



OPEN

A new electoral bottom-up model of institutional governance

Carlos M. Garrido^{1,2}, Francisco C. Santos^{2,3,4}, Elias Fernández Domingos^{5,6}, Ana M. Nunes^{1,7} & Jorge M. Pacheco^{2,3}✉

The sustainable governance of Global Risky Commons (GRC)—global commons in the presence of a sizable risk of overall failure—is ubiquitous and requires a global solution. A prominent example is the mitigation of the adverse effects of global warming. In this context, the Collective Risk Dilemma (CRD) provides a convenient baseline model which captures many important features associated with GRC type problems by formulating them as problems of cooperation. Here we make use of the CRD to develop, for the first time, a bottom-up institutional governance framework of GRC. We find that the endogenous creation of local institutions that require a minimum consensus amongst group members—who, in turn, decide the nature of the institution (reward/punishment) via an electoral process—leads to higher overall cooperation than previously proposed designs, especially at low risk, proving that carrots and sticks implemented through local voting processes are more powerful than other designs. The stochastic evolutionary game theoretical model framework developed here further allows us to directly compare our results with those stemming from previous models of institutional governance. The model and the methods employed here are relevant and general enough to be applied to a variety of contemporary interdisciplinary problems.

Keywords Cooperation, Collective risk dilemma, Evolutionary game theory, Stochastic processes

Global Warming, considered by the United Nations (UN) as one of the most important global problems we face¹, is a typical example of a GRC whose solution requires worldwide cooperation. The Collective Risk Dilemma (CRD) models theoretically this GRC as a risky, threshold public goods game. It has been widely employed in recent years^{2–18}, and represents a theoretical adaptation of the original design employed in behavioral experiments^{19–21}. Cooperation in the CRD means paying a cost to mitigate, whose benefits may or may not become available to all. Thus, the temptation to free-ride on the benefits produced by others at their own expense is an inescapable component of the model. However, in order to secure the provision of the public good, a minimum number of Cooperators is needed, without whom every member of the group is at risk of losing all they have (for details, see Methods).

Previous work⁷ making use of the CRD indicates that sanctioning institutions help to solve the Global Problem provided they are endogenously created at a Local scale, whereas global institutions (such as the UN) do not change the qualitative behavior of the CRD as a function of risk compared to a scenario where no institutions are at work (black solid line in Fig. 3), supporting the widely repeated motto “think globally, act locally”.

Furthermore, even though some global agreements, such as the Montreal Protocol²² and the Kigali Amendment to the Montreal Protocol²³, look promising in tackling global dilemmas related to the production of harmful gases, an all-encompassing agreement addressing Global Warming, reached at a global scale, would still require the implementation of solutions adapted to particular specificities and challenges at a local scale, calling for a polycentric approach to best deal with each particular regional challenge²⁴. Consequently, we shall concentrate, in the following, on local, endogenously created, institutions.

The creation of institutions poses another costly public good, leading to a second-order social dilemma that separates those who contribute to the institution that provides the incentives from those who do not contribute. Our model not only takes these points into consideration but also lets individuals in each group decide, through a voting process, the nature of the institution they create: Either a sanctioning institution (that provides negative

¹BioSystems and Integrative Sciences Institute, Faculdade de Ciências da Universidade de Lisboa Campo Grande, 1749-016 Lisbon, Portugal. ²ATP-group, 2744-016 Porto Salvo, Portugal. ³INESC-ID, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal. ⁴Instituto Superior Técnico, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal. ⁵AI Lab, Computer Science Department, Vrije Universiteit Brussel, Pleinlaan 9, 3rd Floor, 1050 Brussels, Belgium. ⁶MLG, Département D’Informatique, Université Libre de Bruxelles, Boulevard Du Triomphe, CP 212, 1050 Brussels, Belgium. ⁷Departamento de Física, Faculdade de Ciências da Universidade de Lisboa, Campo Grande, 1749-016 Lisboa, Portugal. ✉email: jmpacheco@math.uminho.pt

incentives to anti-social group members) or a rewarding institution (that provides positive incentives to pro-social group members).

Our new Bottom-Up institutional design is illustrated in Fig. 1. To successfully create an institution, a minimum number (n_I) of institution creators needs to be present in each group. The nature (reward/punishment) of the Local institution thus created will be decided by a majority rule, that is, an electoral process performed by the institution creators: A majority of rewarders (cf. Fig. 1 for strategy identification) leads to a rewarding institution, whereas a majority of Punishers originates a sanctioning institution. Mathematically, this implies augmenting the CRD to include, besides Cooperators (C) and Defectors (D), rewarders (R) and Punishers (P) (see Methods). Consequently, the evolutionary dynamics of the 4 types of individuals will unfold in a simplex (phase space) with the shape of a regular tetrahedron, as illustrated in Fig. 2. This setting increases considerably the complexity of the present population dynamics compared to previous approaches^{2–18}.

In the stochastic population dynamics with mutations considered here (see Methods), the evolutionary dynamics is generally characterized by the occurrence of 2 attractors in the interior of the simplex. Figure 2 illustrates the prevailing scenario (for intermediate to high values of risk r , $r \geq 0.4$) where the attractors are depicted with orange solid spheres²⁵. These 2 attractors originate 2 basins of attraction that are typically separated by a “fitness barrier” associated with configurations with a nearly constant number of Ds in the population, leading to a planar-like surface whose intersection with the simplex is illustrated (qualitatively) by black dashed lines.

For low risk, and despite the possible occurrence of 2 attractors, the population is not able to tunnel across the fitness barrier, leading to an average configuration of the population that remains very close to the ALL-D configuration.

At higher risk, the population is now able to explore the other side of the fitness barrier, such that the dynamics becomes dominated by the attractor located in the cooperating region of the simplex, at a configuration dominated by Cs (cf. Fig. 2).

In the Supplementary Information (SI) we provide details of the stationary distribution (defined in Methods) associated with these different scenarios and we also test the robustness of the results shown to changes in the model parameters.

A ubiquitous feature of the CRD (also observed in present model, c.f. Fig. 3) is associated with a transition from a defector dominated dynamics at low-risk to a cooperator dominated dynamics at high-risk, irrespective of the existence of institutions, as well as of their nature. This behavior translates into a S-shape profile when one plots η_G —the population average group achievement, defined in Methods—as a function of risk. Clearly, as shown previously, different institutional designs typically contribute to change the critical values of risk at which the transition from Defection to Cooperation occurs, as well as the rate of this transition.

The quantity η_G not only captures these transitions in a single curve, but it has the additional advantage of allowing one to directly compare the performance of models of different inherent dimensionality. These models were generally developed making use of different parametrizations of costs, benefits, rewards and sanctions. Because our model encompasses all mechanisms introduced before, we are able to compare the performance of all models in the present framework. This is precisely what is shown in Fig. 3 where the main findings of this work are shown.

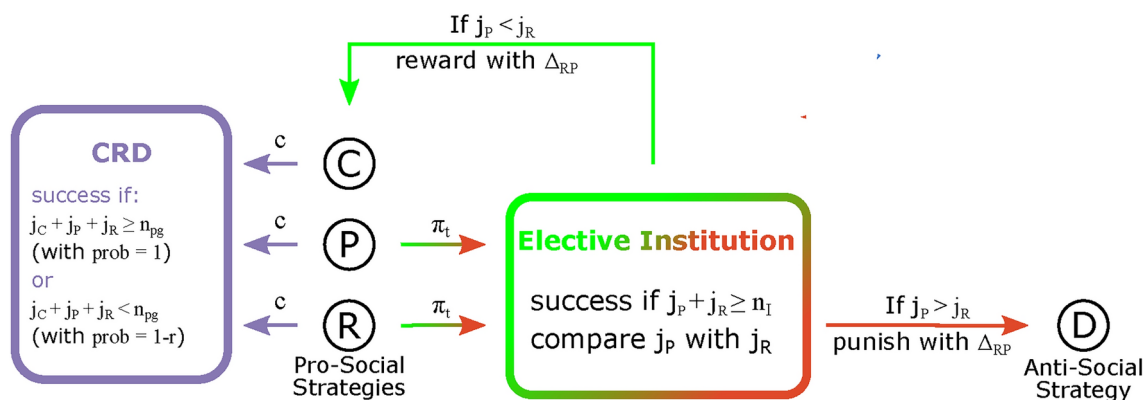


Figure 1. Schematic representation of the Bottom-Up institutional model developed here to address the sustainable governance of GRC making use of the CRD dilemma. There are three pro-social strategies: C (Cooperators) P (Punishers) and R (rewarders) that contribute with a fraction c of their endowment to the CRD (left-pointing arrows); and one anti-social strategy, D (Defectors), that does not contribute. P and R also pay an additional tax π_t to fund an Electoral institution (right-pointing arrows) which decides, based on a majority rule, how to use its revenue Δ_{RP} : Either to Punish the Ds, to reward the pro-social strategies: C, P and R or to reward the pro-social and Punish the anti-social in case of a tie. In the well-mixed approximation, the configuration of the population can be specified by the state vector $i = \{i_C, i_P, i_R, i_D\}$, where i_S is the number of individuals using strategy S in the population. We use a corresponding notation to specify the composition of each group: $j = \{j_C, j_P, j_R, j_D\}$ where j_S is the number of individuals using strategy S in the group—see Methods for full details of the model.

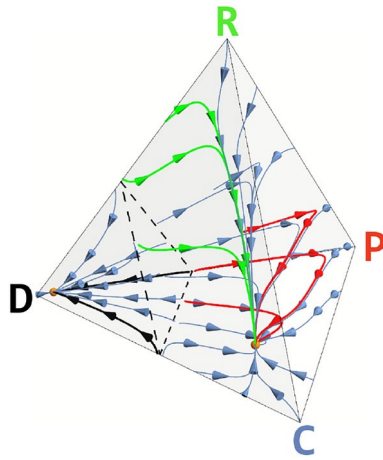


Figure 2. Stochastic Evolutionary dynamics of the Bottom-Up Electoral model developed here, showing that the overall dynamics is dominated, for most parameter values, by 2 interior attractors, both depicted with solid orange spheres: One close to the ALL-D configuration, and another (“cooperative”) at a configuration where Cs clearly dominate. The black and blue arrows (to the left of the dashed triangle) illustrate the most likely paths (in a stochastic sense) that converge to the ALL-D attractor, whereas the blue, green and red arrows (to the right of the dashed triangle) illustrate those paths that converge to the cooperative attractor. Whenever $r < 0.5$ (for the model parameters chosen) the population remains, most of the time, near the ALL-D configuration, well below the fitness barrier illustrated by black dashed lines that qualitatively represent the intersection of the (quasi-planar) fitness barrier top surface with the surface of the simplex. Whenever $r > 0.7$ (see Fig. 3) the population is now able both to tunnel through the barrier and to evolve towards the cooperative attractor where (in a total of 70 individuals) there are 52 Cs, 7 Rs, 6 Ps and 5 Ds. Finally, among the plethora of trajectories (in blue) converging to the “cooperative” attractor, we distinguish 2 types belonging to different “classes” of evolutionary paths: 1) In green we illustrate paths where convergence evolves mostly due to an initial rise of Rs without any significant participation of Ps. 2) In red, paths where convergence evolves mostly through a significant increase of Ps without any significant participation of Rs. As discussed in detail in the main text, the first class of trajectories is more abundant than the second class. Parameters used: $Z = 70$, $\mu = 1/Z$, $\beta = 5$, $N = 8$, $n_{pg} = 6$, $n_I = 2$, $b = 1$, $c = 0.1$, $\delta = 2$, $\pi_t = 0.03$, $r = 0.8$.

Results

In the absence of institutions (black dashed line with open circles, where evolutionary dynamics occurs in a one-dimensional simplex²) and for the model parameter values indicated in Fig. 3, one needs a critical risk value $r^* \approx 0.755$ to reach $\eta_G \approx 0.5$.

Local institutions, either sanctioning (red solid line with solid circles) or rewarding (green dashed line with open circles), whose evolutionary dynamics unfolds in a two-dimensional simplex^{7,12,18}) considerably improve the overall prospects of cooperation in the CRD: Now $\eta_G \approx 0.5$ is reached at $r^* \approx 0.695$ and $r^* \approx 0.675$, respectively.

Further improvement (blue solid line with solid circles, where evolutionary dynamics unfolds in a three-dimensional simplex—see Fig. 2) is obtained via the present model employing local, electoral institutions, of a nature (reward/punishment) that is decided at a group level, case by case, leading to $r^* \approx 0.640$.

The results in Fig. 3 show that the present model, which encompasses electoral institutions of a dual nature, outperforms previous approaches considering reward-only or punishment-only institutions (see also¹⁸), more so if we take the rate of change from defection to cooperation, estimated by computing, at the r^* values reported before, the quantity $(d\eta_G/dr)_{r=r^*}$, which is maximal for the present model.

Figure 2 proves helpful in developing an intuition to understand the superior performance of the present model compared to others, in particular to models of reward-only institutions. Indeed, in reward-only models, the evolutionary dynamics proceeds exclusively along the triangular D-C-R face of the tetrahedron. This will constrain possible paths towards the cooperative attractor to those pertaining to the green class illustrated in Fig. 2. The paths belonging to the red class, also illustrated in Fig. 2 to be accessible in the present Electoral model, are not available in reward-only models, which explains the less cooperative dynamics observed for these models. A similar argument is valid for punishment-only institutions which, per se (and for most model parameters), lead to poorer overall performance compared to reward-only institutions. Indeed, if we estimate the fraction of red-type to green-type evolutionary trajectories obtained via computer simulations of time evolution, starting from the vicinity of the ALL-D configuration, we obtain a (risk-dependent) value of ~ 0.2 (for $r = 0.8$), which supports the intuition that punishment-only institutions are less efficient than reward-only institutions, although their combination, via the electoral process developed here, provides superior results. Similar conclusions can be drawn if we compute the fraction of time punishment institutions prevail over rewarding institutions and vice-versa, as shown in the SI.

It is also worth pointing out that, as illustrated in Fig. 2, Cs clearly dominate in the configuration defined by the cooperative attractor of the stochastic evolutionary dynamics; however, it is important to realize that there

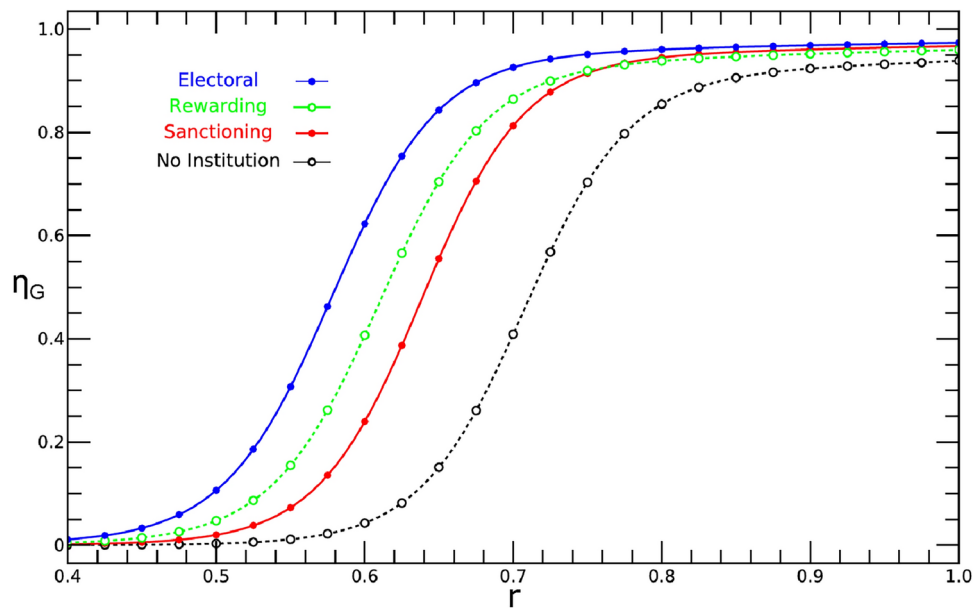


Figure 3. The population average group achievement η_G (defined in Methods) is plotted as a function of risk for the case of i) No institutions (black dashed line with open circles); ii) Local sanctioning (punishing) institutions (red solid line with solid circles); iii) Local rewarding institutions (green dashed line with open circles) and iv) Local Electoral institutions (blue solid line with solid circles). Clearly, the present bottom-up approach leads to higher overall cooperation, for all values of risk, compared to other models developed previously. Parameters used: $Z = 140$, $\mu = 1/Z$, $\beta = 2N = 8$, $n_{pg} = 6$, $n_I = 2$, $b = 1$, $c = 0.1$, $\pi_t = 0.03$, $\delta = 2$.

are small, but non-negligible, numbers of both Rs and Ps present in the population in this configuration. These numbers (note that, for the parameters employed in Fig. 2 the average number of Ps and Rs grow from 1 at $r = 0.5$ to 5 and 6 at $r = 0.8$, respectively) ensure the existence of a reservoir of institutional individuals that are able to maintain some policing of groups, rewarding/sanctioning their members accordingly.

Discussion

As stated in the beginning, sanctions are harder to implement than rewards, even when they result from the celebration of International Agreements. Therefore, it is gratifying to realize that rewarding institutions alone, which have an easier job in implementing their goals, are generally more efficient to promote cooperation than sanctioning institutions, cf. Fig. 3.

The present model, however, reinforces the idea that sanctioning institutions are important in “policing” free-riders^{7,12,26}, here at the group level, where sanctioning is more likely to be effective.

In the present model, our analysis concentrated on the long time distribution of strategies in the population, by computing the stationary distribution and associated observables. This methodology allowed us to compare the performance of different rewarding and sanctioning models with different inherent dimensionality, which was our primary goal. This said, we did not address in detail other aspects of the present model which are also of relevance, namely the behavior of strategies in time. As is well known, evolutionary game models may lead to oscillatory behaviour, both under deterministic^{27–29} and stochastic^{30,31} dynamics. In the present model, the prevailing scenario portrayed in Fig. 2 does not lead to periodic oscillations, although one cannot rule out the occurrence of oscillatory behavior for particular parameter combinations or in situations where spatial effects are added to the model.

To summarize, we developed a new bottom-up model of institutional governance of risky commons which allows the creation of flexible, local institutions, by introducing a threshold requirement for their formation at a local, group level, and whose nature (reward or punishment) is decided, also locally, via a voting process (majority rule).

This model not only promotes the self-organization of cooperation in the population for values of risk significantly lower than previous models, but also keeps reward as the dominant mechanism while keeping sanctioning mechanisms available at all times, and applied at a local level (that is, in smaller groups), which helps to make their efficiency more feasible.

Methods

In all cases, we shall employ evolutionary game theory of finite populations, where the (stochastic) dynamics proceeds in discrete time through a sequence of one-step selection-mutation processes, with mutation probability μ and selection pressure β (see below).

We further assume a well-mixed population of size Z where individuals form groups of size N and engage in a CRD with group threshold n_{pg} . Institutions are endogenously created via the contributions from pro-social individuals in each group (except Cooperators), a process that requires a minimum threshold n_I of those individuals present in each group.

Individuals adopt, each, one of the 4 following strategies: Cooperators (Cs), Punishers (Ps), rewarders (Rs) and Defectors (Ds).

Each possible configuration of the population can be represented by a state vector $\mathbf{i} = \{i_C, i_P, i_R, i_D\}$, where i_S is the number of individuals using strategy S in the population.

Individuals have an initial endowment of b and meet in groups of size N , whose composition can be specified by a corresponding state vector $\mathbf{j} = \{j_C, j_P, j_R, j_D\}$ where j_S represents the number of individuals using strategy S in the group.

Three strategies—C, R and P are pro-social, and thus individuals who employ them contribute a fraction c of their endowment b to the CRD. D, in turn, is an anti-social strategy, as individuals choose not to contribute and thus keep all of their initial endowment—see Fig. 1.

For the CRD to succeed in producing a public good, the number of pro-social individuals in a group needs to exceed a group threshold n_{pg} .

Whenever the threshold is not met, then everyone in the group will lose their endowments with a risk probability r . Therefore the base payoff related to the CRD can be written

$$\pi_0(\mathbf{j}) = b \cdot \theta(j_{CPR}; n_{pg}) + (1 - r)b \cdot \bar{\theta}(j_{CPR}; n_{pg}) \quad (1)$$

where $\theta(x; y)$ is the Heavide step function ($\theta(x; y) = 1$ if $x \geq y$ and 0 otherwise), $\bar{\theta} = 1 - \theta$ and $j_{CPR} = j_C + j_P + j_R$.

Ps and Rs contribute an additional tax, π_t , to a local electoral institution at the group level. This institution will be created and will have an impact proportional to the total contributions of pro-social individuals provided a minimum threshold of participants n_I is reached in a group of size N

$$\Delta_{RP}(j_P, j_R) = \delta \cdot \pi_t \cdot j_{PR} \cdot \theta(j_{PR}; n_I) \quad (2)$$

where $\delta \geq 1$ is a multiplication factor that reflects a possible return on the amount contributed via taxes to the formation of an institution.

Each local institution is here an electoral institution because the decision on the nature of the institution is based on a majority rule:

If the number of Ps is larger than the number of Rs then all anti-social individuals (Ds) in the group will lose some amount (same to all) totaling Δ_{RP} ; if the number of Ps is smaller than the number of Rs then Δ_{RP} will be evenly distributed by all the pro-social individuals (Cs, Ps and Rs); if a tie is reached, i.e. the number of Ps is the same as Rs, then $\Delta_{RP}/2$ is used to reward pro-social individuals and the other $\Delta_{RP}/2$ is used to punish anti-social individuals. Figure 1 illustrates the workings of the present model.

The resulting payoffs associated with each strategy read,

$$\pi_C(\mathbf{j}) = \pi_0(\mathbf{j}) - cb + \frac{\Delta_{RP}(j_P, j_R)}{j_{CPR}} \theta_{el}(j_R; j_P) \quad (3a)$$

$$\pi_P(\mathbf{j}) = \pi_C(\mathbf{j}) - \pi_t \quad (3b)$$

$$\pi_R(\mathbf{j}) = \pi_P(\mathbf{j}) \quad (3c)$$

$$\pi_D(\mathbf{j}) = (j)\pi_0(\mathbf{j}) - \frac{\Delta_{RP}(j_P, j_R)}{N - j_{CPR}} \theta_{el}(j_P; j_R) \quad (3d)$$

where $\theta_{el}(x; y)$ is a threshold function which encodes the elective institution described above

$$\theta_{el}(x; y) = \begin{cases} 0 & \text{if } x < y \\ 1/2 & \text{if } x = y \\ 1 & \text{otherwise} \end{cases} \quad (4)$$

Stochastic Evolutionary Dynamics

In the framework of Evolutionary Game Theory of finite, well-mixed populations, the fitness $f_{S_k}(\mathbf{i})$ of an individual adopting strategy S_k in a population with a configuration \mathbf{i} is given by the average payoff of strategy S_k for that particular configuration, given by

$$f_{S_k}(\mathbf{i}) = \binom{Z-1}{N-1}^{-1} \sum_{\substack{j=0 \\ j_{DCPR}=N-1}}^{j=N-1} \pi_{S_k}(\mathbf{j}) \binom{i_{S_k}-1}{j_{S_k}} \prod_{\substack{l=1 \\ l \neq k}}^4 \binom{i_{S_l}}{j_{S_l}} \quad (5)$$

where a hypergeometric sampling of the population is performed, $j_{DCPR} = j_D + j_C + j_P + j_R$ and $\mathbf{j} = A$ stands for $j_D = j_C = j_P = j_R = A$.

The evolution of the population takes place via a series of discrete one-step selection-mutation processes; as a result, at any discrete time-step, the state of the population depends only on its present configuration, which means that its evolution can be described by a Markov Process, whose probability density function (PDF) p_i satisfies the discrete Master Equation³²

$$p_i(t + \tau) - p_i(t) = \sum_{i'} \{T_{ii'} p_i(t) - T_{i'i} p_i(t)\} \quad (6)$$

where $T_{ii'}$ and $T_{i'i}$ are the transition probabilities between states \mathbf{i} and \mathbf{i}' and vice-versa respectively.

The non-diagonal transition probabilities can be readily calculated given the discrete one-step nature of the selection-mutation process,

$$T_{S_l \rightarrow S_k}(\mathbf{i}) = (1 - \mu) \frac{i_l}{Z} \frac{i_k}{Z-1} P(f_{S_k} - f_{S_l}) + \mu \frac{i_l}{(S-1)Z} \quad (7)$$

where μ is the mutation probability and the notation $T_{S_l \rightarrow S_k}(\mathbf{i})$ refers to the transition probability starting from state \mathbf{i} and having one individual with strategy S_l change its strategy to strategy S_k via the pairwise comparison rule^{33,34} $P(x) = [1 + \exp(-\beta x)]^{-1}$ which employs the Fermi function from statistical physics, where the inverse temperature β ($\beta \in \mathbb{R}_0^+$) represents here the strength of natural selection. The diagonal transition probabilities can be calculated using the relation $T_{ii} = 1 - \sum_{i' \neq i} T_{i'i}$.

We shall be interested in determining the stationary distribution (\bar{p}_i) which is given by⁸ the eigenvector associated with the highest eigenvalue (of value 1) for the transition matrix T^T : $\bar{p}_i(\mathbf{i}) = [T_{ii'}^T]^T \bar{p}_i(\mathbf{i})$.

The transition probabilities further allow us to compute the most likely path the population will follow starting from a given configuration, by means of the gradient of selection ∇_i

$$\nabla_i = \sum_{k=1}^s (T_i^{S_k+} - T_i^{S_k-}) \mathbf{u}_k \quad (8)$$

where $T_i^{S_k+}$ and $T_i^{S_k-}$ represent the probability the number of individuals adopting strategy S_k increases or decreases by one, respectively and the unit vectors \mathbf{u}_k define the basis of the phase space dynamics.

Finally, making use of the stationary distribution \bar{p}_i we can compute other quantities of interest such as η_G , the population average group achievement,

$$\eta_G = \sum_i \bar{p}_i(\mathbf{i}) a_G(\mathbf{i}) \quad (9)$$

where $a_G(\mathbf{i})$ is the so-called group achievement, computed for each population configuration \mathbf{i} by averaging over all possible group configurations for which success in the CRD is granted

$$a_G(\mathbf{i}) = \binom{Z}{N}^{-1} \sum_{\substack{j=0 \\ j_{DCPR}=N}}^{j=N} \theta(j_{CPR}; n_{pg}) \prod_{l=1}^4 \binom{i_{S_l}}{j_{S_l}} \quad (10)$$

Naturally, we may construct similar formulas to compute the incidence of reward or punishment for each configuration \mathbf{i} and subsequently compute the population average amount of reward and punishment for each set of model parameters—see SI. A freely available implementation of the general equations above can be found in the repository³⁵. For a detailed account of the methods employed see³⁶.

Data availability

All data supporting the findings of this study are available in the Plots presented within the paper and its Supplementary Information. Files with raw data used to generate the plots are available from the corresponding author on reasonable request.

Received: 14 November 2024; Accepted: 17 January 2025

Published online: 31 January 2025

References

1. Nations, U. Global issues. <https://www.un.org/en/global-issues> (2023).
2. 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. <https://doi.org/10.1073/pnas.1015648108> (2011).
3. Chen, X., Szolnoki, A. & Perc, M. Risk-driven migration and the collective-risk social dilemma. *Phys. Rev. E* **86**, 036101 (2012).
4. Abou Chakra, M. & Traulsen, A. Evolutionary dynamics of strategic behavior in a collective-risk dilemma. *PLOS Comput. Biol.* **8**, 1–7. <https://doi.org/10.1371/journal.pcbi.1002652> (2012).
5. Santos, F. C., Vasconcelos, V. V., Santos, M. D., Neves, P. & Pacheco, J. M. Evolutionary dynamics of climate change under collective-risk dilemmas. *Math. Models Methods Appl. Sci.* **22**, 1140004 (2012).
6. Hilbe, C., Abou Chakra, M., Altrock, P. M. & Traulsen, A. The evolution of strategic timing in collective-risk dilemmas. *PLoS One* **8**, e66490 (2013).
7. Vasconcelos, V. V., Santos, F. C. & Pacheco, J. M. A bottom-up institutional approach to cooperative governance of risky commons. *Nat. Clim. Chang.* **3**, 797–801. <https://doi.org/10.1038/nclimate1927> (2013).
8. Vasconcelos, V. V., Santos, F. C., Pacheco, J. M. & Levin, S. A. Climate policies under wealth inequality. *Proc. Natl. Acad. Sci.* **111**, 2212–2216 (2014).
9. Chen, X., Zhang, Y., Huang, T.-Z. & Perc, M. Solving the collective-risk social dilemma with risky assets in well-mixed and structured populations. *Phys. Rev. E* **90**, 052823 (2014).
10. Abou Chakra, M., Bumann, S., Schenk, H., Oschlies, A. & Traulsen, A. Immediate action is the best strategy when facing uncertain climate change. *Nat. Commun.* **9**, 2566. <https://doi.org/10.1038/s41467-018-04968-1> (2018).
11. Chen, X. & Szolnoki, A. Punishment and inspection for governing the commons in a feedback-evolving game. *PLoS Comput. Biol.* **14**, e1006347 (2018).
12. Góis, A. R., Santos, F. P., Pacheco, J. M. & Santos, F. C. Reward and punishment in climate change dilemmas. *Sci. Rep.* **9**, 1. <https://doi.org/10.1038/s41598-019-52524-8> (2019).
13. He, N., Chen, X. & Szolnoki, A. Central governance based on monitoring and reporting solves the collective-risk social dilemma. *Appl. Math. Comput.* **347**, 334–341 (2019).
14. Barfuss, W., Donges, J. F., Vasconcelos, V. V., Kurths, J. & Levin, S. A. Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse. *Proc. Natl. Acad. Sci.* **117**, 12915–12922 (2020).
15. Domingos, E. F. et al. Timing uncertainty in collective risk dilemmas encourages group reciprocation and polarization. *iScience* **23** (2020).
16. Couto, M. C., Pacheco, J. M. & Santos, F. C. Governance of risky public goods under graduated punishment. *J. Theor. Biol.* **505**, 1. <https://doi.org/10.1016/j.jtbi.2020.110423> (2020).
17. Dong, Y., Ma, S., Zhang, B., Wang, W.-X. & Pacheco, J. M. Financial incentives to poor countries promote net emissions reductions in multilateral climate agreements. *One Earth* **4**, 1141–1149 (2021).
18. Sun, W., Liu, L., Chen, X., Szolnoki, A. & Vasconcelos, V. V. Combination of institutional incentives for cooperative governance of risky commons. *iScience* **24**, 102844. <https://doi.org/10.1016/j.isci.2021.102844> (2021).
19. 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.* **103**, 3994–3998 (2006).
20. 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.* **105**, 2291–2294. <https://doi.org/10.1073/pnas.0709546105> (2008).
21. 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.* **108**, 11825–11829 (2011).
22. Velders, G. J. M., Andersen, S. O., Daniel, J. S., Fahey, D. W. & McFarland, M. The importance of the montreal protocol in protecting climate. *Proc. Natl. Acad. Sci.* **104**, 4814–4819 (2007).
23. Velders, G. J. M. et al. Projections of hydrofluorocarbon (hfc) emissions and the resulting global warming based on recent trends in observed abundances and current policies. *Atmos. Chem. Phys.* **22**, 6087–6101. <https://doi.org/10.5194/acp-22-6087-2022> (2022).
24. Ostrom, E. A polycentric approach for coping with climate change. *World Bank Policy Research Working Paper Series* (2009).
25. Only at low risk values one observes the occurrence of a single attractor located near the ALL-D configuration, leading to a complete demise of cooperation.
26. Chen, X., Sasaki, T., Brännström, A. & Dieckmann, U. First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation. *J. R. Soc. Interface* **12**, 20140935. <https://doi.org/10.1098/rsif.2014.0935> (2015).
27. Moreira, J. A., Pinheiro, F. L., Nunes, A. & Pacheco, J. M. Evolutionary dynamics of collective action when individual fitness derives from group decisions taken in the past. *J. Theor. Biol.* **298**, 8–15 (2012).
28. Sasaki, T., Uchida, S. & Xiaojie C. Voluntary rewards mediate the evolution of pool punishment for maintaining public goods in large populations. *Sci. Rep.* **5**, 8917 (2015).
29. Liu, L., Chen, X. & Szolnoki, A. Coevolutionary dynamics via adaptive feedback in collective-risk social dilemma game. *eLife* **12**, e82954 (2023).
30. Szolnoki, A. & Chen, X. Strategy dependent learning activity in cyclic dominant systems. *Chaos Solitons Fract.* **138**, 109935 (2020).
31. Pacheco, J. M. & Santos, F. C. Co-evolution of risk and cooperation in climate policies under wealth inequality. *PNAS Nexus* **3**, 550 (2024).
32. Risken, H. *The Fokker-Planck Equation* Vol. 18 (Springer, Berlin, 1989).
33. Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74**, 011909 (2006).
34. Traulsen, A., Pacheco, J. M. & Nowak, M. A. Pairwise comparison and selection temperature in evolutionary game dynamics. *J. Theor. Biol.* **246**, 522–529. <https://doi.org/10.1016/j.jtbi.2007.01.002> (2007).
35. Fernández Domingos, E. Egttools: Toolbox for evolutionary game theory. <https://github.com/Socrats/EGTTools>. <https://doi.org/10.5281/zenodo.3687125> (2020).
36. Fernández Domingos, E., Santos, F. C. & Lenaerts, T. Egttools: Evolutionary game dynamics in python. *iScience* **26**, 1. <https://doi.org/10.1016/j.isci.2023.106419> (2023).

Acknowledgements

This research was supported by Fundação para a Ciência e Tecnologia (FCT) of Portugal through grants PTDC/MAT-APL/6804/2020 and UIDB/50021/2020 - DOI: 10.54499/UIDB/50021/2020. F.C.S. also acknowledges

support from TAILOR, a project funded by the EU Horizon 2020 (grant 952215). E.F.D. is supported by an F.W.O. Senior Postdoctoral Grant (12A7825N).

Author contributions

All authors contributed equally to this work and all authors reviewed the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1038/s41598-025-87322-y>.

Correspondence and requests for materials should be addressed to J.M.P.

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.

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

© The Author(s) 2025

A New Electoral Bottom-Up Model of Institutional Governance

Carlos M. Garrido^{1, 2}, Francisco C. Santos^{3, 4, 2}, Elías Fernández Domingos^{5, 6}, Ana M. Nunes^{1, 7}, and Jorge M. Pacheco^{3, 2}

¹BioSystems and Integrative Sciences Institute, Faculdade de Ciências da Universidade de Lisboa Campo Grande, 1749-016 Lisboa, Portugal

²ATP-group, P-2744-016 Porto Salvo, Portugal

³INESC-ID, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal

⁴Instituto Superior Técnico, Universidade de Lisboa, 2744-016 Porto Salvo, Portugal

⁵AI lab, Computer Science Department, Vrije Universiteit Brussel, Pleinlaan 9, 3rd Floor, 1050 Brussels, Belgium

⁶MLG, Département D'Informatique, Université Libre de Bruxelles, Boulevard Du Triomphe, CP 212, 1050 Brussels, Belgium

⁷Departamento de Física, Faculdade de Ciências da Universidade de Lisboa Campo Grande, 1749-016 Lisboa, Portugal

Supplementary Information

In the following, we explore the impact of changing the different model parameters. The results are quite intuitive, in the sense that they correspond to what one expects from general threshold public goods games:

Increasing population size Z and selection pressure β will act to sharpen the transition from defection to cooperation as a function of risk, as shown in **SI** Figs. 1 and 2, respectively. Correspondingly, larger group sizes N will require a higher risk for the same transition to occur, as shown in **SI** Fig. 3, the same happening if, as shown in **SI** Fig. 4, one increases the public good threshold n_{pg} . Finally, in what concerns the institution parameters (n_I, π_i and δ) we show in **SI** Fig. 5 how increasing the institutional threshold n_I acts to increase the risk at which the transition from defection to cooperation takes place. This transition, in turn, is affected in different ways as we change the individual taxes π_i incurred by both P s and R s, as shown in **SI** Fig. 6 (maintaining all other model parameters invariant). Indeed, for very small values of π_i , small tax increases will be beneficial to overall cooperation. Further increases, however, inevitably become detrimental to overall cooperation. Note, however, that high values of π_i are generally beneficial to cooperation at low-risk. Finally, as shown in **SI** Fig. 7, increasing the institutional return δ helps cooperation to emerge at a lower risk.

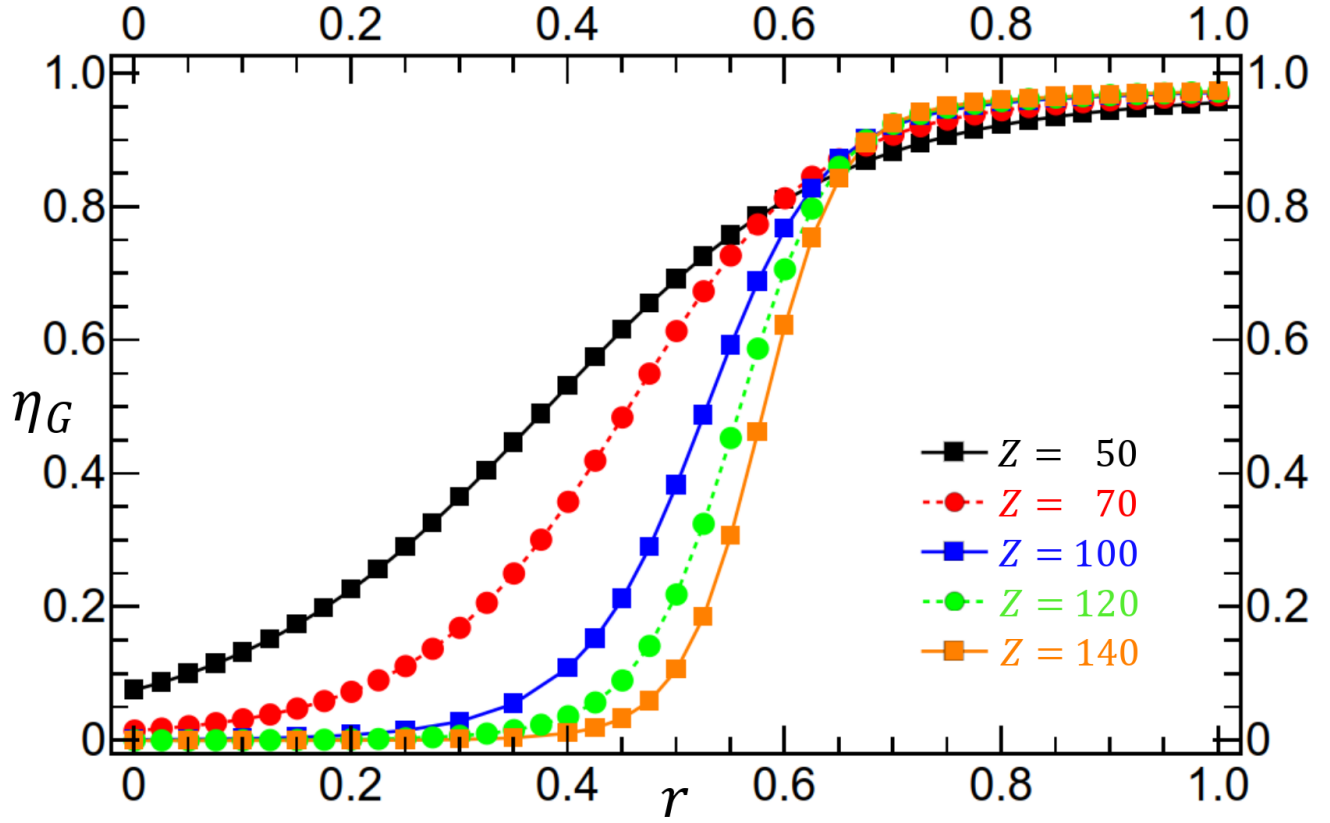


Figure 1. η_G versus risk for different population sizes. We changed population size Z maintaining all other model parameters constant (see below). With increasing Z one observes that the transition from defection ($\eta_G \rightarrow 0$) to cooperation ($\eta_G \rightarrow 1$) occurs in an increasingly narrower interval of risk. At the same time, the value of risk at which $\eta_G = 0.5$ increases. In the following figures, we shall use the value $Z = 100$ as the reference population size. As the figure shows, this value shows an intermediate behavior between small and very large population sizes, retaining the computational feasibility of the calculations, which become prohibitively expensive for large populations. Parameter values used: $b = 1$, $c = 0.1$, $\mu = 1/Z$, $\beta = 2$, $N = 8$, $n_{pg} = 6$, $n_I = 2$, $\pi_t = 0.03$, $\delta = 2$.

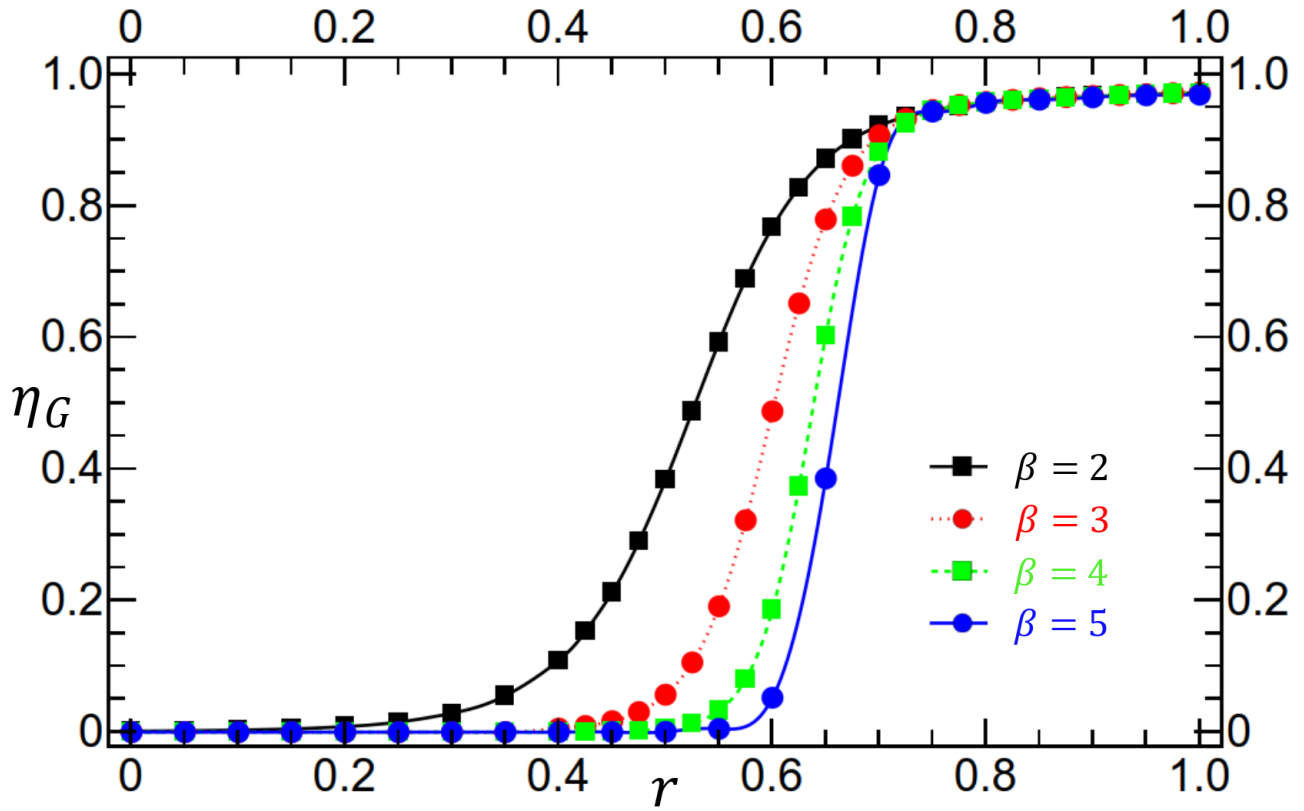


Figure 2. η_G versus risk for different values of the selection pressure β . We changed the selection pressure β while maintaining all other model parameters constant (see below). Similar to changing population size Z , increasing β sharpens the transition from defection ($\eta_G \rightarrow 0$) to cooperation ($\eta_G \rightarrow 1$), also shifting to higher values the value of risk at which $\eta_G = 0.5$. Parameter values used: $b = 1$, $c = 0.1$, $Z = 100$, $\mu = 1/Z$, $\beta = 2$ $N = 8$, $n_{pg} = 6$, $n_I = 2$, $\pi_t = 0.03$, $\delta = 2$.

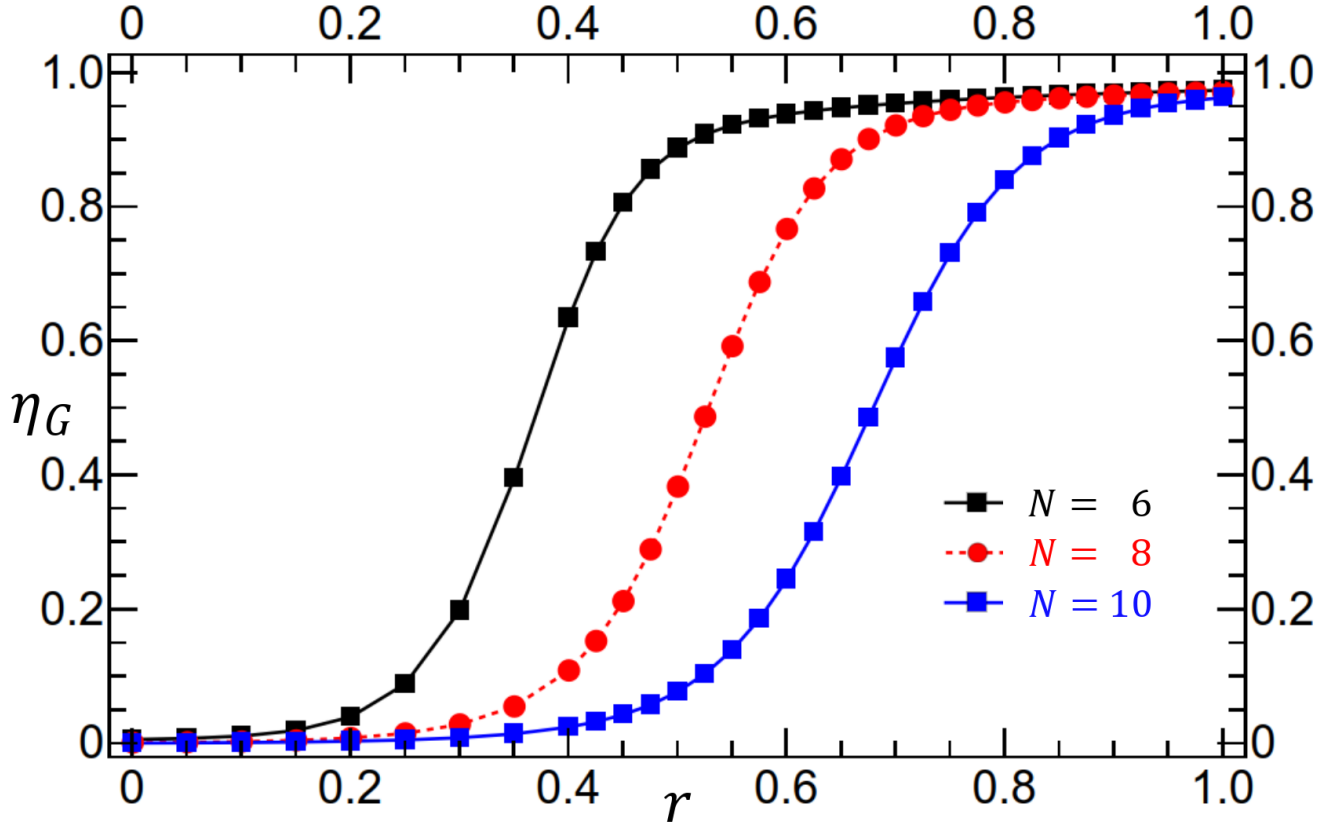


Figure 3. η_G versus risk for different values of the group size N . We changed group size N maintaining all other model parameters constant except n_{pg} (see below), which we also changed to keep the ratio n_{pg}/N in a range of comparative values. This means that we use $n_{pg} = 4$ for $N = 6$ ($n_{pg}/N = 0.67$), $n_{pg} = 6$ for $N = 8$ ($n_{pg}/N = 0.75$) and $n_{pg} = 8$ for $N = 10$ ($n_{pg}/N = 0.80$). With increasing N (and, correspondingly, n_{pg}) one observes that the transition from defection ($\eta_G \rightarrow 0$) to cooperation ($\eta_G \rightarrow 1$) occurs at increasingly higher values of risk. This is a general feature of threshold Public Goods Games that is retained in the present model.

Parameter values used: $b = 1$, $c = 0.1$, $Z = 100$, $\mu = 1/Z$, $\beta = 2$, $n_I = 2$, $\pi_t = 0.03$, $\delta = 2$.

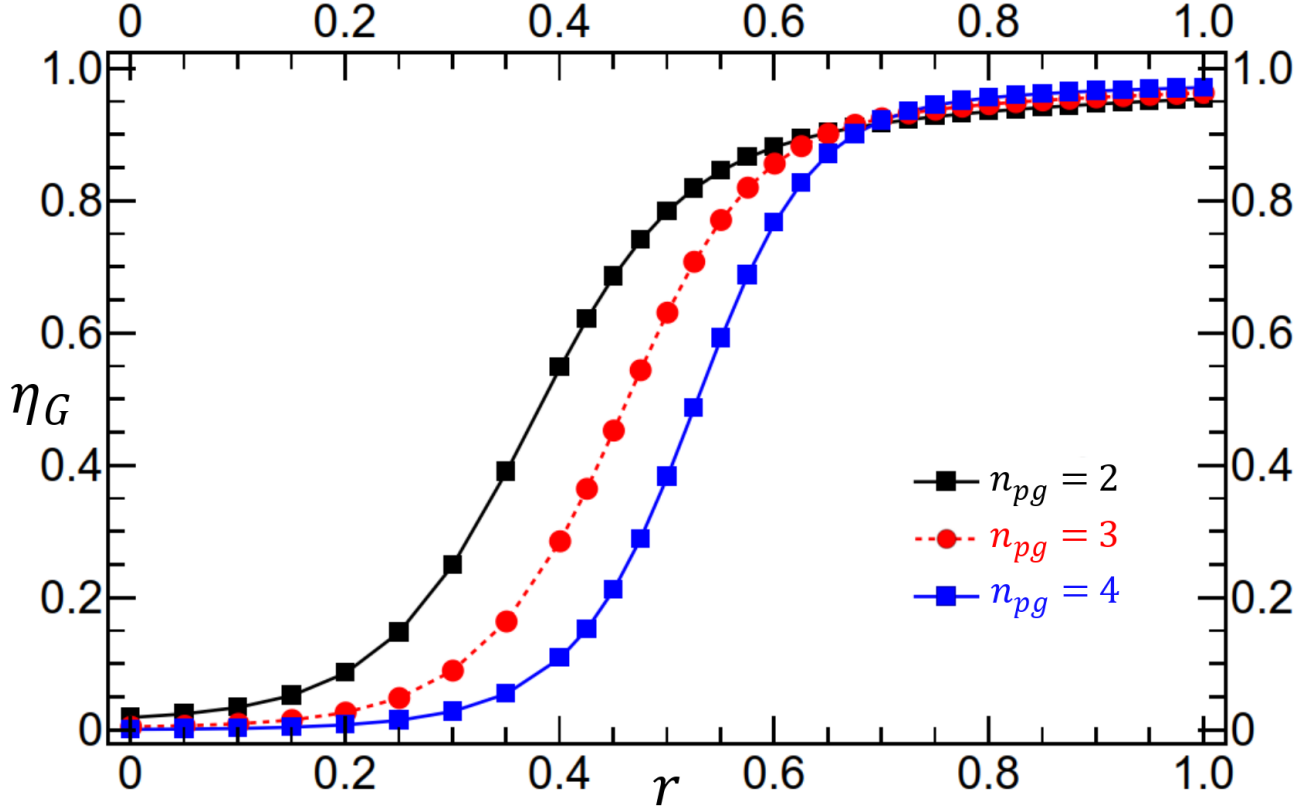


Figure 4. η_G versus risk for different group threshold values. We changed the group threshold value n_{pg} maintaining all other model parameters constant (see below). With increasing n_{pg} one observes that the transition from defection ($\eta_G \rightarrow 0$) to cooperation ($\eta_G \rightarrow 1$) occurs at increasingly higher values of risk. This is a general feature of threshold Public Goods Games that is retained in the present model.

Parameter values used: $b = 1$, $c = 0.1$, $Z = 100$, $\mu = 1/Z$, $\beta = 2$, $N = 8$, $n_{pg} = 6$, $n_I = 2$, $\pi_t = 0.03$, $\delta = 2$.

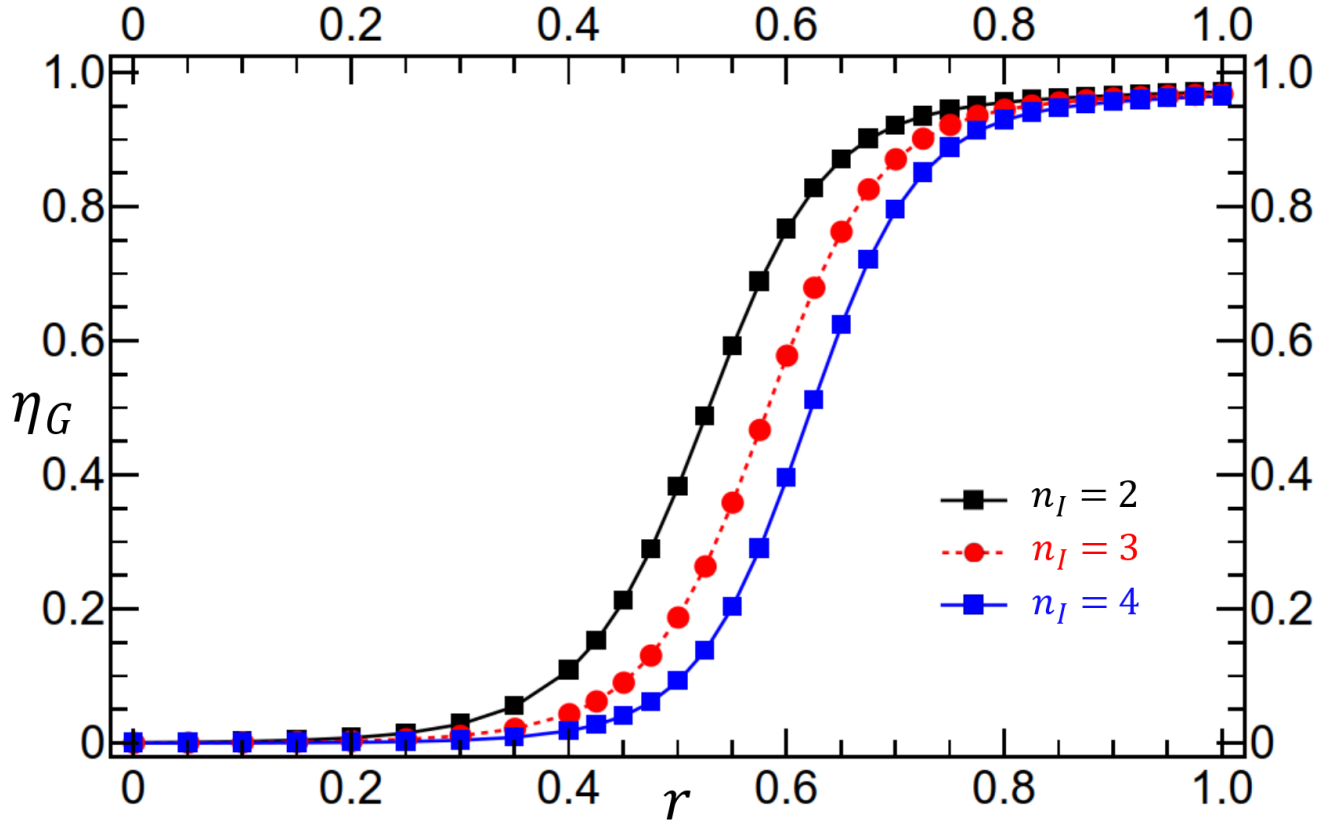


Figure 5. η_G versus risk for different institutional threshold sizes. We changed institutional threshold (n_I) value maintaining all other model parameters constant (see below). With increasing n_I one observes that the transition from defection ($\eta_G \rightarrow 0$) to cooperation ($\eta_G \rightarrow 1$) occurs at increasingly higher values of risk, as one would expect. Parameter values used: $b = 1$, $c = 0.1$, $Z = 100$, $\mu = 1/Z$, $\beta = 2$, $N = 8$, $n_{pg} = 6$, $\pi_t = 0.03$, $\delta = 2$.

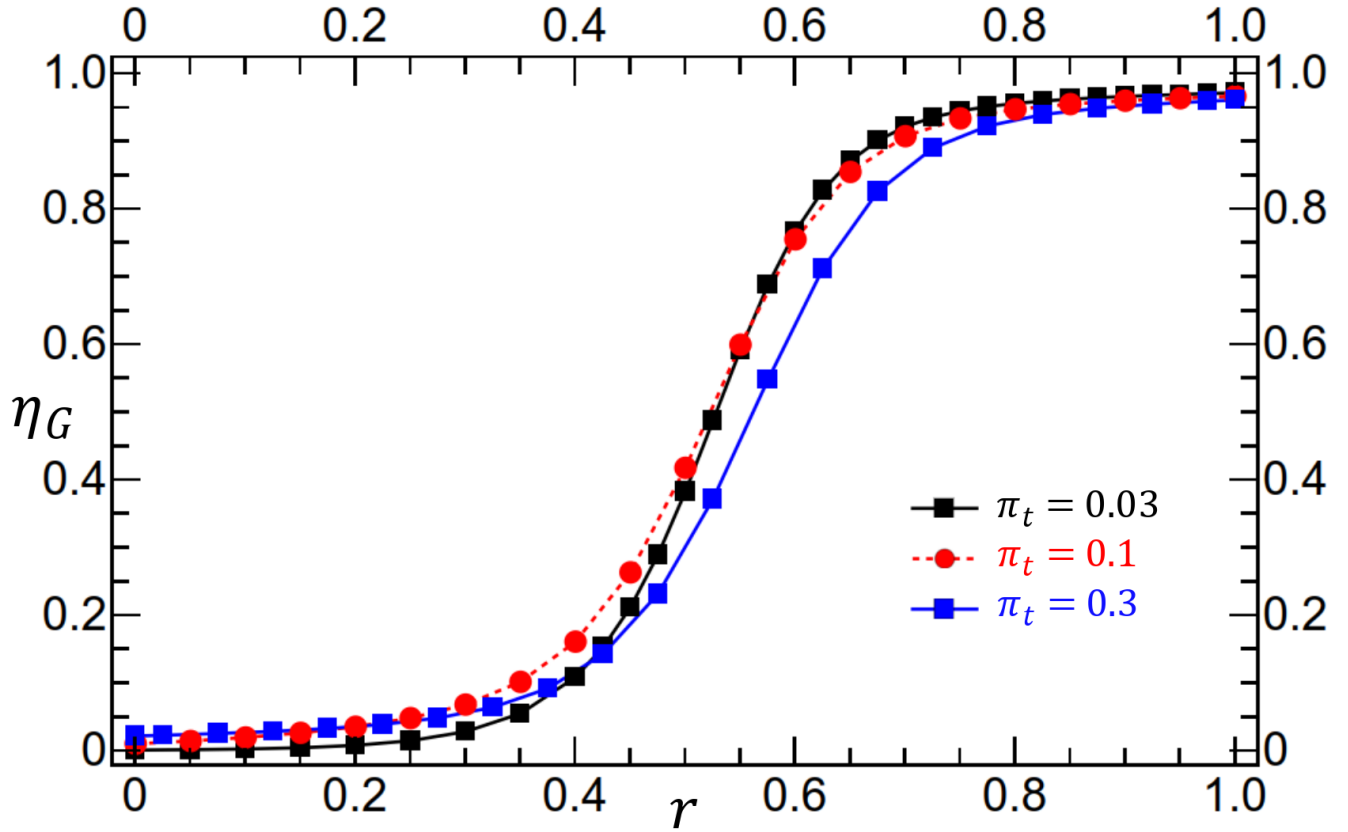


Figure 6. η_G versus risk for different individual institutional tax values. We changed the amount that pro-institutional individuals (P s and R s) must contribute to create an institution at the group level (π_t). We maintained all other model parameters constant (see below). The lowest value plotted ($\pi_t = 0.03$) corresponds to the most used value throughout the manuscript. For smaller values of π_t (not shown) the transition from defection ($\eta_G \rightarrow 0$) to cooperation ($\eta_G \rightarrow 1$) occurs for slightly larger values of risk compared to $\pi_t = 0.03$. In other words, increasing π_t is beneficial to cooperation for low values of π_t . With increasing π_t we still observe an overall benefit regarding the emergence of cooperation (see the results for $\pi_t = 0.1$), but this benefit disappears when the amount contributed via π_t becomes too large, as shown in the figure for an order of magnitude increase of π_t . This said, increasing taxes is beneficial at very low risk, being detrimental at high risk. Parameter values used: $b = 1$, $c = 0.1$, $Z = 100$, $\mu = 1/Z$, $\beta = 2$, $N = 8$, $n_{pg} = 6$, $n_I = 2$, $\delta = 2$.

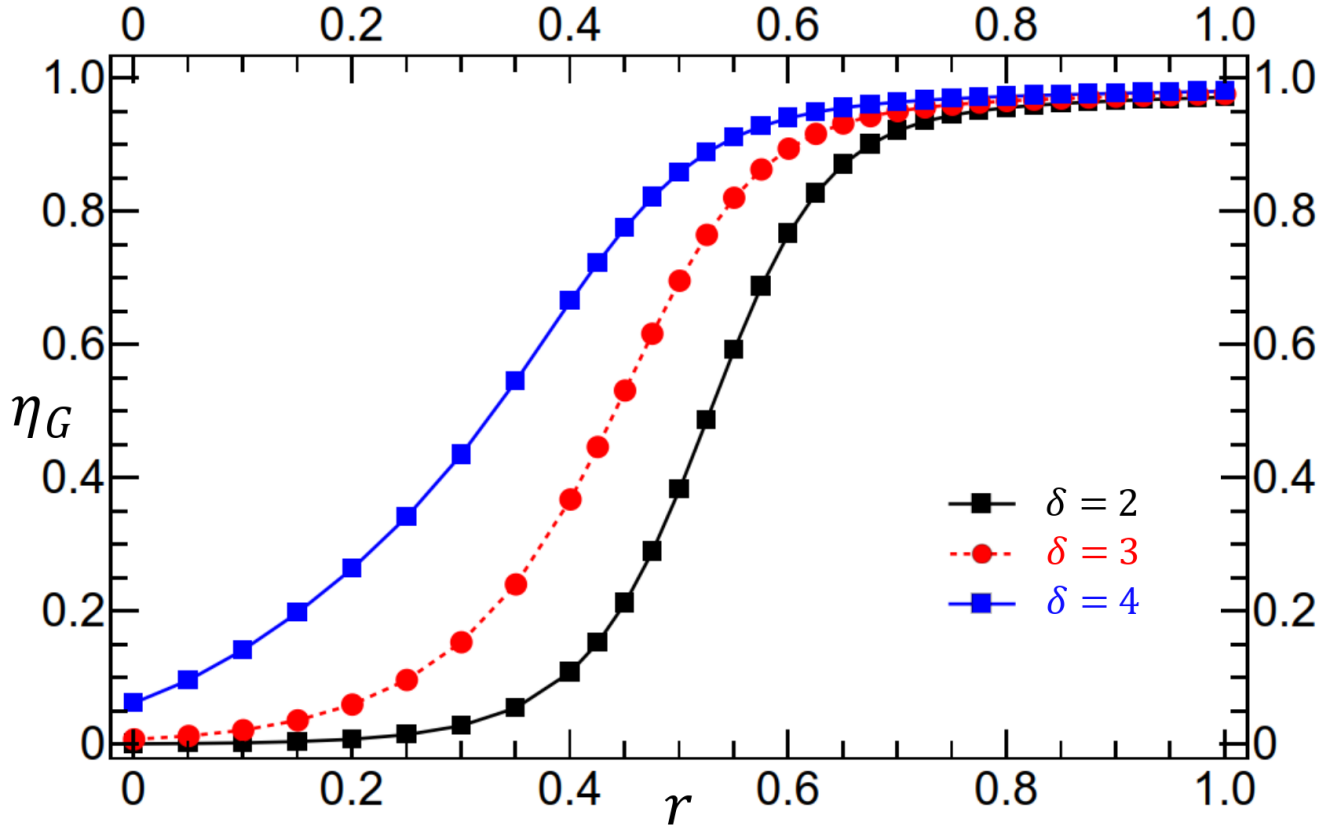


Figure 7. η_G versus risk for different institutional return values δ . We changed the value of the institutional return δ maintaining all other model parameters constant (see below). With increasing δ one observes that, as expected the positive role of institutions is greatly enhanced, sizably reducing the risk value at which $\eta_G = 0.5$ occurs. Parameter values used: $Z = 100$, $\mu = 1/Z$, $\beta = 2$, $N = 8$, $n_I = 2$, $b = 1$, $c = 0.1$, $\pi_r = 0.03$, $\delta = 2$.

Stationary Distributions

In SI Fig. 8 we plot the stationary distribution associated with a particular example, at high risk ($r = 0.7$) where overall cooperation dominates the dynamics. For $S = 4$ strategies, the visualization of the stationary distribution is non-trivial, and therefore we divide the figure in three panels, plotting the stationary distribution in domains of increasingly lower dimensionality.

In panel **a**) we show the full simplex, choosing a perspective that best illustrates the accumulation of strength in the vicinity of the **D-C-R** plane, although it is clear that, in the vicinity of the cooperative attractor, the stationary distribution extends somewhat inside the simplex, towards the "P"-vertex. This becomes clearly visible if we make a triangular cut of the simplex at a constant value of number of **Ds** in the population, in this case $i_D = 9$ (the number of **Ds** at which the maximum of the stationary distribution occurs). The result is shown in panel **b**), for the triangle joining the configurations $\mathbf{R}^* \rightarrow i = (09, 00, 00, 61)$, $\mathbf{C}^* \rightarrow i = (09, 61, 00, 00)$ and $\mathbf{P}^* \rightarrow i = (09, 00, 61, 00)$, also illustrated in panel **a**) by means of a blue triangle, showing how strength penetrates into the interior of the simplex. Finally, panel **c**) shows the relative intensity of the two attractors referred in the main text – the cooperative attractor at high number of **Cs** and the defective attractor at high number of **Ds**. We thus plot the values of the stationary distribution along the edge **D-C**, illustrating the fact that the population spends more time in the vicinity of the cooperative attractor than in the vicinity of the other attractor.

SI Fig. 9 shows, for the parameters indicated, all configurations with a probability of reward (in green) and punishment (in red) greater than 90%; This three-dimensional distribution encompasses the **P-R** edge, along which we plot the probability distributions to punish (in red) to reward (in green) as well as to punish-and-reward (in black) along this edge. Note that, similarly to what was done when computing η_G , we can also investigate the behavior of the population average probabilities of reward and punishment. Indeed, we can compute the population average probabilities

$$\mathbf{Prob}(P) = \sum_i \bar{p}_i(i) a_P(i) \quad (1a)$$

$$\mathbf{Prob}(R) = \sum_i \bar{p}_i(i) a_R(i) \quad (1b)$$

where

$$a_P(i) = \sum_{(j: j_k=0)}^{(j: j_k=N)} \theta(j_{PR}; n_I) \cdot \theta(P; R) \cdot H(j; i, Z, N) \quad (2a)$$

$$a_R(i) = \sum_{(j: j_k=0)}^{(j: j_k=N)} \theta(j_{PR}; n_I) \cdot \theta(R; P) \cdot H(j; i, Z, N) \quad (2b)$$

In SI Fig. 10 we plot $\mathbf{Prob}(R)$, $\mathbf{Prob}(P)$ and the ratio $\mathbf{Prob}(P)/\mathbf{Prob}(R)$ as a function of risk. Clearly, the reward probability is always larger than punishment, and they both grow with increasing risk. However, as risk increases, the ratio decreases and stabilizes as η_G approaches 1. The figure clearly shows that, for any value of risk, both reward and punishment are important and contribute to the success of the electoral model.

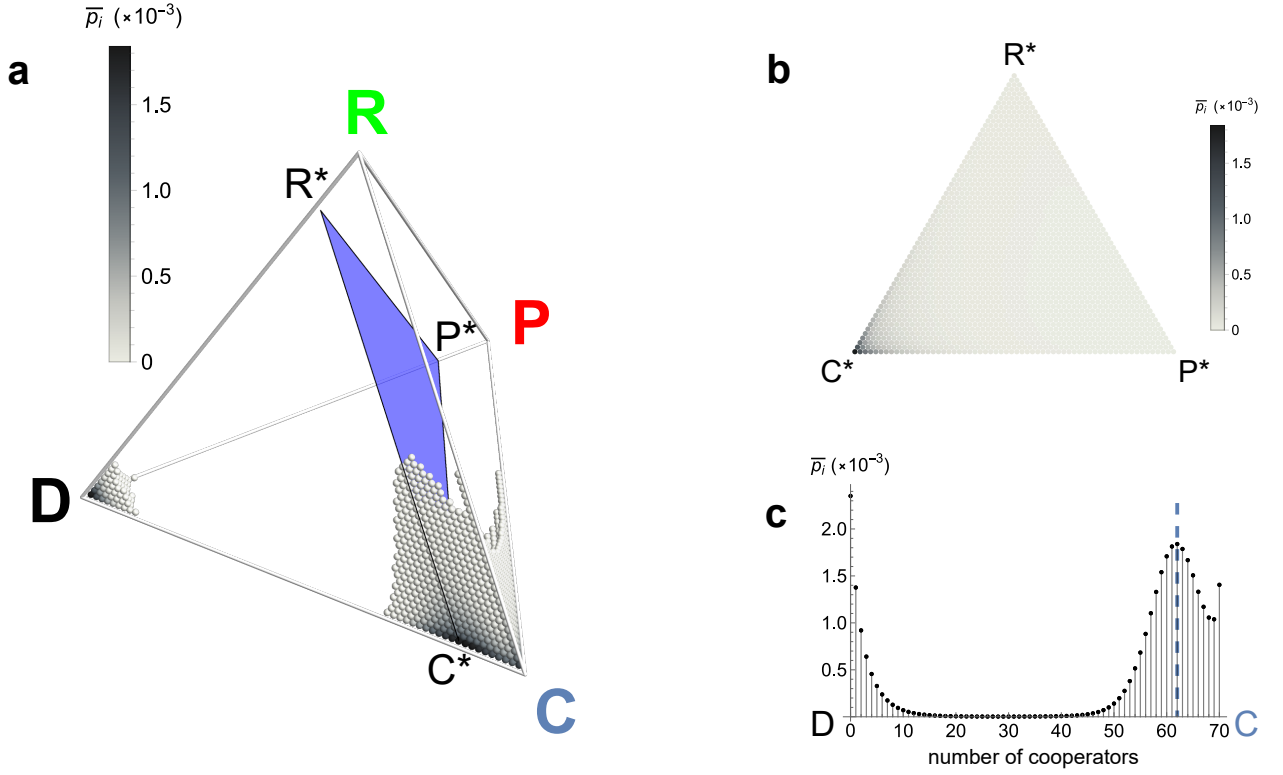


Figure 8. Stationary distribution for a specific set of model parameters. We plot the stationary distribution \bar{p}_i (see Methods in main text for details) for: **a)** All possible configurations associated with a finite population of size $Z = 70$ (see model parameters at the end). Since we selected $r = 0.7$, the cooperative attractor associated with a high number of Cs dominates the stochastic evolutionary dynamics, indicating that the population spends more time in its vicinity. Furthermore, the picture also illustrates the fact that the stationary distribution is mostly concentrated on the **D-C-R** planar face of the tetrahedron, although it is also apparent that it extends towards the interior of the simplex. Panel **b)** illustrates the extent to which the stationary distribution penetrates into the interior of the simplex, by defining the blue triangle shown in panel **a)** that joins the configurations $R^* \rightarrow i = (08, 00, 00, 62)$, $C^* \rightarrow i = (09, 62, 00, 00)$ and $P^* \rightarrow i = (09, 00, 62, 00)$, characterized by, in all cases, the number of **Ds** remaining constant and equal to 9, precisely the value at which the stationary distribution is maximal in the vicinity of the cooperative attractor.

Finally, panel **c)** shows the stationary distribution along the edge "D-C", illustrating the fact that the population spends more time in the vicinity of the cooperative attractor than in the vicinity of the defective attractor.

Parameter values used: $b = 1$, $c = 0.1$, $r = 0.7$, $Z = 70$, $\mu = 1/Z$, $\beta = 5$, $N = 8$, $n_{pg} = 6$, $n_l = 2$, $\pi_t = 0.03$, $\delta = 2$.

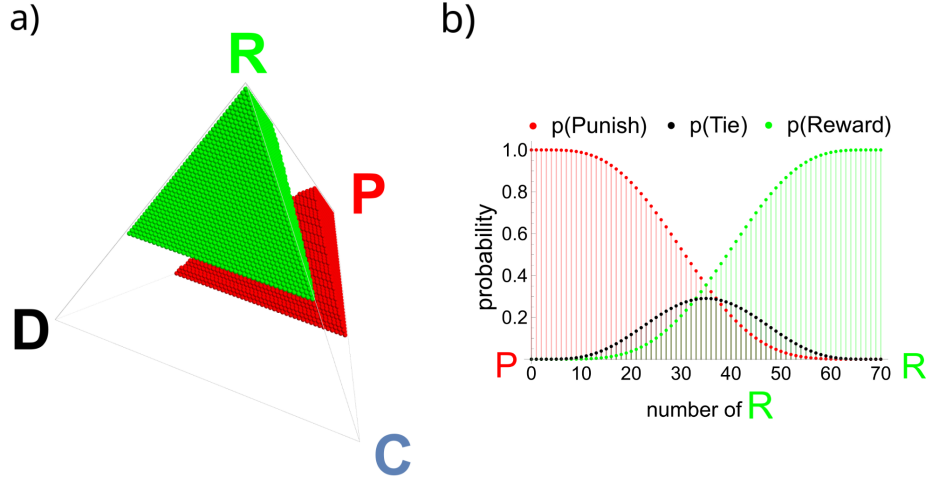


Figure 9. Probability of each incentive type for a specific set of model parameters. We plot: **a)** All the configurations with a probability of reward (in green) and punishment (in red) higher than 90% associated with a finite population of size $Z = 70$ (see model parameters at end). Panel **b)** shows the probability that each type of incentive will be applied as a function of population configuration along the **P-R** edge. $p(\text{Tie})$ represents the probability that a tie happens in a group ($j_P = j_R$) for a given configuration of the population. Parameter values used: $Z = 70$, $N = 8$, $n_{pg} = 6$, $n_I = 2$.

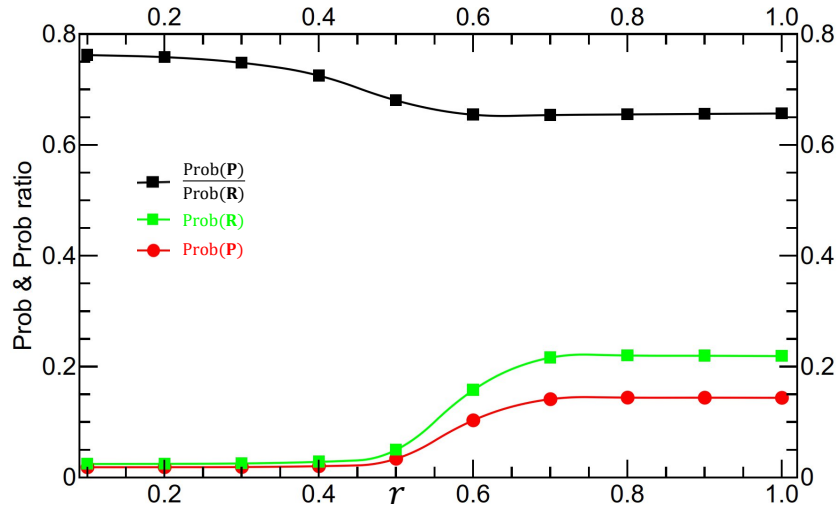


Figure 10. Probability of reward, punishment and their ratio as a function of risk. We plot the probability of punishment (red line with solid circles), of reward (green line with solid squares) as well as the ratio between these two probabilities (black line with solid squares) as a function of risk. Both $\text{Prob}(R)$ and $\text{Prob}(P)$ exhibit a similar behavior with risk, stabilizing at high risk, when η_G approaches 1. However, for all values of risk we have that $\text{Prob}(R) > \text{Prob}(P)$, although the ratio decreases with increasing risk. Same parameters as in Figure 3 of main text.