World Scientific
www.worldscientific.com

Cooperation dynamics of polycentric climate governance

Vítor V. Vasconcelos

Centro de Biologia Molecular e Ambiental,
Universidade do Minho, 4710-057 Braga, Portugal

INESC-ID and Instituto Superior Técnico,
Universidade de Lisboa, IST-Taguspark,
2744-016 Porto Salvo, Portugal

ATP-Group, Instituto Para a Investigação Interdisciplinar,
P-1649-003 Lisboa Codex, Portugal
vvvasconcelos@fc.ul.pt

Francisco C. Santos

INESC-ID and Instituto Superior Técnico,
Universidade de Lisboa, IST-Taguspark,
2744-016 Porto Salvo, Portugal

ATP-Group, Instituto Para a Investigação Interdisciplinar,
P-1649-003 Lisboa Codex, Portugal
franciscocsantos@tecnico.ulisboa.pt

Jorge M. Pacheco*

Centro de Biologia Molecular e Ambiental, and
Departamento de Matemática e Aplicações,
Universidade do Minho, 4710-057 Braga, Portugal

ATP-Group, Instituto Para a Investigação Interdisciplinar,
P-1649-003 Lisboa Codex, Portugal
jmpacheco@math.uminho.pt

Global coordination for the preservation of a common good, such as climate, is one of the most prominent challenges of modern societies. In this manuscript, we use the framework of evolutionary game theory to investigate whether a polycentric structure of multiple small-scale agreements provides a viable solution to solve global dilemmas as climate change governance. We review a stochastic model which incorporates a threshold game of collective action and the idea of risky goods, capturing essential features unveiled in recent experiments. We show how reducing uncertainty both in terms of the perception

*Corresponding author

1

of disaster and in terms of goals induce a transition to cooperation. Taking into account wealth inequality, we explore the impact of the homophily, potentially present in the network of influence of the rich and the poor, in the different contributions of the players. Finally, we discuss the impact of polycentric sanctioning institutions, showing how such a scenario also proves to be more efficient than a single global institution.

Keywords: Governance; cooperation; complex systems; game theory; stochastic processes.

AMS Subject Classification: 22E46, 53C35, 57S20

1. Introduction

Polycentric governance is based on the concurrent learning and action of multiple agents that pursue their interests and that act at a smaller scale than that of the problem at stake.[1–3] In cases when a central, top-down, large-scale agreement or treaty fails to be sought, the polycentric approach has been invoked[2] as a possible means to mitigate the issue or to pave way to a global solution, bringing progress to the stalled process of resolution. We use the powerful framework of evolutionary game theory (EGT) to investigate whether this possibility provides a viable solution, that is, whether polycentricity can be used to solve global dilemmas. In this context, we focus on N-person public goods games (PGG), and on the mechanisms that act to uphold cooperation based on joint decisions made by groups. EGT makes use of the learning process of a multiplicity of interacting agents, facing problems of cooperation. The individuals in these populations are able to revise their strategies depending on their outcome but they have limited information, observing only the acts of others and their final or average outcome, without ever knowing the whole process: this is a challenging setup for cooperation, making it hard to strive. In what follows we show that polycentricity allows for cooperation to emerge even in this adverse scenario.

Despite scientific consensus of its negative impacts in many natural ecosystems, with immediate consequences in human life,[4–6] climate change is one of the examples in which global treaties have failed so far and polycentricity comes up as an alternative. The dilemma over climate issues comes from the fact that regions or nations are tempted to make no effort themselves, while reaping the benefit from the possible efforts of others. Besides this dilemma, akin to many others that humans face, cooperation is sought by world leaders reuniting every year and trying to reach some consensus. The failure of these global summits has been attributed to many factors, of which we distinguish: (i) overall perception of risk is too small, with decision-makers not taking fully into account the effects of missing the targets and discounting[7] its effects; (ii) (scientific/political) uncertainties regarding the exact values of the targets to be met;[8–14] (iii) conflicting policies between rich and poor parties, with a possible segregation of behaviors involving developed countries, on one side, and developing countries, on the other; and (iv) absence of institution(s) to monitor and sanction those not abiding to agreements. Here we show how a polycentric approach may ameliorate the impact of this plethora of effects.

With this in mind, our base model considers a threshold N-person game, in which individuals have to contribute a minimum to effectively contribute to reduce green-house-gas (GHG) emissions, facing two kinds of uncertainty: uncertainty in the outcome and uncertainty in the threshold that defines the collective goal to be met. The first uncertainty concerns the risk that effective efforts prove worthless, meaning that even if the threshold is not met there may still be a chance that nothing catastrophic happens and everyone keeps whatever they have; an effective value attributed to this risk allows EGT to operate for a population under a given risk perception: high levels of risk perception translate into a calculus where existing benefits will likely be lost when contributions are below the threshold, the opposite happening otherwise. The second uncertainty concerns the amount of contributions required for individuals to be sure to retain whatever they have.

In Secs. 2 and 3, we analyze the combined impact of each of these uncertainties when agreements are set at various scales, showing how a polycentric approach to such a global problem helps in reducing the detrimental effect created by uncertainty. Subsequently, in Sec. 4, we augment the base model by splitting the population into two wealth classes: those with high endowments, metaphorically representing the rich, and those with low endowments, representing the poor describing the feedback dynamics between the poor and the rich, and how it acts to build up or diminish the chances of reaching cooperation in each class. We allow these classes to (partially or totally) segregate their behavior, and hence we study the impact of homophily between these classes. Finally, in Sec. 5, we investigate the impact of different kinds of sanctioning institutions to regulate the contribution to GHG reduction, and explain how those institutions created locally, with a smaller range of effect, have a sizeable impact on the overall eagerness to cooperate.

2. Effect of Risk and Scale

On the issues of environmental sustainability and mitigation of the impacts of climate change, one must not overlook uncertainty. Especially when looking at Environmental Agreements, which are typically non-binding,[8,15] uncertainty becomes ubiquitous, as collective investment and its forecast becomes truly unknown: whether it is due to political hesitation or to reservations regarding reversibility, the timings, goal temperatures and even consequences of GHG-induced climate change. Whereas the former are somehow manageable, through negotiation and research, the latter, since it refers to future events, leads to a lot more discussion resulting in (or from) a mindset in which the possibility that nothing damaging happens is non-negligible. Thus, the study of collective action cannot be detached from the overall perception of risk conveyed by climate change,[16,17] a remark that has been reinforced by recent experiments[12,18] and which our base model captures as we show next.[19–23]

We consider a population of Z individuals that are randomly sampled from a population and assembled into groups of size N. Each group is set to play a game

in which a target of M contributions is to be met. Each individual starts with an initial endowment b that can be used to contribute. We start with only two kinds of players: those whose strategy is to contribute (only) to the public good, paying a cost c, the Cooperators (Cs), and those who do not, Defectors (Ds). Consider the possibility that a given group does not reach the predetermined threshold: we call risk, r, the probability of losing the endowment in that situation ($0 \leq r \leq 1$), such that $r = 0$ means that endowments will never be lost, whereas $r = 1$ means that loss of the endowment is certain. Hence, the payoffs of players in a group with k Cs (and $N - k$ Ds) can be written as

$$\Pi_D(k) = bP + (1 - r)b(1 - P),$$
$$\Pi_C(k) = \Pi_D(k) - c \tag{2.1}$$

with P standing for the probability that the group achieves the threshold and $1 - P$ the probability that it does not. We start with the case when there is no uncertainty δ in the threshold, $\delta = 0$. In that case, $P = \Theta(k - M)$, where $\Theta(x)$ is the Heaviside function that is 1 for $x \geq 0$ and 0 otherwise. In Sec. 3, we relax this assumption. Group interactions give individuals a certain payoff, depending on their strategy, whose average value we designate by fitness. We compute the fitness in a well-mixed scenario (the mean-field approximation), where each individual as a fixed (same) probability of interacting with all others and, in a population with i Cs (and $Z - i$ Ds), is given by

$$f_D(i) = \sum_{j=0}^{N-1} \binom{Z-1}{N-1}^{-1} \binom{i}{j} \binom{Z-i-1}{N-1-j} \Pi_D(k),$$
$$f_C(i) = \sum_{j=0}^{N-1} \binom{Z-1}{N-1}^{-1} \binom{i-1}{j} \binom{Z-i}{N-1-j} \Pi_C(k+1). \tag{2.2}$$

Assuming that time evolves in discrete steps, at every step one individual A compares her/his average payoff with that of another randomly chosen individual B and, the larger the payoff of individual B, selected as role model, the more likely it is that A imitates her/his behavior, with a probability given by $p(A, B) = (1 + e^{\beta(f_A - f_B)})^{-1}$, where β controls for learning errors (being usually associated, in the realm of evolutionary dynamics, with the intensity of natural selection). Additionally, individuals can explore the other strategies due to other exogenous factor(s) — technically equivalent to a mutation — controlled by μ. If we let i_A (i_B) be the number of individuals with strategy $A(B)$, then the probability that an individual with strategy A changes to the (different) strategy of B is

$$T_{A \to B} = \frac{i_A}{Z} \left(\frac{i_B}{Z-1}(1 - \mu)p(A, B) + \mu \right). \tag{2.3}$$

The configuration of the population will evolve according to a birth–death process in discrete time, a time-independent Markov process, allowing us to describe the dynamics by means of a Markov chain characterized by the transition probabilities

from a state with i Cs (and $Z - i$ Ds) to a state with i' Cs, $T_{i',i}$. The nonzero transitions are written in Eq. (2.4).

$$
\begin{aligned}
T_{i+1,i} &= T_{D \to C}, \\
T_{i-1,i} &= T_{C \to D}, \\
T_{i,i} &= 1 - T_{D \to C} - T_{C \to D}.
\end{aligned}
\tag{2.4}
$$

We analyze the stationary distribution, given by the eigenvector of matrix $T_{i',i}$, corresponding to the eigenvalue one.[24] Additionally, we compute the most likely direction of evolution of the system (also called the gradient of selection[20,23,25]) as the first Kramers–Moyal coefficient of the expansion of the M-Equation of the process (the so-called drift term). The remaining coefficients tend to zero as the population increases which means that it governs the dynamics in very large populations.[26,27] This coefficient is computed as the difference of the probabilities that the number of individuals of a given strategy goes up and that it goes down, in each independent direction. In this case, it is just $g(i) = T_{i+1,i} - T_{i-1,i}$.

Figure 1 depicts the dynamics and average behavior of a population of individuals for different values of risk for a dilemma played in groups of different sizes. Naturally, in the absence of risk of disaster, there is no point in contributing and thus, apart for random contributions due to some exogenous reason, the dynamics will favor the demise of Cs, with the gradient of selection being always negative. As the risk increases, it leads to the emergence of two roots of the gradient,
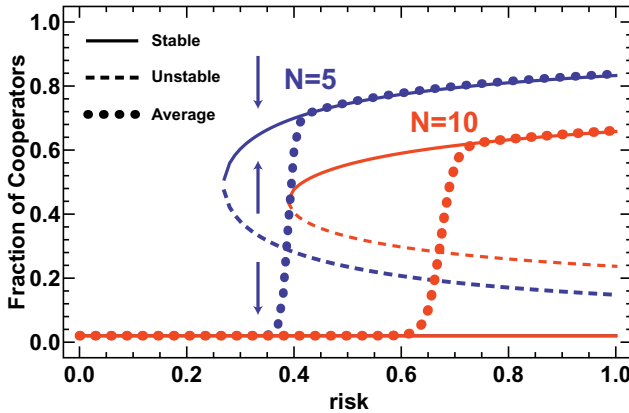

Fig. 1. (Color online) Effect of risk and scale of the agreement. Evolution occurs in populations for different values of risk and group size (N). The dotted lines represent the average fraction of Cs. The solid and dashed lines, together with the arrows, are determined by the gradient of selection which rules the deterministic dynamics. Following the average fraction of Cs, one finds a rapid transition for a critical value of risk, below which there is no cooperation. This transition is triggered by the appearance of a stable fixed point in the deterministic dynamics. The blue and red data denote groups of different sizes, with larger groups being less cooperative both in the value of the transition and in the level of cooperation for high risk. Model parameters: $Z = 200$, $M = N/2$, $b = 1$, $c = 0.1$, $\beta = 5$, $\mu = 1/Z$.

corresponding to unstable and stable mixed internal equilibria in the deterministic dynamics: a coordination between individuals to cooperate is necessary before a stable fraction is able to be robust to changes in strategy. Above a critical value, the average fraction of Cs will increase steadily. Furthermore, under high-risk, this collective coordination becomes easier to achieve and the final level of cooperation is also higher. These results, together with available behavioral experiments,[12,18] demonstrates the key role played by risk perception in favoring the dynamics of Cs.

Given the extending intrinsic global nature of the problem at stake, it is natural to extend this analysis to different group sizes. Larger group sizes implicitly consider less partitioning and, thus, decisions that involve simultaneously a larger and larger fraction of the population. In practice, one can think about group size as the scale at which the decision is being made: smaller group sizes consider several local decisions as opposed to a single large group with the world's fate on its hands. In this model we have not established any relation between the level at which the decision is happening and the risk perception, considering both independent parameters. With this in mind, we can compare, for a given level of risk, if larger groups do better or worse than small ones.

Figure 1 covers two different group sizes with a threshold fixed at half of Cs in the group and it is very clear that smaller groups do better. Not only the transition into a level of high cooperation happens for a lower value of risk for the smaller group but also the overall level that Cs attain is higher. Our results confirm that when the group size becomes comparable to the population size ($N = Z$), cooperation is effectively harder to achieve, suggesting that present world summits may set harder conditions for cooperation than, for instance, a combination of multiple, small-scale, agreements.[20] This effect becomes particularly relevant when collective perception of risk is low, and when economic, and technologic constraints still require sizeable costs from the parties involved, as it is most likely the case in climate negotiations.

3. Threshold Uncertainty

As discussed, the role played by uncertainties associated with incomplete information regarding targets is an unavoidable issue. Experiments observed that threshold uncertainty is detrimental to cooperation.[13,14] This is distinguishable from risk, since it acts not on the consequence of not achieving the threshold but on the threshold itself. Here, we show how our model captures this feature. With all else kept the same, we now introduce variability on the threshold which is now being sampled from a uniform distribution with range $[M - \delta/2, M + \delta/2]$ (other distributions would produce identical results), leading to a change in P, in Eq. (2.1), such that $P = \int_{M-\delta/2}^{M+\delta/2} \frac{1}{\delta}\Theta(k - m)dm$, which is 0 for $k < M - \delta/2$, 1 for $k \geq M + \delta/2$, and $\frac{1}{\delta}(k - M + \delta/2)$ otherwise. This changes the average payoffs of the players, and necessarily their behavior, introducing a region where individuals cannot know what will happen. This region volume in the number space of Cs is δ. Looking at Fig. 2,
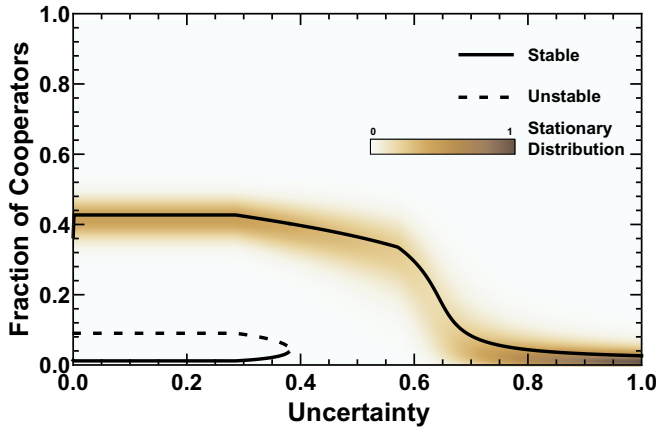
Fig. 2. (Color online) Effect of threshold uncertainty. The solid (dashed) lines correspond to stable (unstable) fixed points in the deterministic dynamics. The color scheme represents the time the population spends in each configuration, given by the stationary distribution. Model parameters: $Z = 200$, $N = 5$, $M = N/2$, $b = 1$, $c = 0.1$, $r = 0.4$, $\beta = 5$, $\mu = 1/Z$. We represent uncertainty as δ/N.

we see that when uncertainty increases the probability that the population remains in the state with high levels of cooperation drops. This corroborates the impetus of the recent report of the Intergovernmental Panel for Climate Change[28] emphasizing research in order to narrow down the amount of threshold uncertainty, namely indicating that humans are the main cause of climate change and, consequently, our actions directly affect the levels needed to reach the targets.

4. Effect of Homophily and Wealth Inequality

Besides risk and uncertainty, lack of consensus in climate summits has also been attributed to conflicting policies between developed and developing countries. In what follows, we introduce wealth inequality in the contributions to the Public Good, mimicking the world's patent wealth inequality and diversity of roles played by different countries. In the light of what was previously found in experiments,[12,29] one might investigate how these roles influence both the distribution of contributions and the effect of homophily in the behavioral dynamics.[30,31] The economic experiments involved groups of six students from western, educated, industrialized, rich, and democratic (WEIRD) countries and were performed using the PGG we described before, introducing different kinds of players: rich and poor, whose initial endowment was higher and lower, respectively. All groups were now composed by three poor individuals and three rich individuals, and showed that in some cases rich would compensate for the smaller tendency to cooperate by the poor, mostly when binding agreements were monitored along the way. In our model, we consider a population with a 1:4 distribution of rich to poor players, roughly reflecting the

present-day status in what concerns the wealth asymmetry of nations. Additionally, instead of a perfectly well-mixed imitation, we allow for the two classes of wealth to limit individuals sphere of influence,[32–34] what is often called homophily[32–34]: high homophily means rich (poor) players only influence and are influenced by rich (poor) players, whereas low homophily means every player is equally likely to influence every other player, independently of their wealth status.

More specifically, when we introduce wealth inequality, we spill the Z individuals into Z_R rich and $Z_P = Z - Z_R$ poor; b and c, in Eq. (2.1), now depend on the class, with b_R (b_P) and c_R (c_P) standing for the initial endowment and cost paid by the rich (poor), respectively. The payoffs of the classes $X = R, P$ are thus written as

$$
\begin{aligned}
\Pi_D^X(k_R, k_P) &= b_X P + (1-r) b_X (1-P), \\
\Pi_C^X(k_R, k_P) &= \Pi_D^X(k_R, k_P) - c_X.
\end{aligned}
\tag{4.1}
$$

The different contributions must now add up to a certain value, which implies a change in P and, since we study the effect in the absence of threshold uncertainty, in a group with k_R rich Cs and k_P poor Cs

$$
P = \Theta(k_R c_R + k_P c_P - M\bar{c}),
\tag{4.2}
$$

with $Z\bar{c} = Z_R c_R + Z_P c_P$. The two classes also introduce a splitting in the sampling for the calculus of the fitness. In the results presented, we do not restrict the fraction of rich and poor in the groups, despite the results being robust to that change.

$$
\begin{aligned}
f_D^R(i_R, i_P) &= \sum_{j=0}^{N-1} \sum_{l=0}^{N-1-j} \binom{Z-1}{N-1}^{-1} \binom{i_R}{j} \binom{i_P}{l} \binom{Z-i_R-i_P-1}{N-1-j-l} \Pi_D^R(j,l), \\
f_C^R(i_R, i_P) &= \sum_{j=0}^{N-1} \sum_{l=0}^{N-1-j} \binom{Z-1}{N-1}^{-1} \binom{i_R-1}{j} \binom{i_P}{l} \binom{Z-i_R-i_P}{N-1-j-l} \Pi_C^R(j+1,l), \\
f_D^P(i_R, i_P) &= \sum_{j=0}^{N-1} \sum_{l=0}^{N-1-j} \binom{Z-1}{N-1}^{-1} \binom{i_R}{j} \binom{i_P}{l} \binom{Z-i_R-i_P-1}{N-1-j-l} \Pi_D^P(j,l), \\
f_C^P(i_R, i_P) &= \sum_{j=0}^{N-1} \sum_{l=0}^{N-1-j} \binom{Z-1}{N-1}^{-1} \binom{i_R}{j} \binom{i_P-1}{l} \binom{Z-i_R-i_P}{N-1-j-l} \Pi_C^P(j,l+1).
\end{aligned}
\tag{4.3}
$$

The imitation dynamics occurs in two sub-populations, eventually restricted by the homophily parameter, h, that incorporates the idea that individuals of a given class $X = P, R$ may be more likely to choose to imitate individuals of the same class than individuals of the opposite class Y. Thus, we can build the transition matrix such that going from a state with a given number of rich and poor Cs (i_R, i_P) to (i_R', i_P'), $T_{(i_R', i_P')(i_R, i_P)}$, can be written using $T_{(i_R \pm 1, i_P)(i_R, i_P)} = T_R^\pm$ and
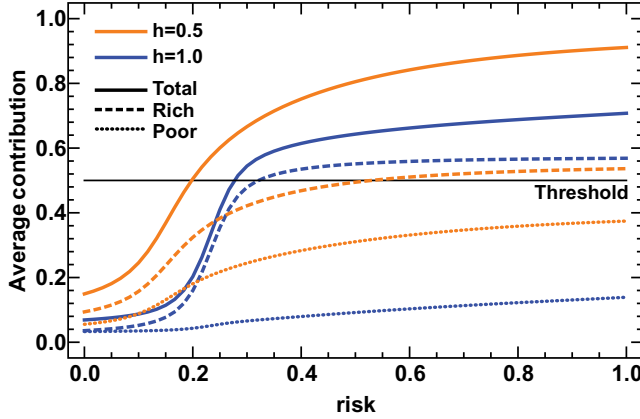
Fig. 3.   Effect of homophily in heterogeneous populations. The solid lines represent the total average contribution per group normalized by the average maximum possible contribution of the groups. The dashed and dotted lines correspond to the decoupling of the total contribution into what is contributed by the rich and poor, respectively. $Z = 200$, $Z_{\mathrm{R}} = 40$, $N = 6$, $M = 3c\bar{b}$, $c = 0.1$, $b_{\mathrm{R}} = 1.7$, $b_{\mathrm{P}} = 0.3$, $(\bar{b} = 1)$, $\beta = 5$, $\mu_X = 1/Z_X$.

$T_{(i_{\mathrm{R}},i_{\mathrm{P}}\pm1)(i_{\mathrm{R}},i_{\mathrm{P}})} = T_{\mathrm{P}}^{\pm}$, with

$$
T_X^+ = \frac{Z_X - i_X}{Z}\left(\left(\frac{i_X}{Z-1-hZ_Y}p(D_X,C_X)\right.\right.
$$
$$
\left.\left.+ \frac{(1-h)i_Y}{Z-1-hZ_Y}p(D_X,C_Y)\right)(1-\mu_X)+\mu_X\right),
$$
$$
T_X^- = \frac{i_X}{Z}\left(\left(\frac{Z_X-i_X}{Z-1-hZ_Y}p(C_X,D_X)\right.\right.
$$
$$
\left.\left.+ \frac{(1-h)(Z_Y-i_Y)}{Z-1-hZ_Y}p(C_X,D_Y)\right)(1-\mu_X)+\mu_X\right). \tag{4.4}
$$

Figure 3 shows that, under the premises of our model, and in agreement with existing experiments,[12,29] the rich generally contribute more than the poor. This effect is even stronger in the presence of high homophily, given that the contribution of the poor is very sensitive to homophily and tends to go down in that case. This, in turn, means that the rich will often compensate for the lower contribution of the other class, a feature which will happen to a limited extent, being dependent on risk. Overall, this also indicates that homophily, if widespread, may lead to a collapse of cooperation, especially in the transition of low to high risk.

5. Sanctioning Institutions

To conclude the list of effects we proposed to address, we explore the effects of global punishment institutions versus locally arranged ones and investigate at which scale

sanctioning should happen.[35] Naturally, given the pros and cons of some procedures against others, agreeing on the way punishment should be implemented is far from reaching a consensus.[36] Here, we discuss two configurations. Institutions need not be global (such as the United Nations), supported by all members willing to punish/sanction, or punishers, that overview all group interactions in the population; they may also be local, group-wide institutions, created to enforce cooperation within a particular group of individuals. While the establishment of global institutions will depend on the total number of punishers in the population, setting up local institutions relies solely on those that exist within a group. Moreover, one does not expect that all the parties (e.g. countries, regions or cities) will be willing to incur a cost in order to sanction others, despite being willing to undertake the necessary measures to mitigate the climate change effects (or, in the language used so far, to cooperate). In other words, one may expect to witness, in general, the three behaviors simultaneously in the population. Figure 2 represents these three behaviors, cooperate, defect and punish as C, D and P providing an overall portrayal of the evolutionary dynamics of the population, in the presence of these three possible behaviors.

With this in mind, in this last section we go back to a population comprising players of the same average wealth and explore the effects of this additional strategy, the punishers (Ps). As before, Cs, but also, Ps contribute a certain fraction of their endowment, in order to reach a common goal, whereas Ds do not contribute. Ps will also contribute to an institution incurring in an additional cost π_t (punishment tax) adding to that associated with cooperation. This cost is paid to the institution to make it able to punish defectors by an amount π_f (punishment fine), whenever the institution reaches a total of contributions $M_I \pi_t$. This creates a second game, by introducing an additional efficiency threshold that must be achieved, now in terms of punishers that contribute both to the public good and to the sanctioning institution. Thus, the punishment pool, or institution, acts as a second-order public good that indirectly increases the investment in the original public good, which, as before, is seen as the health and stability of climate. This leads to a modification of the payoffs in Eq. (2.1) such that for a group with k_C Cs, k_P Ps and $N - k_C - k_P$ Ds the payoffs are

$$\Pi_D(k_C, k_P) = b\mathrm{P} + (1-r)b(1-\mathrm{P}) - \Pi_{\text{scale}},$$

$$\Pi_C(k_C, k_P) = b\mathrm{P} + (1-r)b(1-\mathrm{P}) - c, \tag{5.1}$$

$$\Pi_P(k_C, k_P) = \Pi_C(k_C, k_P) - \pi_t$$

with $\mathrm{P} = \Theta(k_C + k_P - M)$ and the scale being either local or global. In the first case, $\Pi_{\text{local}} = \pi_f \Theta(k_P - M_I)$ and in the later $\Pi_{\text{global}} = \pi_f \Theta(i_P - M_I)$, with i_P being the number of punishers in the whole population. The averages are also computed with the hypergeometric sampling for a given number of Cs, i_C and Ps, i_P (and

Ds, $i_{\mathrm{D}} = Z - i_{\mathrm{C}} - i_{\mathrm{P}}$) in the population:

$$f_{\mathrm{D}}(i_{\mathrm{C}}, i_{\mathrm{P}}) = \sum_{j_1=0}^{N-1} \sum_{j_2=0}^{N-1-j_1} \binom{Z-1}{N-1}^{-1} \binom{i_{\mathrm{C}}}{j_1} \binom{i_{\mathrm{P}}}{j_2} \binom{Z-i_{\mathrm{C}}-i_{\mathrm{P}}-1}{N-1-j_1-j_2}$$
$$\times \Pi_{\mathrm{D}}(j_1, j_2),$$

$$f_{\mathrm{C}}(i_{\mathrm{C}}, i_{\mathrm{P}}) = \sum_{j_1=0}^{N-1} \sum_{j_2=0}^{N-1-j_1} \binom{Z-1}{N-1}^{-1} \binom{i_{\mathrm{C}}-1}{j_1} \binom{i_{\mathrm{P}}}{j_2} \binom{Z-i_{\mathrm{C}}-i_{\mathrm{P}}}{N-1-j_1-j_2}$$
$$\times \Pi_{\mathrm{C}}(j_1+1, j_2),$$

$$f_{\mathrm{P}}(i_{\mathrm{C}}, i_{\mathrm{P}}) = \sum_{j_1=0}^{N-1} \sum_{j_2=0}^{N-1-j_1} \binom{Z-1}{N-1}^{-1} \binom{i_{\mathrm{C}}}{j_1} \binom{i_{\mathrm{P}}-1}{j_2} \binom{Z-i_{\mathrm{C}}-i_{\mathrm{P}}}{N-1-j_1-j_2}$$
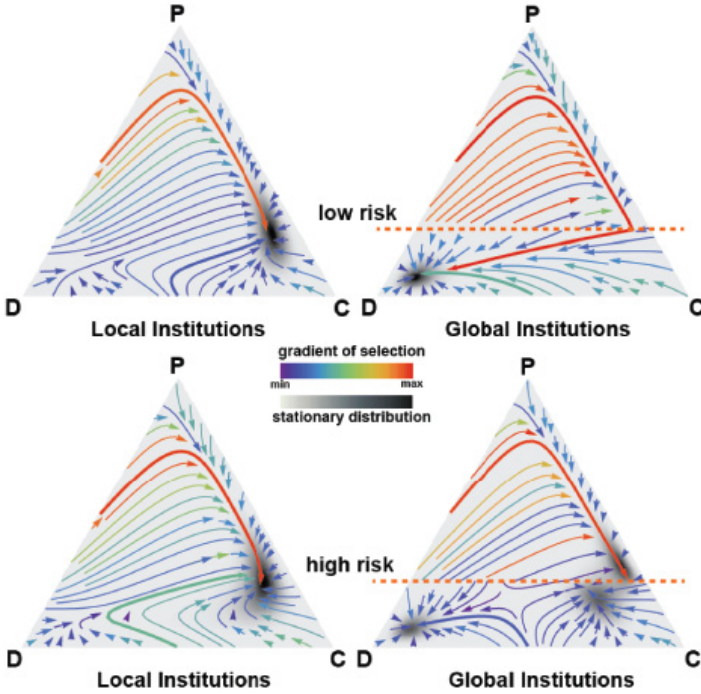$$\times \Pi_{\mathrm{P}}(j_1, j_2+1). \tag{5.2}$$

Fig. 4. (Color online) Effect of local versus global sanctioning institutions. Left panels represent the dynamics with local institution and right panels the dynamics with global institutions, with the orange line representing the threshold in the creation of the institution. The top panels are evaluated for low risk and the bottom panels for high risk. The gradient of selection is represented as a stream indicating in each point the most likely direction and which corresponds, in the deterministic dynamics, to paths of the behavior of the population. $Z = 200$, $\mu = 0.05$, $N = 8$, $M = 6$, $c = 0.1$, $b = 1$, $\beta = 5$, $M_I = 25\%$, $\pi_t = 0.03$, $\pi_f = 0.3$, $r = 0.2$ (low risk), $r = 0.5$ (high risk).

Finally, the transitions that build the transition matrix are given by the transition between any pair of strategies. The probability that an individual with strategy $A = C, D, P$ changes to another strategy, B, is given by Eq. (2.3), with the mutation term $i_A/Z\mu$ now divided by two. This, again, allows us to build the transition matrix from which the stationary distribution is extracted.

Figure 4 contains the key elements of the dynamics of this complex system. The local institutions are always more efficient than the global one. In the depicted case, the local institutions are able to cancel the effect of the attractor close to configuration where everyone defects, pushing the population to a cooperative state with the average number of punishers just over the threshold. On the other hand, the global institution keeps that attractor, since it exists when the institution is not working. For high risk though, the institution seems to work quite on the verge, allowing for sequences of small invasions of defectors before becoming effective, creating three areas of behavior: full defection, on one side and, on the other, full cooperation with enough punishers to maintain the institution which sequentially goes to a mixed state of defectors and Cs that goes back to an effective institution. Thus, local institutions act as a second-order public good, thus being more effective locally.

6. Conclusions

Our approach defines a mathematical framework to study what has been coined as polycentric governance[2] based on empirical work on fisheries.[1] Our results show that high levels of cooperation can indeed be reached globally via such a polycentric approach, when a multiplicity of small groups face local dilemmas.[20,23] However, this mechanism is not bullet proof and suffers from fragilities which are of the same stance as those stemming from the usual top-down approach. We discuss this in the context of climate change negotiations and introduce four of the main issues that are pointed as causes to their failure. We show how uncertainty is a key factor to this threshold game, whether in terms of the perception of disaster or in terms of the targets to be met. Both low levels of risk perception or high threshold uncertainty induce a critical transition into a state where everyone involved ceases to cooperate. Additionally, we consider the possibility that individuals have different abilities to contribute, taking into account wealth inequality. We show how homophily in terms of the network of influence of the rich and the poor leads to an added effort from the rich and, in some cases, to a collapse of cooperation. In any of these cases, all of which are detrimental to cooperation, the cooperative state is more robust the smaller the groups, an observation that strongly supports the polycentric approach. Finally, we study the possibility of supervision of these dilemmas via institutions and again pose the question of at which scale should it be implemented, comparing, in the setup of EGT, the efficiency of a global institution overviewing all interactions versus that of locally arranged ones. Our results show that polycentric sanctioning is more efficient in monitoring the multiple interactions.

In any case, the chance of failing to solve the climate change problem is still very high, something slightly amended whenever demanding thresholds are adopted.[20,23] Moreover, if intermediate tasks are designated,[29] or if individuals have the opportunity to pledge their contribution before actual action,[12] cooperation is also more prominent. Nonetheless, some of these results need to be tested in the field. This should happen not only in terms of their validity but also in order to pursue the identification of the critical parameters that define the regime we are in. How close are we to the transition in terms of risk perception in a specific issue? How much can we reduce threshold uncertainty? Up to which scale can we implement local institutions and keep the cost:size proportionality? Answering these practical questions is what, ultimately, will render feasible the empirical implementation of our results.

Acknowledgments

References

1. E. Ostrom, *Governing the Commons*: *The Evolution of Institutions for Collective Action* (Cambridge Univ. Press, 1990).
2. E. Ostrom, A polycentric approach for coping with climate change, *Policy Research Working Paper*, Background Paper to the 2010 World Development Report, WPS5095, 2009.
3. V. Ostrom, *The Quest to Understand Human Affairs*: *Natural Resources Policy and Essays on Community and Collective Choice* (Lexington Books, 2011).
4. R. B. Alley, B. Hewitson, B. Hoskins, F. Joos, J. Jouzel, V. Kattsov, U. Lohmann, M. Manning, T. Matsuno and M. Molina, *Summary for Policy Makers*; Global Humanitarian Forum, 2007.
5. G. A. Meehl, T. F. Stocker, W. D. Collins, P. Friedlingstein, A. T. Gaye, J. M. Gregory, A. Kitoh, R. Knutti, J. M. Murphy and A. Noda, *2007*: *Global Climate Projections*, (Cambridge Univ. Press, 2007), pp. 747–845.
6. P. A. Stott, N. P. Gillett, G. C. Hegerl, D. J. Karoly, D. A. Stone, X. Zhang and F. Zwiers, Detection and attribution of climate change: A regional perspective, *Wiley Interdisciplinary Rev.*: *Climate Change* **1** (2010) 192–211.
7. S. Levin, The trouble of discounting tomorrow, *Solutions* **4** (2012) 20–24.
8. B. C. O'Neill and M. Oppenheimer, Dangerous climate impacts and the Kyoto Protocol, *Science* **296** (2002) 1971–1972.
9. S. Barrett, *Environment and Statecraft*: *The Strategy of Environmental Treaty-Making* (Oxford Univ. Press, 2005).
10. M. Oppenheimer and R. B. Alley, Ice sheets, global warming and Article 2 of the UNFCCC, *Climatic Change* **68** (2005) 257–267.
11. S. Barrett, *Why Cooperate?*: *The Incentive to Supply Global Public Goods* (Oxford Univ. Press, 2007).

12. A. Tavoni, A. Dannenberg, G. Kallis and A. Lschel, Inequality, communication and the avoidance of disastrous climate change in a public goods game, *Proc. Natl. Acad. Sci. USA* **108** (2011) 11825–11829.

13. S. Barrett and A. Dannenberg, Climate negotiations under scientific uncertainty, *Proc. Natl. Acad. Sci. USA* **109** (2012) 17372–17376.

14. S. Barrett and A. Dannenberg, Sensitivity of collective action to uncertainty about climate tipping points, *Nature Climate Change* **4** (2014) 36–39.

15. K. Jaimet, A primer on Kyoto, 2002.

16. G. Heal, *Formation in International Environmental Agreements* (Kluwer, 1993).

17. G. Heal and B. Kristrom, Uncertainty and climate change, *Environ. Res. Econ.* **22** (2002) 3–39.

18. 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. USA* **105** (2008) 2291–2294.

19. J. Wang, F. Fu, T. Wu and L. Wang, Emergence of social cooperation in threshold public goods games with collective risk, *Phys. Rev. E* **80** (2009) 016101.

20. 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* **108** (2011) 10421-5.

21. M. Abou Chakra and A. Traulsen, Evolutionary dynamics of strategic behavior in a collective-risk dilemma, *PLoS Comput. Biol.* **8** (2012) e1002652.

22. X. Chen, A. Szolnoki and M. Perc, Risk-driven migration and the collective-risk social dilemma, *Phys. Rev. E* **86** (2012) 036101.

23. F. C. Santos, V. V. Vasconcelos, M. D. Santos, P. N. B. Neves and J. M. Pacheco, Evolutionary dynamics of climate change under collective-risk dilemmas, *Math. Models Methods Appl. Sci.* **22** (2012) 1140004.

24. N. V. Kampen, *Stochastic Processes in Physics and Chemistry* (North-Holland, 2007).

25. J. M. Pacheco, F. C. Santos, M. O. Souza and B. Skyrms, Evolutionary dynamics of collective action in n-person stag hunt dilemmas, *Proc. Roy. Soc. Lond. B*: *Biol. Sci.* **276** (2009) 315–321.

26. A. Traulsen, M. A. Nowak and J. M. Pacheco, Stochastic dynamics of invasion and fixation, *Phys. Rev. E* **74** (2006) 011909.

27. J. M. Pacheco, V. V. Vasconcelos and F. C. Santos, Climate change governance, cooperation and self-organization, *Phys. Life Rev.* **11** (2014) 573–586.

28. IPCC (2013). *Climate Change 2013*: *The Physical Science Basis*. Contribution of Working Group Fifth Assessment Report of the Intergovernmental Panel on Climate Change, eds. T. F. Stocker, D. Qin, G.-K. Plattner, M. Tignor, S. K. Allen, J. Boschung, A. Navels, Y. Xia, V. Bex and P. M. Midgley (Cambridge Univ. Press, Cambridge, UK and New York, USA).

29. M. Milinski, T. Röhl and J. Marotzke, Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets, *Climatic Change* **109** (2011) 807–814.

30. M. Abou Chakra and A. Traulsen, Under high stakes and uncertainty the rich should lend the poor a helping hand, *J. Theor. Biol.* **341** (2014) 123–130.

31. V. V. Vasconcelos, F. C. Santos, J. M. Pacheco and S. A. Levin, Climate policies under wealth inequality, *Proc. Natl. Acad. Sci. USA* **111** (2014) 2212–2216.

32. M. McPherson, L. Smith-Lovin and J. M. Cook, Birds of a feather: Homophily in social networks, *Ann. Rev. Soc.* **27** (2001) 415–444.

33. J. H. Fowler and N. A. Christakis, Cooperative behavior cascades in human social networks, *Proc. Natl. Acad. Sci. USA* **107** (2010) 5334-8.

34. F. L. Pinheiro, M. D. Santos, F. C. Santos and J. M. Pacheco, Origin of peer influence in social networks, *Phys. Rev. Lett.* **112** (2014) 098702.

35. V. V. Vasconcelos, F. C. Santos and J. M. Pacheco, A bottom-up institutional approach to cooperative governance of risky commons, *Nature Climate Change* **3** (2013) 797–801.
36. A. Traulsen, T. Röhl and M. Milinski, An economic experiment reveals that humans prefer pool punishment to maintain the commons, *Proc. Roy. Soc. B*: *Biol. Sci.* **279** (2012) 3716–3721.