Evolution of Cooperation through Graduated Punishment

Marta Gomes da Cunha Couto
martacouto@tecnico.ulisboa.pt

Instituto Superior Técnico, Lisboa, Portugal

November 2018

Abstract

Understanding the mechanisms that promote and maintain cooperative behavior is widely recognized as a major theoretical problem. Knowing how human beings should cooperate may help us to address real global and complex issues, like environmental protection, which demand a collective commitment. The present study focuses on the role of graduated punishment conducted by self-organized institutions in the emergence of cooperation in social dilemmas. Previous studies analyzed the effect of strict (unconditional) punishment in Public Goods Games and Collective Risk Dilemmas. Oppositely, graduated punishment consists of a sanction whose severity is gradually adjusted to the collective returns. This type of sanctioning system is a design principle empirically observed in long-enduring common-pool resource institutions. Using Game Theory, Evolutionary Dynamics and Stochastic Processes, we propose a new model of evolution in finite populations, where individuals engage in N-person games with three possible strategies — cooperator, defector, and (graduated) punisher. We conclude that graduated punishment is better at promoting and preserving cooperation than a strict form of punishment. This improvement is enhanced if the costs required to maintain the external sanctioning institution are also graduated. Plus, local institutions are more effective than a global one. Finally, if rewards are considered instead of punishment, we show that they should be graduated as well.

Keywords: evolution of cooperation, collective risk dilemma, graduated punishment, environmental management

1. Introduction

One of the biggest mysteries about human behavior is *cooperation* [1]. Perhaps surprisingly to some, Physics may help solve this problem [2, 3], if we look at social systems as complex systems, where global patterns and collective phenomena are more than the sum of individual contributions.

Cooperation is a concept that sounds quite familiar to us human beings. Nature and human life show us every day that cooperation is in fact important. However, if we think thoroughly, the idea of offering assistance to others is rather odd from a purely rational point of view. Why would one lose something (like time, money or energy) to cooperate with a stranger? Or why should we pay taxes if we know that many don't but still benefit from the common goods? Why do we care to do well if that act may represent a cost to ourselves? These questions reveal the mystery associated with cooperation [1], recently identified by Science's invited panel of scientists as one of the major scientific challenges of our century [4].

We see that cooperation is widespread in Nature at all scales and levels of complexity. Yet, we also know that *Darwin's natural selection* is about competition. In a nutshell, individuals are in conflict (for instance, to get food or to mate), the stronger ones survive and, as a consequence, their genes are preserved in the following generations. This is the so-called *survival of the fittest*. Natural selection favors the selfish and strong even though cooperative interactions prevail in living systems. Cooperation has been identified as the third building block of evolution, next to selection and mutation, working from the level of cells to large societies [5].

The present challenge is to understand the mechanisms that enabled the emergence of cooperation over time and hopefully apply the insights attained to human endeavors where cooperation is not achieved as, for instance, the present worldwide problem of the climate change and environmental protection. It is thereby fundamental to understand the contexts that prevent selfishness and conflict while allowing pro-sociality to be sustained (or induced, when absent).

Here, we study the role of *punishment* in the context of a *Collective Risk Dilemma* (CRD) [6]. Punishment should be regarded as a costly tool — applying sanctions to cheaters bears a cost since it requires monitorization and the act of punish-

ing itself. There must be agents responsible for those actions, either an external institution or peers (pool- and peer-punishment, respectively). It has been shown that *strict* punishment helps increasing cooperation in some scenarios [7]. However, as it represents a relatively high cost to one faction, a second-order free-riding problem arises and the sanctioning institution may not be sustained. To overcome this problem, we propose *graduated* punishment as an alternative to strict.

2. Background

Before addressing our specific problem, it is useful to describe the theoretical apparatus which has been used, important models and also some of the state of the art main results.

2.1. Evolutionary Game Theory

Cooperation is an interaction involving at least two individuals where an individual is ready to pay a cost for others to have a benefit [8]. This can be conveniently formalized resorting to the mathematics of *Game Theory*, which is the formal way to quantitatively describe situations of conflict of interests.

A game or interaction is defined by a set of *players*, the options or *strategies* available to the players and a *payoff rule* which determines the gain (or payoff) of each player after the game. Note that the players don't know each other's choices beforehand and the payoffs depend on everyone's choices. Many dilemmas involve simultaneous decisions of several individuals. In a classical N-player interaction called a *Public Goods Game* (PGG), each individual chooses how much to contribute to a common pot and how much to save; then the amount contributed is increased by some factor and equally redistributed. Here, an individual is tempted to defect (not to contribute) because he or she will get the common benefit without any cost. However, if no one contributes there is no benefit at all, a situation often called the *Tragedy of the Commons* [9]. Therefore, the optimal choice for an individual is to defect, whereas for the group of players (socially speaking) is to cooperate. When there is this kind of antagonism between strategies — one which is good for the population but not the best for the individual - we are in the presence of a *social dilemma* [8].

To study the evolution of cooperation in large populations we may rely on *Evolutionary Dynamics*. As already mentioned, natural selection means that the fittest agents will out-compete others: the weak die and the strong prevail. Interestingly, dynamics of peer-influence can be formally equivalent to natural selection, in the sense that strategies that provide better payoffs will be imitated and will spread (*social learning*). That means that fitness may be related not only with a genetic background but also with cultural traits or chosen strategies. This way, we can use evolutionary dynamics to describe and model populations of players — individuals provided with strategies that play with each other. For that, we need a selection rule or, better said in this context, an imitation rule, which takes into account the strategies' payoffs and how sensitive to imitation the agents are. This combination of game theory and evolutionary dynamics is called *Evolutionary Game Theory* (EGT).
Importantly, once we wish to look into finite populations, some randomness is introduced in the systems. The adoption of new strategies by individuals can be described as a stochastic process, where errors and random exploration of strategies are explicitly considered.

2.2. Collective Risk Dilemma

The main purpose of studying cooperation is to explore its grounds, in which conditions it emerges and endures. There are several known mechanisms that promote cooperation [8], but let us focus on risk.

A *Collective Risk Dilemma* (CRD) is an N-person game where there is a threshold in the number of cooperators, such that if this number is not attained or surpassed, everyone will lose their earnings, with a probability r, which is called the risk. Here, we can say that people cooperate so that they don't lose what they have, contrary to the PGG, where players immediately get a positive payoff after each round. It is proved analytically [6] and experimentally [10] that the awareness of the risk increases the number of cooperators. Despite being quite simple, this formulation might mimic worldwide conventions on climate governance, where agents (for example, countries) cooperate (or not) for the environment protection and the risk of failure is not to be disregarded. At the present time, the perception of risk concerning climate matters is low [11], which in certain conditions won't be sufficient to keep cooperation. We can either increase the risk awareness or find other mechanisms to have a part along risk in boosting cooperation (for instance, sanctioning institutions [7]).

2.3. Punishment

Punishment of the defectors [12, 13] is a significant way of increasing cooperation, which occurs not only in human societies but also among other species' societies [14]. Clearly, we may find punishment for bad behaviors on several occasions. Yet, this is not so simple to formulate theoretically nor is to prove that punishment is an explanation for human cooperation.

The first aspect to bear in mind, is that the act

of punish has itself a cost (*costly punishment* or *altruistic punishment*) [12, 15, 16]. Thereby, we must introduce a third type of player or strategy in our EGT framework, the *punisher* (P), besides cooperator (C) and defector (D). As punishing directly the opponents or creating external sanctioning institutions is costly, a second-order dilemma arises, that is, some people do not contribute to the punishment establishment — they are cooperators but not punishers. Thus, even effective institutional sanctioning may be unstable from a dynamical point of view [7].

Another question is how punishment initially emerges. In order to be sustained, punishment needs to attain enough levels such that the induced cooperation compensates the punishment cost. When there are only a few people willing to punish defectors (an extreme situation being a state where there are much fewer punishers than defectors) the costs are too high and maybe vain [17, 13]. Boyd *et al.* remark that "*these problems are an artifact of the unrealistic way that punishment is implemented in existing models and in most experiments*" [17]. The authors criticize models and experiments that consider unconditional and uncoordinated punishment because there is empirical evidence that punishment may be coordinated if punishers communicate with each other. That way, they can predict when the sanctions will be effective depending on the number of punishers in that group.

One way of establishing a conditional punishment (as opposed to *strict*) is using the concept of *graduated punishment*. When we say strict we mean that the punishment applied is constant and does not depend on the state of the system. On the contrary, graduated punishment depends on the harm caused to the society by choosing defection which can be defined in different manners. For instance, it can be proportional to one's offense [18] (if that can be measured) or it can be related to the number of defectors in the population (the more defectors, the bigger should be the sanctions) as pointed out in [19]: "*models where sanctioning is considered as an unchanging part of one's strategy fail to acknowledge a common real-life observation, which is that an increase in antisocial behavior will frequently trigger an increase in both willingness as well as severity of sanctioning amongst those who feel threatened by the negative consequences. Extreme examples thereof include terrorist attacks and other malicious acts, upon which the security measures in the affected areas are often tightened rather drastically*".

Curiously, graduated punishment was empirically observed in specific systems, *Common Pool Resources* (CPR) systems. CPR systems enclose resources which are finite (they can have an end) and non-excludable (it is difficult to prohibit someone to make use of them), as for instance fisheries. By observing and analyzing numerous such systems, Elinor Ostrom, a political economist, found that in many successful communities, that is, communities where a self-management of the resources had sustainability emerged, eight design principles were applied. Graduated punishment is one of those key principles [15]:

> *Appropriators who violate operational rules are likely to be assessed graduated sanctions (depending on the seriousness and context of the offense) by other appropriators, by officials accountable to these appropriators, or by both.*

One of the reasons that Ostrom points out for graduated punishment to work better than strict and severe is that, in the case of a singular act of defection (without precedents and in an unusual problematic situation), a harsh sanction can cause "*resentment and unwillingness to conform to the rules in the future*". Ostrom also notices that "*the appropriators in these CPRs somehow have overcome the presumed problem of the second-order dilemma*". What may explain this achievement is that either the costs of monitoring and sanctioning activities are low or the benefits from it are high (or even both). Particularly, in certain systems, the costs of monitoring are low because they are a natural result of the set of rules that are being used.
Graduated punishment is also recommended in the United States and Europe to cope with environmental offenses [20].

Interestingly, Iwasa and Lee [18], explored graduated punishment mathematically, using the concept of *total welfare function* ϕ, which is the total gain of the community (simply the sum of the payoffs of all individuals). This function is maximized when punishment is gradual, more concretely, when a punishment that grows with the square root of harm caused to the society by the defector is applied. But this is so under specific conditions: possible false accusation (a certain probability of an innocent being punished or a cheater getting away with a bad action) and heterogeneity among people (different sensibilities when adopting a strategy).

Likewise, other authors have wondered about the optimal way of punishing, if it should be strict and severe or graduated [21, 22], and there are several concepts and models [19, 23]. Some even reach contradictory conclusions.

Our hypothesis is that graduated sanctions may have opened an evolutionary route for costly institutions to prevail. We would like to understand how

graduated punishment have evolved with cooperation and how we can apply these ideas to real issues like the present problem of the climate change. This is likely the most important collective dilemma we face [24, 10, 25], and the one where carefully designed institutions and incentives may help us to coordinate efforts towards the preservation of the planet Earth.

3. Methods

We start by following [6] and [7], which study the evolutionary dynamics of a finite population under a CRD (briefly explained in section 2.2). Then, we adapt the model in order to contemplate graduated punishment.

The model goes as follows. In a finite population of size Z, individuals interact in groups of size N, each one choosing one of the three strategies (C, D or P) and starting with an initial endowment or benefit b. Cs and Ps will contribute a fraction of their endowment, c (the cost), to a common amount, while Ds do not contribute. If the total number of contributors (the sum of the number of Cs and Ps) is below a certain threshold n_{pg}, everyone in the group will lose their remaining endowments with a probability r, the perception of risk. Besides this (classical CRD, so far), Ps also contribute with a punishment tax π_t to an external institution that effectively punishes Ds with a fine π_f, if it has enough funding. The sanctioning institution constitutes a second-order public good, only achieved if there are at least n_p punishers (the only contributors in this PGG). We distinguish the type of institution by the scale at which it is formed and upon which it acts — a global one (like the United Nations, for instance) concerns the entire population (is sustained by all Ps and punishes all Ds), whereas a local institution concerns only one group (supported by Ps belonging to that group and punishes Ds inside the same group). These are the two cases considered in [7].

The payoff functions for Cs, Ds, and Ps in a group where there are j_C Cs, j_P Ps, and $j_D \equiv N - j_P - j_C$ Ds can be written as

$$\Pi_C = -c + b\Theta(j_C + j_P - n_{pg}) + \\ +(1-r)b[1-\Theta(j_C + j_P - n_{pg})] \quad (1a)$$

$$\Pi_P = \Pi_C - \Delta_t \quad \text{where} \quad \Delta_t = \pi_t \quad (1b)$$

$$\Pi_D = \Pi_C + c - \Delta_f \Theta(j_P - n_p) \quad \text{where} \quad \Delta_f = \pi_f \quad (1c)$$

where $\Theta(k)$ is the Heaviside function (being zero for $k < 0$ and one for $k \geq 0$), n_{pg} is a positive integer ($0 < n_{pg} < N$), r is real ($0 < r < 1$) and the parameters c, π_t (*tax*), π_f (*fine*) and b are positive real numbers. Equation 1c is defined for local institutions. For a global institution we substitute j_P by the number of punishers of the entire population (i_P). Note also that whenever n_p is achieved, $\Delta_f \Theta(j_P - n_p)$ is a positive constant (strict punishment) [7].

In our model, this is changed since we want to analyze the effect of graduated punishment. This way, we define a new kind of punishment, where Δ_f is dependent on the number of defectors: the higher the number of defectors, the more severe is the punishment. Actually, we choose to inspect and compare three different *sanctioning policies* — *strict punishment and strict costs*, *graduated punishment and strict costs*, and *graduated punishment and graduated costs* (see equations 2 and figure 1).

$$\Delta_{t(f)}^{strict} = \pi_{t(f)} \quad (2a)$$

$$\Delta_{t(f)}^{graduated} \propto \frac{\pi_{t(f)}}{1 + e^{-g[j_D - (N - n_{pg})]}} \quad (2b)$$

Note that $\pi_{t(f)}$ is the average (over the number of defectors j_D) of $\Delta_{t(f)}^{graduated}$, so that the comparison is possible. The parameter g defines the steepness of the functions — the higher the g, the more abrupt is the variation of the fine or tax around the point $N - n_{PG}$.

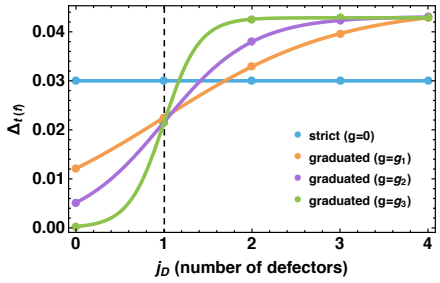

Figure 1: Tax (or fine) $\Delta_{t(f)}$ versus the number of defectors j_D — types of sanctioning policies as defined in equations 2. Here, we define tax (or fine) at local level. Parameter g defines the steepness of the curves ($g_1 < g_2 < g_3$). The areas below lines are equal, so that the comparison is possible. Note that parameters π_f and π_t are not necessarily of the same order; actually, π_t should be smaller than π_f so that punishers do not pay a higher tax than the defectors' fine, and π_f should be higher than the contribution c, otherwise it would still compensate to cheat. For global level, we consider the number of defectors in the entire population and define a global threshold $n_{PG}\prime$ as being $\frac{n_{PG}Z}{N}$. Also, g has to be rescaled. In this particular example, $N = 4$, $n_{PG} = 3$, $g_1 = 1$, $g_2 = 3$, $g_3 = 8$, and $\pi_f = 0.3$.

Now that we have a way to write the payoff of each strategy in an N-person game, what follows is

the population dynamics, that is, how the strategies evolve over time. It is reasonable to think that people choose strategies that provide good earnings through a social learning process equivalent to natural selection. This can be translated by the so-called *replicator equation* [26]

$$\dot{x} = x(1-x)[f_C(x) - f_D(x)] \tag{3}$$

where x is the fraction of cooperators, and f_C and f_D are the *fitness* functions (average payoff) of cooperators and defectors, respectively, which will be defined shortly. Although equation 3 regards only two strategies (Cs and Ds), it can be generalized for any number of strategies.

However, the finiteness of the populations is a feature that requires a different way of tackling the problem because the replicator equation assumes an infinite population and is deterministic (the solution only depends on an initial condition). For instance, sampling a finite population introduces some stochasticity and there could be errors of imitation as well. These effects allow the tunneling through fixed points [27], which may completely alter the outcomes in relation to the deterministic scenario. In order to take these random fluctuations into account, we must use stochastic processes [26].

The state of the system is defined by a vector $\boldsymbol{i} = \{i_1, \ldots, i_k, \ldots, i_s\}$ for the whole population and $\boldsymbol{j} = \{j_1, \ldots, j_k, \ldots, j_s\}$ for the group, where i_k (j_k) is the number of individuals in the population (group) with strategy S_k. Note that there are $s+1$ possible strategies but the state space has only s dimensions because of the restriction $i_1 + \cdots + i_s + i_{s+1} = Z$ ($j_1 + \cdots + j_s + j_{s+1} = N$). We can now write the *fitness* function of a strategy S_k (the average payoff of an individual using that strategy resulting from playing with the other players [28]) in a population with configuration $\boldsymbol{i}$, $f_{S_k}(\boldsymbol{i})$, as [29, 6, 30, 31]

$$f_{S_k}(\boldsymbol{i}) = \binom{Z-1}{N-1}^{-1} \times \\ \times \sum_{(\boldsymbol{j};j_k=N-1)}^{(\boldsymbol{j};j_k=N-1)} \Pi_{S_k}(\boldsymbol{j}) \binom{i_k-1}{j_k} \prod_{l=1(l\neq k)}^{s+1} \binom{i_l}{j_l} \tag{4}$$

where $\Pi_{S_k}(\boldsymbol{j})$ is the payoff of a strategy S_k in a group with composition $\boldsymbol{j}$ and $(\boldsymbol{j}; j_k = q)$ designates any group configuration in which there are specifically q players with strategy S_k. The fitness functions have this form because random sampling without replacement from a finite population leads to groups that follow a hypergeometric distribution. So, fitness is an average payoff over all possible groups in the population (remind that individuals interact within small groups). For infinite populations, we would rather use the binomial distribution.

We can now use equation 4 to compute other interesting variables. For instance, to obtain the average fraction of groups that achieve n_{pg} contributors (that is, that can maintain the public good), $a_G(\boldsymbol{i})$, we must only substitute $\Pi_{S_k}(\boldsymbol{j})$ by $\Theta(j_C + j_P - n_{pg})$. To compute the average fraction of groups that reach n_p Ps (that is, that can support a sanctioning institution), $a_I(\boldsymbol{i})$, we replace $\Pi_{S_k}(\boldsymbol{j})$ by $\Theta(j_P - n_p)$ (for local institutions) or $\Theta(i_P - n_p)$ (for global institutions).

The rule under which strategies evolve (the analogue of the replicator equation) is the *pairwise comparison rule*, equation 5 [32], combined with a stochastic birth-death process [33]. The update is performed as follows: at each time step a random individual is selected to change its strategy (say, X); with a given probability of mutation, μ, this agent will change to a randomly chosen strategy from the space of available strategies; with probability $1-\mu$, another agent is randomly selected (having strategy Y) and the former individual will imitate the strategy of the latter with probability given by

$$\varphi = \frac{1}{1 + e^{\beta(f_X - f_Y)}} \tag{5}$$

where $\beta(\geq 0)$ is a parameter that relates to the intensity of selection and f_X and f_Y are the fitness of strategies X and Y, respectively. This function is the well-known Fermi function from statistical physics. Here, β acts like the inverse of the temperature.

The update process only depends on the current state of the system, thus the dynamics of the vector $\boldsymbol{i}(t)$ (the configuration of the population at time t) corresponds to a *Markov process* over a s-dimensional space [32, 34, 35, 36]. The probability density function, $p_{\boldsymbol{i}}(t)$, is the predominance of configuration $\boldsymbol{i}$ at time t, which evolves under the *master equation* [36]

$$p_{\boldsymbol{i}}(t+\tau) - p_{\boldsymbol{i}}(t) = \sum_{\boldsymbol{i'}} \{T_{\boldsymbol{ii'}} p_{\boldsymbol{i'}}(t) - T_{\boldsymbol{i'i}} p_{\boldsymbol{i}}(t)\} \tag{6}$$

where $T_{\boldsymbol{ii'}}$ and $T_{\boldsymbol{i'i}}$ are the transition probabilities per unit time (transition rates) between configurations $\boldsymbol{i}$ and $\boldsymbol{i'}$.

We are interested in obtaining the stationary distribution ($\overline{p_{\boldsymbol{i}}}$), which gives the probability of each state after a sufficiently long time. For that, we set the left-hand side of equation 6 to zero, which corresponds to an eigenvector search problem [36] (explicitly, the eigenvector associated with the eigenvalue 1 of the transition matrix $\Lambda = [T_{ij}]^T$). The stationary distribution will also allow the computation of relevant quantities such as average group achievement ($\eta_G = \sum_{\boldsymbol{i}} \overline{p_i} a_G(\boldsymbol{i})$) and institution prevalence ($\eta_I = \sum_{\boldsymbol{i}} \overline{p_i} a_I(\boldsymbol{i})$).

Now, to construct the transition matrix, Λ, we need

to compute the transition probabilities among all possible configurations. The way we defined the update process imposes that from one configuration to the next (after one time step) the only transition allowed is: the number of individuals playing a certain strategy is increased by one and the number of individuals playing another strategy is decreased by one, that is, just one agent can change its strategy. It is possible that the former strategy coincides with the latter — in that case, nothing happens. By the pairwise comparison rule (equation 5), the transition rates are then

$$T_{S_l \to S_k} = (1-\mu)\left[\frac{i_l}{Z}\frac{i_k}{Z-1}\left(1+e^{\beta(f_{S_l}-f_{S_k})}\right)^{-1}\right]+\mu\frac{i_l}{sZ} \tag{7}$$

$$T_{\boldsymbol{ii}} = 1 - \sum_{\boldsymbol{i'}\neq\boldsymbol{i}} T_{\boldsymbol{i'i}} \tag{8}$$

where the mutation rate μ is considered and S_l and S_k are two different strategies.
Thus, the probability to increase (decrease) by one the number of individuals with strategy S_k, $T_{\boldsymbol{i}}^{S_k^+}$ ($T_{\boldsymbol{i}}^{S_k^-}$) is

$$T_{\boldsymbol{i}}^{S_k^\pm} = \sum_{i'_1,\ldots,i'_{k-1},i'_{k+1},\ldots,i'_s} T_{\boldsymbol{i}\{i'_1,\ldots,i_k\pm1,\ldots,i'_s\}} \tag{9}$$

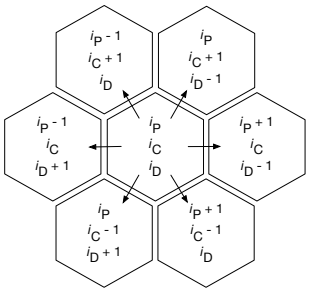

Figure 2: Local representation of two dimensional phase space and possible transitions from state $\boldsymbol{i} = \{i_P, i_C, i_D\}$ (each hexagonal cell represents one state).

These transitions can be used to compute the gradient of selection ($\nabla_{\boldsymbol{i}}$), which indicates the following most likely direction of the phase space when the system is in the state $\boldsymbol{i}$. We can now particularize for the 3-strategy case ($s = 2$ space). In figure 2, the phase space and possible transitions from state $\boldsymbol{i} = \{i_P, i_C, i_D\}$ for a bidimensional one-step process are represented. Hence, the evolutionary dynamics occurs in a 2-dimensional *simplex*, whose basis is defined by the unit vectors $\boldsymbol{u}_C$ and $\boldsymbol{u}_P$ (for instance). The sum of the transition vectors of configuration $\boldsymbol{i}$ (vectors with magnitude $T_{\boldsymbol{ii'}}$ and direction $\boldsymbol{i} \to \boldsymbol{i'}$) corresponds to the gradient of selection or *drift*, which we can write as

$$\nabla_{\boldsymbol{i}} = (T_{\boldsymbol{i}}^{C^+} - T_{\boldsymbol{i}}^{C^-})\boldsymbol{u}_C + (T_{\boldsymbol{i}}^{P^+} - T_{\boldsymbol{i}}^{P^-})\boldsymbol{u}_P \tag{10}$$

The finite population analogues of stable (unstable) fixed points are called probability attractors (repellers), which occur for $\nabla_{\boldsymbol{i}} = 0$.

Summing up, through the described Markov process, we can characterize the system: with the stationary distribution we compute the average group achievement η_G, the institution prevalence η_I, and the average population configuration (population composition averaged over time), with the gradient of selection we know the most probable path and its fastness at each point of the state space.

4. Results and Discussion

In this section, we start by presenting an analysis of 2-strategy CRD and then move on to our main purpose, the outcomes from 3-strategy CRD. Finally, we take a little time thinking about an alternative mechanism.

Having three available strategies, there are three possible 2-strategy games: Ds *versus* Cs, Ds *versus* Ps, and Cs *versus* Ps. In figure 3, we show the effect of punishment in these games.

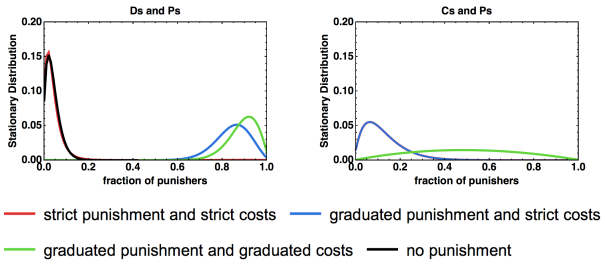

Figure 3: Stationary distribution of 2-strategy games — Ds *versus* Ps (left) and Cs *versus* Ps (right) — for different sanctioning policies (see legend) and local institutions. Note that when sanctions are not applied, a Ds *versus* Ps game is equivalent to a Ds *versus* Cs. Therefore, the black line (no punishment) is equivalent to the stationary distribution of Ds *versus* Cs game. In the right panel, red and blue lines are superimposed because there is no difference from strict to graduated punishment since there are no defectors in that game. Parameters: $Z = 100$, $b = 1$, $c = 0.1$, $r = 0.0$, $N = 4$, $n_{PG} = 0.75 \times N$, $n_P = 0.25 \times N$, $\mu = 1/Z$, $\beta = 5$, $g = 5$ (when graduated), $\pi_f = 0.15$, $\pi_t = 0.03$.

In Ds *versus* Ps game, we can see the red peak placed in the left, while the green and blue are in the right, which is an indication that graduated punishment and costs may effectively be more suitable to attain cooperation than strict in hard conditions. Moreover, in Cs *versus* Ps game, the stationary distribution of graduated costs takes low peak values

and is more evenly distributed through all states (see green line in the right), while for strict costs, the most probable state has few punishers.

But the story cannot end here. Even if in a Ds *versus* Ps game a state with many Ps is reached, it may not be stable. As much as an agent is capable of choosing to contribute or not to the primary public good, he can choose to contribute or not to the sanctioning institution, that is, to be a punisher or a (simple) cooperator. This means that in reality, we must have three strategies because Ps can choose to turn into Cs and that would annihilate the stable states portrayed in the left plot of figure 3 (since to be a cooperator against many punishers is individually beneficial). Hereafter, we consider three strategies (P, C, and D) because only a 3-strategy CRD can embody all possible interactions between agents and consequent effects.

In figure 4, we present the dependence of η_G and η_I on the risk r, for different types of external institutions. Firstly, we confirm that risk increases cooperation (η_G grows with r) as obtained by other authors in [6] and [7]. Newly, we can verify that different sanctioning policies have diverse consequences. The more effective type is graduated punishment and costs (green lines), followed by strict costs and graduated punishment (blue lines), the less effective being the strict punishment and costs (red lines), although still better than not having any sanction institution (black lines). This conclusion can also be checked in figures 5, 6 and 7.

Since for high risk perception it is easy to attain high levels of cooperation, we are more interested in the low risk regime (say, $r \lesssim 0.2$). In that domain, the effect of graduated punishment and costs is even more pronounced, providing a good solution to avoid defection. It is important to notice that the second-order dilemma is avoided — we can see that through the increased values of η_I, which means that an external institution is sustained more frequently, which in its turn grants more cooperation. Still in figure 4, it can be seen that global institutions fail for low risk ($r \lesssim 0.3$) for all three kinds of punishment and costs, which supports the bottom-up institutions philosophy proposed in [7]. For $r > 0.3$, cooperation emerges easily and there is no big variation between the different punishments (all η_G curves fairly overlap) — the risk plays the lead role. Since in the case of graduated costs, the taxes are low when there is little defection, the institution is maintained with minor costs, that is, Cs and Ps are virtually the same, which leads to the rise of η_I green curve.

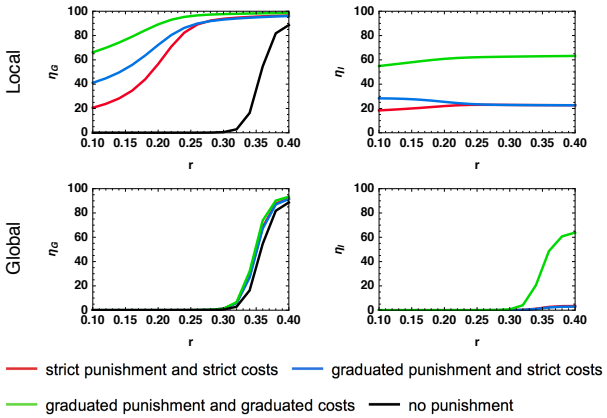

Figure 4: Average group achievement η_G (left) and institution prevalence η_I (right) versus the risk r, for different sanctioning policies, for local (above) and global institutions (below). Parameters: $Z = 100$, $b = 1$, $c = 0.1$, $N = 4$, $n_{pg} = 0.75 \times N$, $n_p = 0.25 \times N$, $\mu = 1/Z$, $\beta = 5$, $g = 2.5$ (local), 0.25 (global), $\pi_f = 0.3$, $\pi_t = 0.03$.

In figure 5, we show simplexes representing the system's dynamics with three strategies: gradient of selection (vectors following a temperature color gradient) and the stationary distribution (grey scale points) at each point of the state space, $\boldsymbol{i}$. Each *vertex* of the simplex is associated with a state entirely populated by just one of the strategies (monomorphic configurations), whereas in each *edge* only two strategies are at stake. Note that the edges of simplexes do not exactly correspond to a 2-strategy game because mutations to a third strategy are included. From figure 5 a) to b), states near vertex C become more probable (darker points). This effect is due to the gradual nature of punishment — when the system is around vertex D (high levels of defection), if the institution is attained, the sanctions on defectors are high, inhibiting their propagation. We can verify this through the gradient of selection near vertex D: while in figure 5 a) the vectors are clearly pointing towards that vertex, in figure b) they are less pronounced, escaping the full defection state and allowing the system to evolve onto highly cooperative states. From figure 5 b) to c), states near C-P edge become more probable and the gradient ceases to bend towards vertex C but strongly points to the C-P edge. This time, also the costs are graduated, therefore in states with few defectors, punishers pay little (or almost no) taxes. In the C-P edge, Cs and Ps are virtually the same and, since the punishers' charges are lowered, the average number of Ps increases (from 11.1% to 40.9%).

The states near C-P are dangerous to defectors because if one pops up due to a mutation, it is almost sure that there will be enough Ps to punish him/her. Hence, η_G increases from 54.2% to 83.7%.

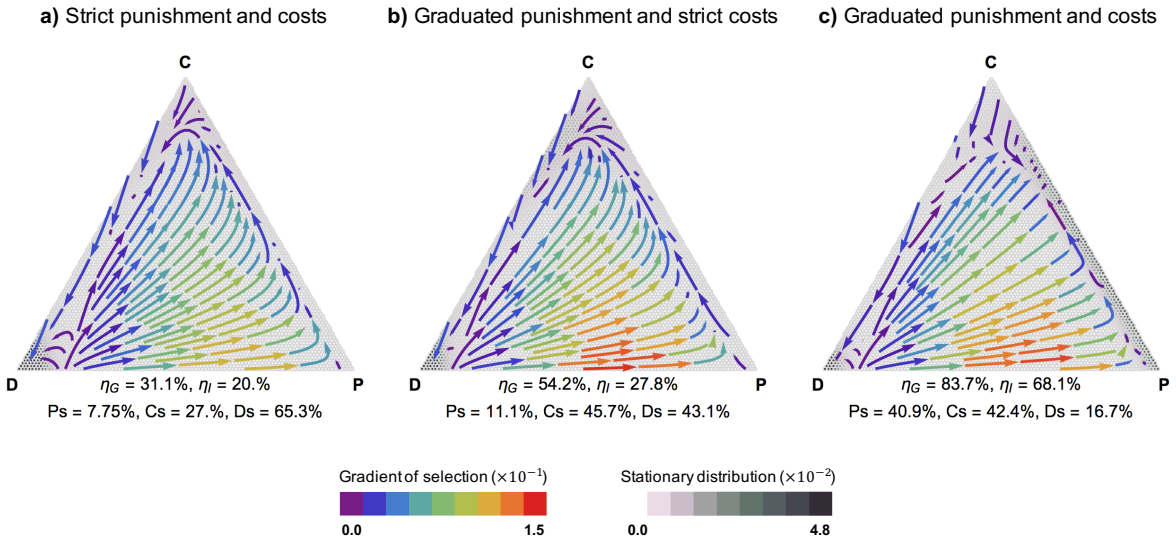

Figure 5: CRD with three strategies — C, P and D. Gradient of selection at each point of the state space $\boldsymbol{i}$ (colored vectors) and stationary distribution (grey scale points) for local institutions and different sanctioning policies. Below each simplex, the corresponding values of η_G, η_I, and the average population configuration are displayed. Parameters: $Z = 100$, $b = 1$, $c = 0.1$, $r = 0.15$, $N = 4$, $n_{pg} = 0.75 \times N$, $n_p = 0.25 \times N$, $\mu = 1/Z$, $\beta = 5$, $g = 5$ (when graduated), $\pi_f = 0.3$, $\pi_t = 0.03$.

One may now investigate how graduate (or steep) should be the punishment and costs. So, we explore the dependence on parameter g introduced in section 3. In figure 6, we can see that the higher this parameter, the more cooperation and institutions are enhanced. This means that the variation of fines and taxes (according to the number of defectors in the group) should be abrupt.

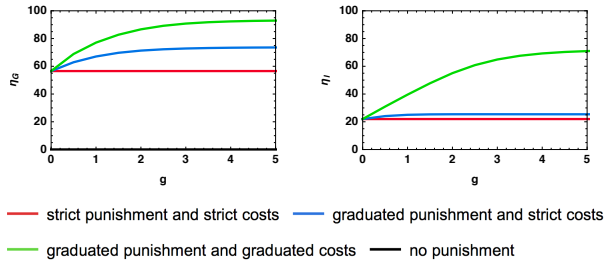

Figure 6: Average group achievement η_G (left) and institution prevalence η_I (right) versus the steepness of graduated punishment/cost g, for local institutions and different sanctioning policies. For $g = 0$, strict punishment/cost is recovered. Parameters: $Z = 100$, $b = 1$, $c = 0.1$, $r = 0.2$, $N = 4$, $n_{pg} = 0.75 \times N$, $n_p = 0.25 \times N$, $\mu = 1/Z$, $\beta = 5$, $\pi_f = 0.3$, $\pi_t = 0.03$

In figure 7, we study the importance of the group size N and conclude that small groups are more favorable to a cooperative behavior (as in [7]) and also to sanctioning institutions' prevalence.

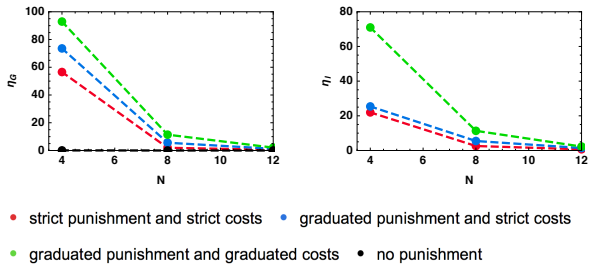

Figure 7: Average group achievement η_G (left) and institution prevalence η_I (right) versus the size of the groups N, for local institutions and different sanctioning policies. Parameters: $Z = 100$, $b = 1$, $c = 0.1$, $r = 0.2$, $n_{pg} = 0.75 \times N$, $n_p = 0.25 \times N$, $\mu = 1/Z$, $\beta = 5$, $g = 5$, $\pi_f = 0.3$, $\pi_t = 0.03$.

The previous results about graduated punishment and costs are robust on the variation of other parameters — c, n_{PG}, n_P, β, μ, π_f, and π_t. If c is high, fewer agents will be willing to contribute, decreasing overall cooperation. Increasing n_p also hinders cooperation, as expected. As for π_f and π_t, increasing π_f enhances cooperation, whereas π_t inhibits it. Despite all these dependencies, graduated punishment and costs work better than strict for a broader set of parameters.

Until now we have just been talking about punishment. However, it is quite natural to think that the reverse kind of incentive, *reward*, might also be a nice tool to instigate a prosocial behavior [37, 38, 39, 40]. We now include a few remarks

about rewards. First, for the same reasons as before, the new incentive is also costly, so, in our EGT framework, we still need agents that pay for it — we call them rewarders (R), although their role is identical to that of punishers. Considering that the ones which should be rewarded are all players except defectors, we write the payoff functions similarly to equations 1, only exchanging the subtractive term of punishment π_f in the payoff of defectors, by an additive term of reward π_r in the payoffs of cooperators and rewarders. Second, rewarding cannot be directly compared to sanctioning in the sense that imposing $\Delta_f = \Delta_r$ wouldn't be fair. The fine should be greater than c but there is no sense in the reward being too — that would unbalance and change the nature of the game. Therefore, our point is not to compare the effect of rewards with sanctions, but rather to assess if graduated rewards lead to better results than strict rewards. So, the idea is to implement graduated rewards analogously to punishment, that is, to give higher rewards to Cs and Rs when these are few. In figure 8, we can see how π_r influences the outcomes.

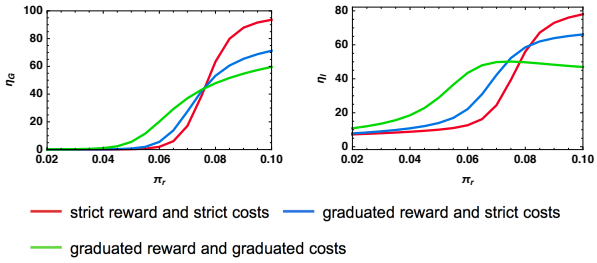

Figure 8: Average group achievement η_G (left) and institution prevalence η_I (right) versus reward π_r for different rewarding policies (see legend) and local institutions. Parameters: $Z = 100$, $b = 1$, $c = 0.1$, $r = 0.2$, $N = 4$, $n_{PG} = 0.75 \times N$, $n_P = 0.25 \times N$, $\mu = 1/Z$, $g = 5$ (when graduated), $\pi_t = 0.03$.

It is clear (and straightforward) that rewards raise cooperation. For $\pi_r \gtrsim 0.08$ (which we consider too high since it almost covers the contribution c), it is easy to understand why strict rewarding best performs: for an already highly cooperative population (guaranteed by rewards in general), graduated rewarding offers smaller prizes comparing to strict. The opposite happens on the small π_r region, namely, for $0.04 \lesssim \pi_r < 0.08$ both graduated reward and strict costs and graduated reward and costs are better policies than strict. The advantage of graduated costs over strict costs is that Rs spend less when Cs and Rs are scant. For that reason, cooperation and institutions emerge more easily for lower values of π_r. On the contrary, the costs become too elevated when there are many agents worthy of prizes, thus η_I starts decreasing at some point. Nevertheless, as we have discussed above, we should not rely on high rewards — they spoil the game as an agent is not supposed to expect a reward of the same order of or higher than contribution c. That said, we conclude that graduated rewards (and costs) promote more cooperation than strict.

5. Conclusions

We conclude that graduated punishment is better than strict at preventing defection in a Collective Risk Dilemma when the conjoint effort needed to maintain the public good is relatively high, especially if the perception of risk is low. This result supports Ostrom's empirical findings [15] — graduated punishment is indeed an enduring collective solution. Thus, it is plausible that it has co-evolved with cooperation. Plus, the proposed formulation of graduated costs solves the problem of second-order free-riding, sustaining external institutions with ease. Our model overcomes some of the problems of precedent ones, namely the artificiality of assuming an unconditional and uncoordinated punishment.

Not surprisingly, rewards also work as a mechanism to avoid the tragedy of the commons. Curiously, we verify that graduated rewards are more effective than strict, similarly to what happens with punishment.

As previously discovered, cooperation emerges more frequently if agents engage in small groups ($N \sim 4$). Newly, graduated punishment works better in such conditions, that is, if regulated by local institutions, instead of a global one. These facts suggest that agreements on climate change should concern regions or a compartmented structure rather than being implemented at a global scale. This is the so-called polycentric or bottom-up approach to environmental governance advocated by several authors [41, 42, 43]. So, using a CRD approach, we prove that polycentricity allied to graduated sanctions is a robust policy to cope with environmental offenses.

Summarizing, our proposal encloses the following sub-achievements: i) a useful (powerful and general) dynamical framework which allows the study of the evolution of strategies in finite populations, ii) a mechanism that promotes cooperation in a risky context, *graduated* incentives, and iii) an application for climate governance.

Acknowledgements

The author would like to thank Professors F. C. Santos and J. M. Pacheco for all their support and motivation, and V. V. Vasconcelos for providing useful code.

References

[1] R. Axelrod, *The Evolution of Cooperation*. Basic Books, Inc., Publishers, 1984.

[2] C. Hauert and G. Szabó, "Game theory and physics," *American Journal of Physics*, vol. 73, pp. 405–414, 5 2005.

[3] M. Perc, J. J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, and A. Szolnoki, "Statistical physics of human cooperation," pp. 1–48, 2017.

[4] E. Pennisi, "How did cooperative behavior evolve?," *Science*, vol. 309, no. 5731, pp. 93–93, 2005.

[5] M. Nowak and R. Highfield, *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed*. Simon and Schuster, 2011.

[6] F. C. Santos and J. M. Pacheco, "Risk of collective failure provides an escape from the tragedy of the commons," *Proc. Natl. Acad. Sci. USA*, vol. 108, no. 26, p. 1042110425, 2011.

[7] V. V. Vasconcelos, F. C. Santos, and J. M. Pacheco, "A bottom-up institutional approach to cooperative governance of risky commons," *Nature Climate Change*, vol. 3, no. 9, pp. 797–801, 2013.

[8] D. G. Rand and M. A. Nowak, "Human cooperation," *Trends Cogn. Sci.*, vol. 17, no. 8, pp. 413–425, 2013.

[9] G. Hardin, "The tragedy of the commons," *Science*, vol. 162, no. 3859, pp. 1243–1248, 1968.

[10] M. Milinski, R. D. Sommerfeld, H.-J. Krambeck, F. A. Reed, and J. Marotzke, "The collective-risk social dilemma and the prevention of simulated dangerous climate change," *Proc. Natl Acad. Sci.*, vol. 105, no. 7, pp. 2291–2294, 2008.

[11] G. Heal and B. Kristrm, "Uncertainty and climate change," *Environ. Resour. Econom.*, vol. 22, no. 1, pp. 3–39, 2002.

[12] R. Axelrod, "An Evolutionary Approach to Norms," *Am. Polit. Sci. Rev.*, vol. 80, no. 4, pp. 1095–1111, 1986.

[13] R. Boyd and P. J. Richerson, "Punishment allows the evolution of cooperation (or anything else) in sizable groups," *Ethology and Sociobiology*, vol. 13, no. 3, pp. 171–195, 1992.

[14] T. H. Clutton-Brock and G. A. Parker, "Punishment in animal societies," *Nature*, vol. 373, pp. 209–216, 1995.

[15] E. Ostrom, *Governing the Commons: The Evolution of Institutions for Collective Action*. New York: Cambridge Univ. Press, 1990.

[16] E. Fehr and S. Gächter, "Altruistic punishment in humans," *Nature*, vol. 415, no. 6868, pp. 137–140, 2002.

[17] R. Boyd, H. Gintis, and S. Bowles, "Coordinated Punishment of Defectors Sustains Cooperation and Can Proliferate When Rare," *Science*, vol. 328, no. 5978, pp. 617–620, 2010.

[18] Y. Iwasa and J.-H. Lee, "Graduated punishment is efficient in resource management if people are heterogeneous," *J. Theor. Biol.*, vol. 333, pp. 117–125, 2013.

[19] M. Perc and A. Szolnoki, "Self-organization of punishment in structured populations," *New J. Phys.*, vol. 14, 2012.

[20] S. Mandiberg and M. Faure, "A graduated punishment approach to environmental crimes: Beyond vindication of administrative authority in the united states and europe," *Columbia J. Environ. Law*, no. 34, pp. 447–511, 2009.

[21] H. Shimao and M. Nakamaru, "Strict or graduated punishment? effect of punishment strictness on the evolution of cooperation in continuous public goods games," *PLoS ONE*, vol. 8, no. 1, pp. 1–10, 2013.

[22] M. Nakamaru and U. Dieckmann, "Runaway selection for cooperation and strict-and-severe punishment," *J. Theor. Biol.*, vol. 257, no. 1, pp. 1–8, 2009.

[23] T. Ohdaira, "Study of the Evolution of Cooperation Based on an Alternative Notion of Punishment Sanction with Jealousy," *Journal of Information Processing*, vol. 24, no. 3, pp. 534–539, 2016.

[24] S. Barrett, *Environment and Statecraft: The Strategy of Environmental Treaty-Making*. Oxford Univ. Press, 2005.

[25] A. Dreber and M. A. Nowak, "Gambling for global goods," *Proc. Natl Acad. Sci.*, vol. 105, no. 7, pp. 2261–2262, 2008.

[26] K. Sigmund, *The Calculus of Selfishness*. Princeton Univ. Press, 2010.

[27] S. H. Strogatz, *Nonlinear Dynamics and Chaos*. Perseus Books, 1994.

[28] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge Univ. Press, 1998.

[29] S. Van Segbroeck, J. M. Pacheco, T. Lenaerts, and F. C. Santos, "Emergence of fairness in repeated group interactions," *Phys. Rev. Lett.*, vol. 108, no. 15, pp. 1–5, 2012.

[30] C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, and K. Sigmund, "Via freedom to coercion: The emergence of costly punishment," *Science*, vol. 316, no. 5833, pp. 1905–1907, 2007.

[31] J. M. Pacheco, F. C. Santos, M. O. Souza, and B. Skyrms, "Evolutionary Dynamics of Collective Action in N-person Stag-Hunt Dilemmas," *Proc. R. Soc. B*, vol. 276, pp. 315–321, 2009.

[32] A. Traulsen, M. Nowak, and J. Pacheco, "Stochastic dynamics of invasion and fixation," *Phys. Rev. E*, vol. 74, no. 1, p. 011909, 2006.

[33] S. Karlin and H. M. Taylor, *A first course in stochastic processes*. Academic Press Inc., 2nd ed., 1975.

[34] A. Traulsen, J. M. Pacheco, and L. A. Imhof, "Stochasticity and evolutionary stability," *Phys. Rev. E*, vol. 74, p. 021905, Aug 2006.

[35] A. Traulsen, J. C. Claussen, and C. Hauert, "Coevolutionary dynamics: From finite to infinite populations," *Phys. Rev. Lett.*, vol. 95, no. 23, pp. 1–4, 2005.

[36] N. G. van Kampen, *Stochastic processes in physics and chemistry*. North-Holland, 3rd ed., 2007.

[37] K. Sigmund, C. Hauert, and M. A. Nowak, "Reward and punishment," *Proc. Natl Acad. Sci.*, vol. 98, no. 19, pp. 10757–10762, 2001.

[38] C. Hauert, "Replicator dynamics of reward reputation in public goods games," *J. Theor. Biol.*, vol. 267, no. 1, pp. 22 – 28, 2010.

[39] T. Sasaki and T. Unemi, "Replicator dynamics in public goods games with reward funds," *J. Theor. Biol.*, vol. 287, no. 1, pp. 109–114, 2011.

[40] T. Sasaki and S. Uchida, "Rewards and the evolution of cooperation in public good games," *Biology Letters*, vol. 10, no. 1, pp. 20130903–20130903, 2014.

[41] K. Sigmund, H. De Silva, A. Traulsen, and C. Hauert, "Social learning promotes institutions for governing the commons," *Nature*, vol. 466, no. 7308, p. 861863, 2010.

[42] E. Ostrom, "Polycentric systems for coping with collective action and global environmental change," *Global Environ. Change*, vol. 20, no. 4, pp. 550 – 557, 2010.

[43] A. Jordan, D. Huitema, H. van Asselt, and J. Forster, eds., *Governing Climate Change: Polycentricity in Action?* Cambridge Univ. Press, 2018.