a given population – the average group achievement, η_G. This value represents the average fraction of groups that will overcome the **CRD**, successfully preserving the public good.

Results

In Fig. 1 we compare the average group achievement η_G (as a function of risk) in four scenarios: (*i*) a reference scenario without any policy (*i.e.*, no reward or punishment, in black); and three scenarios where a budget is applied to (*ii*) rewards, (*iii*) punishment and (*iv*) a combination of rewards and sanctions (see below). Our results are shown for the two most paradigmatic regimes: low (Fig. 1A) and high (Fig. 1B) coordination requirements. Naturally η_G improves whenever a policy is applied. Less obvious is the difference between the various policies. Applying only rewards (blue curves in Fig. 1) is more effective than only punishment (red curve) for low values of risk. The opposite happens when risk is high. On scenarios with a low relative threshold (Fig. 1A), rewards play the key role, with sanctions only marginally outperforming them for very high values of risk. For high coordination thresholds (Fig. 1B) reward and punishment portray comparable efficiency in the promotion of cooperation, with pure-Punishment ($w = 0$) performing slightly better than pure-Reward ($w = 1$).

Justifying these differences is difficult from the analysis of η_G alone. To better understand the behavior dynamics under Reward and Punishment, we show in Fig. 2 the gradients of selection (top panels) and stationary distributions (lower panels) for each case and different budget values. Each gradient of selection represents, for each discrete state k/Z (*i.e.*, fraction of Cs), the difference $G(k) = T^+(k) - T^-(k)$ among the probability to increase ($T^+(k)$) and decrease ($T^-(k)$) the number of cooperators (see Methods) by one. Whenever $G(k) > 0$ the fraction of Cs is likely to increase; whenever $G(k) < 0$ the opposite is expected to happen. The stationary distributions show how likely it is to find the population in each (discrete) configuration of our system. The panels on the left-hand side show the results obtained for the **CRD** under pure-Reward; on the right-hand side, we show the results obtained for pure-Punishment.

Naturally, both mechanisms are inoperative whenever the per-capita incentives are inexistent ($\delta = 0$), creating a natural reference scenario in which to study the impact of Reward and Punishment on the **CRD**. In this case, above a certain value of risk (r), decision-making is characterized by two internal equilibria (*i.e.*, adjacent finite population states with opposite gradient sign, representing the analogue of fixed points in a dynamical system characterizing evolution in infinite populations). Above a certain fraction of cooperators the population overcomes the coordination barrier and naturally self-organizes towards a stable co-existence of cooperators and defectors. Otherwise, the population is condemned to evolve towards a monomorphic population of defectors, leading to the tragedy of the commons[9]. As the budget for incentives increases, using either Reward or Punishment leads to very different outcomes, as depicted in Fig. 2.

Contrary to the case of linear Public Goods Games[12], in the **CRD** coordination and co-existence dynamics already exist in the absence of any reward/punishment incentive. Reward is particularly effective when
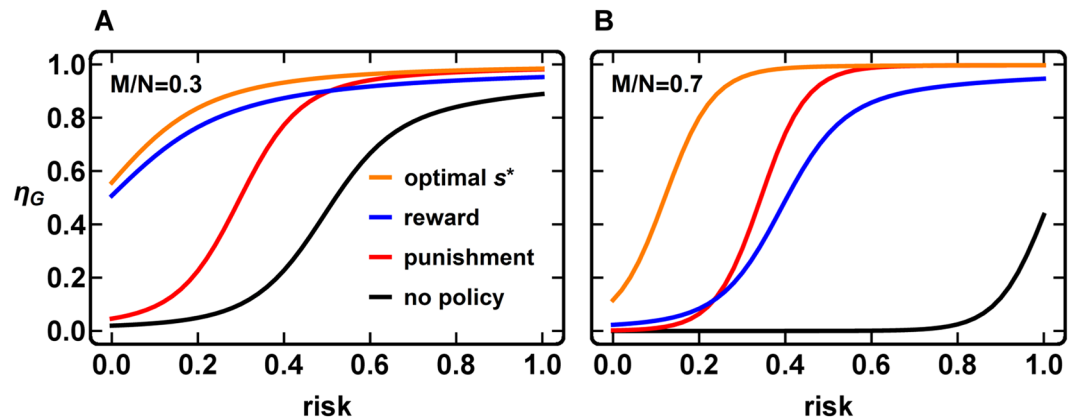
Figure 1. Average group achievement η_G as a function of risk. Left: Group relative threshold M/N = 3/10. Right: Group relative threshold M/N = 7/10. In both panels, the black line corresponds to a reference scenario where no policy is applied. The red line shows η_G in the case where all available budget is applied to pure-Punishment ($w = 0$), whereas the blue line shows results for pure-Reward ($w = 1$). Pure-Reward is most effective at low risk values, while pure-Punishment is marginally the most effective policy at high risk. These features are more pronounced for low relative thresholds (left panel), and only at high thresholds does pure-Punishment lead to a sizeable improvement with respect to pure-Reward. Finally, the orange line shows the results using the combination of Reward and Punishment, leading (naturally) to the best results. In this case, we adopt pure-Reward ($w = 1$) when there are few cooperators and, above a certain critical point $k/Z = s = 0.5$, we switch to pure-Punishment ($w = 0$). As detailed in the main text (see Fig. 3 and Methods), $s = 0.5$ provides the optimal switching point s^* for cooperation to thrive. Other parameters: Population size $Z = 50$, group size $N = 10$, cost of cooperation $c = 0.1$, initial endowment $B = 1$, budget $\delta = 0.025$, reward efficiency $a = 1$, punishment efficiency $b = 1$, intensity of selection $\beta = 5$, mutation rate $\mu = 0.01$.

cooperation is low (small k/Z), showing a significant impact on the location of the finite population analogue of an unstable fixed point. Indeed, increasing δ lowers the minimum number of cooperators required to reach the cooperative basin of attraction (as well as increasing the prevalence of cooperators in co-existence point on the right), which ultimately disappears for high δ (Fig. 2A). This means that a smaller coordination effort is required before the population dynamics start to naturally favor the increase of cooperators. Once this initial barrier is surpassed, the population will naturally tend towards an equilibrium state, which does not improve appreciably under Reward. The opposite happens under Punishment. The location of the coordination point is little affected, yet once this barrier is overcome, the population will evolve towards a more favorable equilibrium (Fig. 2B). Thus, while Reward seems to be particularly effective to bootstrap cooperation towards a more cooperative *basin of attraction*, Punishment seems effective in sustaining high levels of cooperation.

As a consequence, the most frequently observed configurations are very different when using each of the policies. As shown by the stationary distributions (Fig. 2C,D), under Reward the population visits more often states with intermediate values of cooperation (*i.e.*, where *C*s and *D*s co-exist). Intuitively, this happens because the coordination effort is eased by the rewards, causing the population to effectively overcome it and reach the coexistence point (the equilibrium state with an intermediate amount of cooperators) thus spending most of the time near it. On the other hand, Punishment will not ease the coordination effort, and thus the population will spend most of the time in states of low cooperation, failing to overcome this barrier. Notwithstanding, once surpassed, the population will stabilize on higher states of cooperation. This is especially evident for high budgets, as shown with $\delta = 0.02$ (blue line). Moreover, since $N\delta$ corresponds to a fixed total amount which is distributed by the existing cooperators/defectors, this causes the per-cooperator/defector budget to vary depending on the number of existing cooperators/defectors (*i.e.*, each of the j cooperators receives $w\delta N/j$ and each defector loses $(1-w)\delta N/(N-j)$). In other words, positive (negative) incentives become very profitable (or severe) if defection (cooperation) prevails within a group. In particular, whenever the budget is significant (see, *e.g.*, $\delta = 0.02$ in Fig. 2) the punishment becomes so high when there are few defectors within a group, that a new equilibrium emerges close to full cooperation.

The results in Fig. 2 show that Reward can be instrumental in fostering pro-social behavior, while Punishment can be used for its maintenance. This suggests that, to combine both policies synergistically, pure-Reward ($w = 1$) should be applied at first, when there are few cooperators (low k/Z); above a certain critical point ($k/Z = s$) one should switch to pure-Punishment ($w = 0$). In the Methods section, we demonstrate that, similar to linear Public Goods Games[12], in **CRD**s this is indeed the policy which minimizes the advantage of the defector, even if we consider the alternative possibility of applying both policies simultaneously. In Methods, we also compute a general expression for the optimal switching point s^*, that is, the value of k above which Punishment should be applied instead of Reward to maximize cooperation and group achievement. By using such policy — that we denote by s^* — we obtain the best results shown with an orange line in Fig. 1. We propose, however, to explore what happens in the context of a **CRD** when s^* is not used. How much cooperation is lost when we deviate from s^* to either of the pure policies, or to a policy which uses a switching point different from the optimal one?
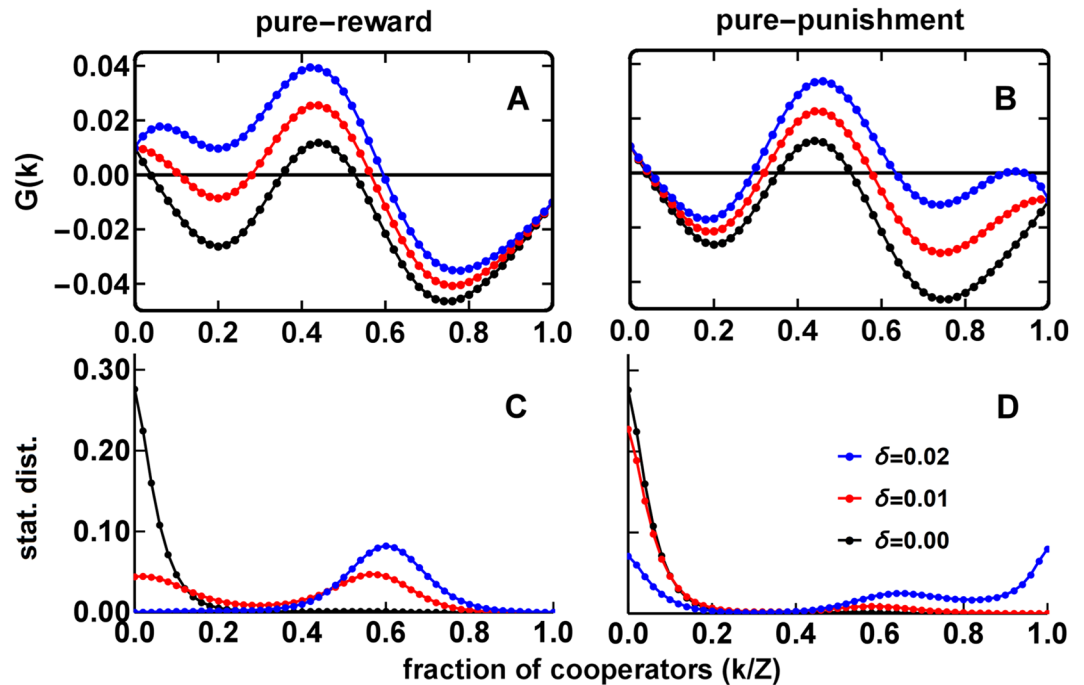
Figure 2. Gradient of selection (top panels, A and B) and stationary distribution (bottom panels, C and D) for the different values of per-capita budget δ indicated, using either pure-Reward ($w = 1$, left panels) or pure-Punishment ($w = 0$, right panels). The black curve is equal on the left and right panels, since in this case $\delta = 0$. As δ increases, the behaviour under Reward and Punishment is qualitatively similar, by displacing the (unstable) coordination equilibrium towards lower values of k/Z, while displacing the (stable) coexistence equilibrium towards higher values of k/Z. This happens, however, only for low values of δ. Indeed, by further increasing δ one observes very different behaviours under Reward and Punishment: Whereas under Punishment the equilibria are further moved apart (in accord with what happened for low δ) under Reward the coordination equilibrium disappears, and the overall dynamics becomes characterized by a single coexistence equilibrium which consistently shifts towards higher values of k/Z with increasing δ. This difference in behaviour, in turn, has a dramatic impact in the overall prevalence of configurations achieved by the population dynamics, as shown by the stationary distributions: On panel C (pure-Reward) the population spends most of the time on intermediate states of cooperation. On panel D (pure-Punishment) the population spends most of the time on both extremes (high and low cooperation) but especially on low cooperation states. Other parameters: $Z = 50$, $N = 10$, $M = 5$, $c = 0.1$, $B = 1$, $r = 0.5$, $a = b = 1$, $\beta = 5$ and $\mu = 0.01$ (see Methods for details).

Figure 3 illustrates how the choice of the switching point s impacts the overall cooperation, as evaluated by η_G, for different values of risk. For a switching point of $s = k/Z = 1.0$ (0.0) a static policy of always pure-Reward (pure-Punishment) is used. This can be seen on the far right (left) of Fig. 3. Figure 3 suggests that, for low thresholds, an optimal policy switching (which, for the parameters shown, occurs for $s = 50\%$, see Methods) is only marginally better than a policy solely based on rewards ($s = 1$). Figure 3 also allows for a comparison of what happens when the switching point occurs too late (excessive rewards) or too early (excessive sanctions) in a low-threshold scenario. A late switch is significantly less harmful than an early one. In other words, our results suggest that when the population configuration cannot be precisely observed, it is preferable to keep rewarding for longer. This said, whenever the perception of risk is high (an unlikely situation these days) an early switch is slightly less harmful than a late one. In the most difficult scenarios, where stringent coordination requirements (large M) are combined with a low perception of risk (low r), the adoption of a combined policy becomes necessary (see right panel of Fig. 1).

Discussion

One might expect the impact of Reward and Punishment to lead to symmetric outcomes – Punishment would be effective for high-cooperation the same way that Reward is effective for low-cooperation. In low-cooperation scenarios (under low risk, threshold or budget) Reward alone plays the most important role. However, in the opposite scenario, Punishment alone does not have the same impact. Either a favourable scenario occurs, where any policy yields a satisfying result, or Punishment cannot improve outcomes on its own. In the latter case, the synergy between both policies becomes essential to achieve cooperation. Such optimal policy involves a combination of the single policies, Reward and Punishment, which is dynamic, in the sense that the combination does not remain the same for all configurations of the population. It corresponds to employing pure Reward at first, when cooperation is low, switching subsequently to Punishment whenever a pre-determined level of cooperation is reached.
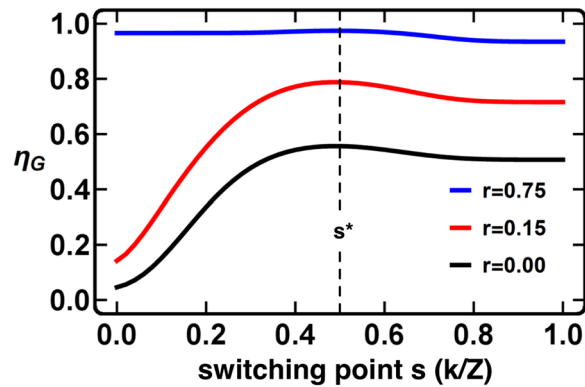
Figure 3. Average group achievement η_G as a function of the location of the switching point s. The switching point s corresponds to the configuration (fraction of **C**s in the population, k/Z) above which w suddenly switches from pure-Reward ($w = 1$) to pure-Punishment ($w = 0$). Assuming both policies are equally efficient, the optimal switching point occurs at 50% of cooperators ($k/Z = 0.5$). The far-left values of s correspond to a static policy of always pure-Punishment – the switch from pure-Reward to pure-Punishment occurs immediately at 0% of cooperators. On the far-right (switching point = 100%) a pure-Reward policy is depicted. We can also see what happens when the switch occurs too late or too early, for different values of risk. For low values of risk, it is significantly less harmful to have a late switch from Reward to Punishment than an early one, meaning that when the population configuration cannot be precisely observed, it is preferable to keep rewarding for longer. See Methods for the calculation of the optimal switching point (s^*) that maximizes cooperation fitness relative to defection – and consequent group achievement. Other parameters: $Z = 50$, $B = 1$, $\mu = 0.01$, $\beta = 5$, $N = 10$, $M = 3$, $c = 0.1$, and $\delta = 0.025$.

The optimal procedure, however, is unlikely to be realistic in the context of Climate Change agreements. Indeed, and unlike other Public Goods Dilemmas, where Reward and Punishment constitute the main policies available for Institutions to foster cooperative collective action, in International Agreements it is widely recognized that Punishment is very difficult to implement[2,42]. This has been, in fact, one of the main criticisms put forward in connection with Global Agreements on Climate Mitigation: They suffer from the lack of sanctioning mechanisms as it is practically impossible to enforce any type of sanctioning at a Global level. In this sense, the results obtained here by means of our dynamical, evolutionary approach, are gratifying, given these a-priori limitations of sanctioning in **CRD**s. Not only do we show that Reward is essential to foster cooperation, mostly when both the perception of risk is low and the overall number of engaged parties is small (low k/Z), but also we show that Punishment mostly acts to sustain cooperation, after it has been installed. Given that low-risk scenarios are more common and harmful to cooperation than high-risk ones, our results in connection with rewards provide a viable way to explore in the quest for establishing Global cooperative collective action. Reward policies may also be very relevant in scenarios where Climate Agreements are coupled with other International agreements from which parties are not interested to deviate from[2,42]. Finally, the fact that rewards ease coordination towards cooperative states suggests that positive incentives should also be used within intervention mechanisms aiming at fostering pro-sociality in artificial systems and hybrid populations comprising humans and machines[46–49].

The model used takes for granted the existence of an institution with a budget available to implement either Reward or Punishment. New behaviours may emerge once individuals are called to decide whether or not to contribute to such an institution, allowing for a scenario where this institution fails to exist[10,28,50,51]. At present, and under the Paris agreement, we are witnessing the potential birth of an informal funding institution, whose goal is to finance developing countries to help them increase their mitigation capacity. Clearly, this is just an example pointing out to the fact that the prevalence of local and global institutional incentives may depend and may be influenced by the distribution of wealth available among parties, in the same way that it influences the actual contributions to the public good[10,29,33]. Finally, several other effects may further influence and/or affect the present results. Among others, if intermediate tasks are considered[33], or if individuals have the opportunity to pledge their contribution before their actual action[7,40,52], it is likely that pro-social behavior may be enhanced. Work along these lines is in progress.

Methods

Public goods and collective risks.
Let us consider a population with Z individuals, where each individual can be a cooperator (**C**) or a defector (**D**). For each round of this game, a group of N players is sampled from the original finite population of size Z, which corresponds to a process of sampling without replacement. The probability of a group comprising any possible combination of **C**s and **D**s is given by the hypergeometric distribution. In the context of a given group, a strategy is associated with a payoff value corresponding to an individual's earnings in that round, which depend on the action of the rest of group. Fitness is the expected payoff of an individual in a population, before knowing to which group he was assigned. This way, for a population with k out of Z **C**s and each group containing j out of N **C**s, the fitness of a **D** and a **C** can be written as:

$$f_D = \binom{Z-1}{N-1}^{-1} \sum_{j=0}^{N-1} \binom{k}{j} \binom{Z-k-1}{N-j-1} \Pi_D(j)$$

(1)

$$f_C = \binom{Z-1}{N-1}^{-1} \sum_{j=0}^{N-1} \binom{k-1}{j} \binom{Z-k}{N-j-1} \Pi_C(j+1)$$

(2)

where $\Pi_C(j)$ and $\Pi_D(j)$ stand for the payoff or a C and a D in a single round, in a group with N players and j Cs. To define the payoff functions, let $\theta(x)$ be a Heaviside step-function distribution, where $\theta(x) = 0$ if $x < 0$ and $\theta(x) = 1$ if $x \geq 0$. Each player can contribute with a fraction c of her endowment B (with $0 \leq c \leq 1$), and in case a group contains less than M cooperators ($0 < M \leq N$) there is a risk r of failure ($0 \leq r \leq 1$), in which case no player obtains her remaining endowment. The payoff of a defector ($\Pi_D(j)$) and the payoff of a cooperator ($\Pi_C(j)$), before incorporating any policy, can be written as[9]:

$$\Pi_D(j) = B\{\theta(j-M) + (1-r)[1-\theta(j-M)]\}$$

(3)

$$\Pi_C(j) = \Pi_D(j) - cB$$

(4)

Reward and punishment. To include a Reward or a Punishment policy, let us follow ref.[12] and consider a group budget $N \cdot \delta$ which can be used to implement any type of policy. The fraction of $N \cdot \delta$ applied to Reward is represented by the weight w, with $0 \leq w \leq 1$. Parameters a and b correspond to the efficiency of Reward and Punishment (for all Figures above it was assumed that $a = b = 1$).

$$\Pi_D^P(j) = \Pi_D(j) - \frac{b(1-w)N\delta}{N-j}$$

(5)

$$\Pi_C^R(j) = \Pi_C(j) + \frac{awN\delta}{j}$$

(6)

Naturally, these new payoff functions can be included into the previous fitness functions (Π_D^P replaces Π_D and Π_C^R replaces Π_C), letting fitness values account for the different policies.

Evolutionary dynamics in finite populations. The fitness functions written above allow us to setup the (discrete time) evolutionary dynamics. Indeed, the configurations of the entire population may be used to define a Markov Chain, where each state is characterized by number of cooperators[9,44]. To decide in which direction the system will evolve, at each step a player i and a neighbour j of her are drawn at random from the population. Player i decides whether to imitate her neighbour j with a probability depending on the difference between their fitness[43,44]. This way, a system with k cooperators may stay in the same state, switch to $k-1$ or to $k+1$. The probability of player i imitating player j can be given by the Fermi function:

$$p_{j,i}(k) \equiv [1 + e^{-\beta(f_j - f_i)}]^{-1}$$

(7)

where β is the intensity of selection. Using this probability distribution, we can fully characterize this Markov process. Let k be the total number of cooperators in the population and Z the total size of the population. $T^+(k)$ and $T^-(k)$ are the probabilities to increase and decrease k by one, respectively[44]:

$$T^{\pm}(k) = \frac{k}{Z} \frac{Z-k}{Z} [1 + e^{\mp\beta[f_C(k) - f_D(k)]}]^{-1}$$

(8)

The most likely direction can be computed using the difference $G(k) \equiv T^+(k) - T^-(k)$. A mutation rate can be introduced by using transition probabilities $T_\mu^+(k) = (1-\mu)T^+(k) + \mu\frac{Z-k}{Z}$ and $T_\mu^-(k) = (1-\mu)T^-(k) + \mu\frac{k}{Z}$. In all cases we used a mutation rate $\mu = 0.01$, this way avoiding the population to fixate in a monomorphic configuration. In this context, the stationary distribution becomes a very useful tool to analyse the overall population dynamics, providing the probability $\overline{p}_k = P(\frac{k}{Z})$ for each of the $Z+1$ states of this Markov Chain to be occupied[53,54]. For each given population state k, the hypergeometric distribution can be used to compute the average fraction of groups that obtain success $-a_G(k)$. Using the stationary distribution and the average group success, the average group achievement (η_G) can then be computed, providing the overall probability of achieving success: $\eta_G = \sum_{k=0}^{Z} \overline{p}_k a_G(k)$.

Combined policies. By allowing the weight w to depend on the frequency of cooperators, we can derive the optimal switching point s^* between positive and negative incentives by minimizing the defector's advantage ($f_D - f_C$). This is done similarly to ref.[12], but using finite populations and therefore a hypergeometric distribution (see Eqs (1), (2), (5), and (6)), to account for sampling without replacement. From Eqs (1) and (2), we get
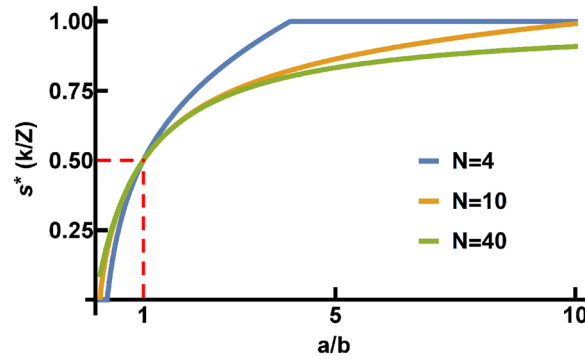
Figure 4. Optimal switching point s^* as a function of the ratio a/b, for different values of N (see Methods).

$$f_D = \sum_{j=0}^{N-1} \frac{\binom{k}{j}\binom{Z-1-k}{N-1-j}}{\binom{Z-1}{N-1}}\left(\Pi_D(j) - \frac{b(1-w)N\delta}{N-j}\right)$$

$$f_C = \sum_{j=0}^{N-1} \frac{\binom{k-1}{j}\binom{Z-1-(k-1)}{N-1-j}}{\binom{Z-1}{N-1}}(\Pi_C(j+1) - c + \frac{awN\delta}{j+1})$$

from which we aim at finding the value of w (with respect to k) that minimizes $F' = f_D - f_C$. Since $\Pi_D(j), \Pi_C(j+1)$ and c do not depend on w, these quantities do not affect the choice of the optimal w, leaving us with the problem of minimizing the following expression:

$$F' = -N\delta\sum_{j=0}^{N-1} \frac{\binom{k}{j}\binom{Z-1-k}{N-1-j}}{\binom{Z-1}{N-1}}\left(\frac{b(1-w)}{N-j}\right) - N\delta\sum_{j=0}^{N-1} \frac{\binom{k-1}{j}\binom{Z-k}{N-1-j}}{\binom{Z-1}{N-1}}\left(\frac{aw}{j+1}\right)$$

Since $\binom{k}{j} = \binom{k-1}{j}\frac{k}{k-j}$ and $\binom{Z-1-k}{N-1-j} = \binom{Z-k}{N-1-j}\frac{Z-k-(N-1-j)}{Z-k}$,

$$F' = -N\delta\sum_{j=0}^{N-1} \frac{\binom{k-1}{j}\binom{Z-k}{N-1-j}}{\binom{Z-1}{N-1}}\left(\frac{aw}{j+1} + \frac{b(1-w)}{N-j}\frac{k}{k-j}\frac{Z-k-N+1+j}{Z-k}\right)$$

$$= -N\delta\sum_{j=0}^{N-1} \frac{\binom{k-1}{j}\binom{Z-k}{N-1-j}}{\binom{Z-1}{N-1}}\left(w\left[\frac{a}{j+1} - \frac{b}{N-j}\frac{k}{k-j}\frac{Z-k-N+1+j}{Z-k}\right]\right)$$

$$-N\delta\sum_{j=0}^{N-1} \frac{\binom{k-1}{j}\binom{Z-k}{N-1-j}}{\binom{Z-1}{N-1}}\frac{b}{N-j}\frac{k}{k-j}\frac{Z-k-N+1+j}{Z-k}$$

The second summation does not depend on w; thus the optimal policy is given by the minimization of:

$$F'' = -N\delta\sum_{j=0}^{N-1} \frac{\binom{k-1}{j}\binom{Z-k}{N-1-j}}{\binom{Z-1}{N-1}}\left(w\left[\frac{a}{j+1} - \frac{b}{N-j}\frac{k}{k-j}\frac{Z-k-N+1+j}{Z-k}\right]\right)$$

Since N and δ are always positive, the whole expression can be divided by $N\delta$ without changing the optimization problem. Moreover, by multiplying the expression by (-1), it can finally be shown that minimizing $f_D - f_C$ is equivalent to maximizing the following expression:

7

$$w\sum_{j=0}^{N-1}\frac{\binom{k-1}{j}\binom{Z-k}{N-1-j}}{\binom{Z-1}{N-1}}\left(\left[\frac{a}{j+1}-\frac{b}{N-j}\frac{k}{k-j}\frac{Z-k-N+1+j}{Z-k}\right]\right)$$

where j represents the number of Cs in a group of size N, sampled without replacement from a population of size Z containing k Cs. Now, let us consider that the optimal switching point $s*$ depends on k. Since this sum decreases as k increases, containing only one root, the solution to this optimization problem corresponds to having w set to 1 (pure Reward) for positive values of the sum, suddenly switching to $w = 0$ (pure Punishment) once the sum becomes negative. The optimal switching point $s*$ depends on the ratio $\frac{a}{b}$, group size N and population size Z. The effect of population size (Z) and group size (N) on $s*$ is limited, while the impact of the efficiency of reward (a) and punishment (b) is illustrated in Fig. 4. For $\frac{a}{b} = 1$ the switching point is $s* = 0.5$ (see Fig. 4). Interestingly, we note that, also in the **CRD**, $s*$ is not impacted by the group success threshold (M) or the risk associated with losing the retained endowment when collective success is not attained (r). This is the case as we assume that the decision to punish or reward is independent on M or r. Notwithstanding, the model that we present can, in the future, be tuned to test more sophisticated incentive tools, such as rewarding or punishing depending on (*i*) how far group contributions remained from (or surpassed) the minima to achieve group success or (*ii*) how soft/strict is the dilemma at stake, given the likelihood of losing everything when collective success is not accomplished.

References

1. Barrett, S. Self-enforcing international environmental agreements. *Oxford Economic Papers* **46**, 878–894 (1994).
2. Barrett, S. Why cooperate?: the incentive to supply global public goods. (Oxford UP, 2007).
3. Dreber, A. & Nowak, M. A. Gambling for Global Goods. *Proc Natl Acad Sci USA* **105**, 2261–2262 (2008).
4. Hardin, G. The Tragedy of the Commons. *Science* **162**, 1243 (1968).
5. Milinski, M., Sommerfeld, R. D., Krambeck, H. J., Reed, F. A. & Marotzke, J. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc Natl Acad Sci USA* **105**, 2291–2294 (2008).
6. Pacheco, J. M., Santos, F. C., Souza, M. O. & Skyrms, B. Evolutionary dynamics of collective action in N-person stag hunt dilemmas. *Proc R Soc Lond B* **276**, 315–321 (2009).
7. Tavoni, A., Dannenberg, A., Kallis, G. & Löschel, A. Inequality, communication and the avoidance of disastrous climate change in a public goods game. *Proc Natl Acad Sci USA* **108**, 11825–11829 (2011).
8. Bosetti, V., Heugues, M. & Tavoni, A. Luring others into climate action: coalition formation games with threshold and spillover effects. *Oxford Economic Papers* **69**, 410–431 (2017).
9. Santos, F. C. & Pacheco, J. M. Risk of collective failure provides an escape from the tragedy of the commons. *Proc Natl Acad Sci USA* **108**, 10421–10425 (2011).
10. Vasconcelos, V. V., Santos, F. C. & Pacheco, J. M. A bottom-up institutional approach to cooperative governance of risky commons. *Nat. Clim. Change* **3**, 797–801 (2013).
11. Sigmund, K., Hauert, C. & Nowak, M. A. Reward and punishment. *Proc. Natl. Acad. Sci. USA* **98**, 10757–10762 (2001).
12. Chen, X., Sasaki, T., Brännström, Å. & Dieckmann, U. First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation. *Journal of The Royal Society Interface* **12**, 20140935 (2015).
13. Hilbe, C. & Sigmund, K. Incentives and opportunism: from the carrot to the stick. *Proceedings of the Royal Society of London B: Biological Sciences* **277**, 2427–2433 (2010).
14. Gneezy, A. & Fessler, D. M. Conflict, sticks and carrots: war increases prosocial punishments and rewards. *Proceedings of the Royal Society of London B: Biological Sciences*, rspb20110805 (2011).
15. Sasaki, T. & Uchida, S. Rewards and the evolution of cooperation in public good games. *Biology letters* **10**, 20130903 (2014).
16. Fehr, E. & Gächter, S. Altruistic punishment in humans. *Nature* **415**, 137–140 (2002).
17. Sigmund, K. Punish or perish? Retaliation and collaboration among humans. *Trends in ecology & evolution* **22**, 593–600 (2007).
18. Masclet, D., Noussair, C., Tucker, S. & Villeval, M.-C. Monetary and nonmonetary punishment in the voluntary contributions mechanism. *Am. Econ. Rev.* **93**, 366–380 (2003).
19. Charness, G. & Haruvy, E. Altruism, equity, and reciprocity in a gift-exchange experiment: an encompassing approach. *Games and Economic Behavior* **40**, 203–231 (2002).
20. Andreoni, J., Harbaugh, W. & Vesterlund, L. The carrot or the stick: Rewards, punishments, and cooperation. *The American economic review* **93**, 893–902 (2003).
21. Szolnoki, A. & Perc, M. Reward and cooperation in the spatial public goods game. *EPL (Europhysics Letters)* **92**, 38003 (2010).
22. Perc, M. *et al.* Statistical physics of human cooperation. *Physics Reports* **687**, 1–51 (2017).
23. Fang, Y., Benko, T. P., Perc, M., Xu, H. & Tan, Q. Synergistic third-party rewarding and punishment in the public goods game. *Proc. Roy. Soc. A* **475**, 20190349 (2019).
24. Milinski, M., Semmann, D., Krambeck, H. J. & Marotzke, J. Stabilizing the Earth's climate is not a losing game: Supporting evidence from public goods experiments. *Proc Natl Acad Sci USA* **103**, 3994–3998 (2006).
25. Chen, X., Szolnoki, A. & Perc, M. Averting group failures in collective-risk social dilemmas. *EPL (Europhysics Letters)* **99**, 68003 (2012).
26. Chakra, M. A. & Traulsen, A. Evolutionary dynamics of strategic behavior in a collective-risk dilemma. *PLoS Comput Biol* **8**, e1002652 (2012).
27. Chen, X., Szolnoki, A. & Perc, M. Risk-driven migration and the collective-risk social dilemma. *Physical Review E* **86**, 036101 (2012).
28. Pacheco, J. M., Vasconcelos, V. V. & Santos, F. C. Climate change governance, cooperation and self-organization. *Phys Life Rev* **11**, 595–597 (2014).
29. Vasconcelos, V. V., Santos, F. C., Pacheco, J. M. & Levin, S. A. Climate policies under wealth inequality. *Proc Natl Acad Sci USA* **111**, 2212–2216 (2014).
30. Hilbe, C., Chakra, M. A., Altrock, P. M. & Traulsen, A. The evolution of strategic timing in collective-risk dilemmas. *PLoS ONE* **8**, e66490 (2013).
31. Barrett, S. Avoiding disastrous climate change is possible but not inevitable. *Proc Natl Acad Sci USA* **108**, 11733 (2011).
32. Barrett, S. & Dannenberg, A. Climate negotiations under scientific uncertainty. *Proc Natl Acad Sci USA* **109**, 17372–17376 (2012).

33. Milinski, M., Röhl, T. & Marotzke, J. Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Climatic change* 1–8 (2011).
34. Boesch, C. Cooperative hunting roles among Tai chimpanzees. Human. *Nature* **13**, 27–46 (2002).
35. Creel, S. & Creel, N. M. Communal hunting and pack size in African wild dogs, Lycaon pictus. *Animal Behaviour* **50**, 1325–1339 (1995).
36. Black, J., Levi, M. D. & De Meza, D. Creating a good atmosphere: minimum participation for tackling the 'greenhouse effect'. *Economica* 281–293 (1993).
37. Stander, P. E. Cooperative hunting in lions: the role of the individual. *Behavioral ecology and sociobiology* **29**, 445–454 (1992).
38. Alvard, M. S. *et al.* Rousseau's whale hunt? Coordination among big-game hunters. *Current anthropology* **43**, 533–559 (2002).
39. Souza, M. O., Pacheco, J. M. & Santos, F. C. Evolution of cooperation under N-person snowdrift games. *J Theor Biol* **260**, 581–588 (2009).
40. Pacheco, J. M., Vasconcelos, V. V., Santos, F. C. & Skyrms, B. Co-evolutionary dynamics of collective action with signaling for a quorum. *PLoS Comput Biol* **11**, e1004101 (2015).
41. Skyrms, B. *The Stag Hunt and the Evolution of Social Structure*. (Cambridge Univ Press, 2004).
42. Barrett, S. *Environment and statecraft: the strategy of environmental treaty-making*. (Oxford UP, 2005).
43. Sigmund, K. *The Calculus of Selfishness*. (Princeton Univ Press, 2010).
44. Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74**, 011909 (2006).
45. Traulsen, A., Hauert, C., De Silva, H., Nowak, M. A. & Sigmund, K. Exploration dynamics in evolutionary games. *PNAS* **106**, 709–712 (2009).
46. Paiva, A., Santos, F. P. & Santos, F. C. Engineering pro-sociality with autonomous agents in *Thirty-Second AAAI Conference on Artificial Intelligence*, pp. 7994–7999 (2018).
47. Shirado, H. & Christakis, N. A. Locally noisy autonomous agents improve global human coordination in network experiments. *Nature* **545**, 370 (2017).
48. Santos, F. P., Pacheco, J. M., Paiva, A. & Santos, F. C. Evolution of collective fairness in hybrid populations of humans and agents in *Proceedings of the Thirty-Third AAAI Conference on Artificial Intelligence*, Vol. 33, pp. 6146–6153 (2019).
49. Rahwan, I. *et al.* Machine behaviour. *Nature* **568**, 477 (2019).
50. Powers, S. T., van Schaik, C. P. & Lehmann, L. How institutions shaped the last major evolutionary transition to large-scale human societies. *Philosophical Transactions of the Royal Society B: Biological Sciences* **371**, 20150098 (2016).
51. Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **466**, 861 (2010).
52. Santos, F. C., Pacheco, J. M. & Skyrms, B. Co-evolution of pre-play signaling and cooperation. *J Theor Biol* **274**, 30–35 (2011).
53. Kulkarni, V. G. Modeling and analysis of stochastic systems. (Chapman and Hall/CRC, 2016).
54. Hindersin, L., Wu, B., Traulsen, A. & García, J. Computation and simulation of evolutionary game dynamics in finite populations. *Sci. Rep.* **9**, 6946 (2019).

Acknowledgements

Author contributions

A.R.G., F.P.S, J.M.P. and F.C.S. designed and implemented the research; A.R.G., F.P.S, J.M.P. and F.C.S prepared all the Figures; A.R.G., F.P.S, J.M.P. and F.C.S. wrote the manuscript; A.R.G., F.P.S, J.M.P. and F.C.S reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Correspondence and requests for materials should be addressed to F.C.S.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.