nature climate change

LETTERS

PUBLISHED ONLINE: 14 JULY 2013 | DOI: 10.1038/NCLIMATE1927

A bottom-up institutional approach to cooperative governance of risky commons

Vítor V. Vasconcelos[1,2,3], Francisco C. Santos[1,3] and Jorge M. Pacheco[1,4,5]*

Avoiding the effects of climate change may be framed as a public goods dilemma[1], in which the risk of future losses is non-negligible[2–7], while realizing that the public good may be far in the future[3,7–9]. The limited success of existing attempts to reach global cooperation has been also associated with a lack of sanctioning institutions and mechanisms to deal with those who do not contribute to the welfare of the planet or fail to abide by agreements[1,3,10–13]. Here we investigate the emergence and impact of different types of sanctioning to deter non-cooperative behaviour in climate agreements. We show that a bottom-up approach, in which parties create local institutions that punish free-riders, promotes the emergence of widespread cooperation, mostly when risk perception is low, as it is at present[3,7]. On the contrary, global institutions provide, at best, marginal improvements regarding overall cooperation. Our results clearly suggest that a polycentric approach involving multiple institutions is more effective than that associated with a single, global one, indicating that such a bottom-up, self-organization approach, set up at a local scale, provides a better ground on which to attempt a solution for such a complex and global dilemma.

To investigate the role of sanctioning institutions, let us consider a finite (and small[1,3]) population of size Z where individuals interact through what has been coined the collective-risk dilemma (CRD), a threshold public goods game—akin to an N-person stag-hunt or coordination game[14]—that mimics the problem at stake[2–4,6]. Individuals organize groups of size N, in which each participant may act as a cooperator (C), defector (D) or punisher (P). Each individual starts with an initial endowment or benefit b. Cs and Ps contribute a fraction c of this benefit, the cost, to reach a common goal, whereas Ds do not contribute. If the overall contribution in the group is insufficient—that is, if the joint number of Cs and Ps in the group is below n_{pg}—everyone in that group will lose their remaining endowments with a probability r (here understood as the perception of risk of collective disaster[2]); otherwise, everyone will keep whatever they have.

The scenario of present-day summits, in which all countries meet in a single group with the aim of establishing long-term goals and commitments by which all must abide[3], is known to be detrimental to cooperation[6]. Hence, it is better to establish smaller groups focused on overcoming shorter-term goals, meant to be revised and reassessed frequently. To this end, we model individual decision-making as an interacting dynamical process, where individuals are embedded in a behavioural ecosystem[15–17], such that decisions and achievements of others influence one's own decisions through time[18–21] (Methods and Supplementary Information for further details). Behavioural experiments[4,5,22], as well as other theoretical models[23,24], have implemented thresholds through repeated interactions, and other authors have highlighted the role played by pledges and communication during negotiations[1,5,25], bringing about additional layers of complexity to this problem (details and comparison with other models in the Supplementary Information).

Besides contributing to this public good, Ps also contribute with a punishment tax (π_t) to an institution that, whenever endowed with enough funding ($n_p\pi_t$) will effectively punish Ds by an amount Δ. Hence, establishing an institution stands as a second-order public good[17,20], which is only achieved above a certain threshold number of contributors n_p (ref. 14). The fact that, in both cases, contributors may pay a cost in vain increases the realism (and the inherent complexity) of the decision process modelled here.

The institution need not be a global one (such as the United Nations)—supported by all Ps in the population—that overviews all group interactions in the population. Institutions may also be local, group-wide, created by Ps within each group to enforce cooperation in that group of individuals. Here we shall consider both cases.

In the absence of Ps, this model reduces to the evolutionary game theoretical model[6] developed to investigate the general role of risk in climate change agreements, and inspired in economic experiments[4] that provided evidence on the unavoidable role of risk perception in the context of climate change. Indeed, the theoretical model not only corroborates the results of the economic experiments[4], but also allows one to extend the analysis to arbitrary group size, risk perception and even group-networked agreements[6]. The new, fundamental changes stemming from the introduction of Ps in this behavioural ecosystem will allow us to assess the role of sanctioning institutions in the presence of risk, a feature that has not been studied before, neither theoretically nor experimentally.

The stochastic evolutionary dynamics of the population occurs in the presence of errors, both in terms of errors of imitation[21] and in terms of behavioural mutations[26], the latter accounting for a free exploration of the possible strategies. We calculate the pervasiveness in time of each possible behavioural composition of the population, the so-called stationary distribution (Methods), which allows the computation of the average fraction of groups that successfully produce (or maintain) the public good—a quantity we designate as group achievement, η_G—and the prevalence in time of a given type of institution—that is, the fraction of time the population witnesses the presence of sanctioning institutions (local or global)—a quantity we designate as institutions prevalence, η_I. It is important to note that both quantities can be directly

[1]ATP-Group, CMAF, Instituto para a Investigação Interdisciplinar, P-1649-003 Lisboa, Portugal, [2]Centro de Física da Universidade do Minho, 4710-057 Braga, Portugal, [3]INESC-ID and Instituto Superior Técnico, Universidade Técnica de Lisboa, IST-Taguspark, 2744-016 Porto Salvo, Portugal, [4]Centro de Biologia Molecular e Ambiental, Universidade do Minho, 4710-057 Braga, Portugal, [5]Departamento de Matemática e Aplicações, Universidade do Minho, 4710-057 Braga, Portugal. *e-mail: pacheco@cii.fc.ul.pt

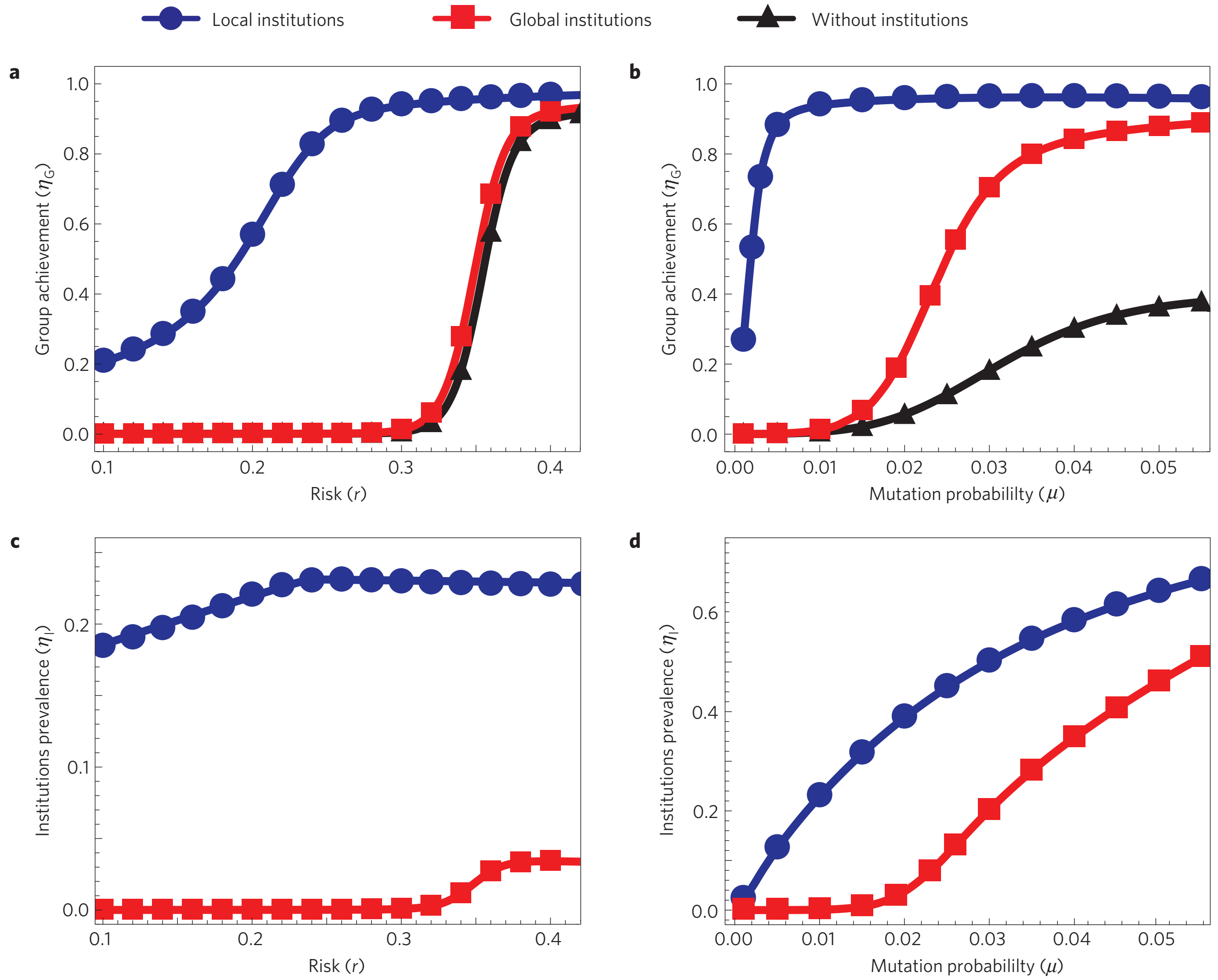

Figure 1 | Group achievement η_G and institutions prevalence η_I. a,b, The average fraction of groups that attain the public good (η_G) as a function of perception of risk (r; **a**) and behavioural exploration probability (μ; **b**). Sanctions are enacted by a global institution (red lines and squares) or by local institutions (blue lines and circles). Black lines and triangles: results obtained in the absence of any institution. **c**,**d**, Results for η_I as a function of risk r (**c**) and exploration μ (**d**). Unlike global institutions, often associated with marginal improvements of cooperation, local institutions promote group coordination to avoid a collective disaster, mostly for low perception of risk. The coordination threshold n_{pg} is set to 75% of the group size, whereas local (global) institutions are created whenever 25% of the group (population) contributes to its establishment. Other parameters: $Z = 100$, $N = 4$, $c/b = 0.1$, $\mu = 1/Z$, $\pi_f = 0.3$, $\pi_t = 0.03$ and $r = 0.3$.

compared with data extracted from experiments[2,4]. In particular, the empirical results obtained for the risk dependence[4] (in the absence of any sanctioning) show that the group achievement (η_G in our model) increases with the value of risk, correlating nicely with the dependence shown in Fig. 1a with black lines and symbols.

In Fig. 1a the behaviour of η_G as a function of risk is shown in the absence of any institutions (in black), under one global institution (in red) and under local institutions (in blue). Comparison of the black and red curves shows that global institutions provide, at best, a marginal improvement compared with no institutions at all. This result is surprising, given that most climate agreements attempt to involve all countries at once[1,3,27], in which case a single, global institution constitutes the most natural candidate (further details in the Supplementary Information).

On the contrary, under local, group-wide, sanctioning institutions, associated with a distributed scenario in which global sanctions will result from the joint role of a variety of institutions, group achievement is substantially enhanced, in particular when it is most needed: for low values of the perception of risk and whenever individuals face stringent requirements to avoid a collective disaster (Fig. 1a), as has been pointed out to be the case in the context of climate treaties[1]. One can also show (Supplementary Information) that this result is even more pronounced in a scenario encompassing (many) small groups (and institutions). This aspect is particularly important, as the group size (N) defines both the scale at which agreements should be attempted and the overall scope of each institution.

The success of local institutions is closely connected with their resilience. As shown in Fig. 1c, local institutions prevail for longer periods than a (single) global one, always promoting more widespread cooperation than global ones. The efficiency and prevalence of both kinds of institution, however, can be significantly enhanced for high behavioural mutations (Fig. 1b,d), associated with situations in which participants change their decisions more frequently. This scenario may be relevant, given the multitude of (often conflicting) factors that contribute to the process of decision-making[12,13,19].

The dynamics associated with each type of institution is best characterized by the full stationary distributions, plotted in Figs 2 and 3 and covering the entire configuration space mapped onto the triangular simplexes, in which each (discrete) configuration is represented by a circular dot. Darker dots indicate those configurations visited more often, according to the colour gradient scale indicated in each panel. In each dot the relative frequencies

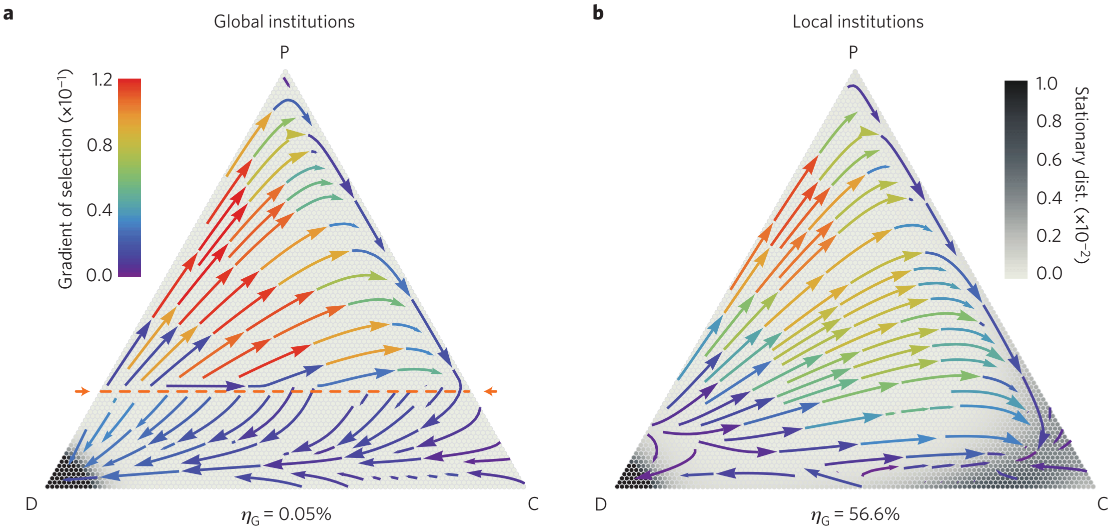

Figure 2 | Behaviour of the CRD in the full configuration space with three strategies—Cs, Ps and Ds—for the same parameters as in Fig. 1 and low risk ($r = 0.2$). a, Global institutions. **b**, Local institutions. The value of the stationary distribution at each configuration is shown using a grey scale; the magnitude of the gradient of selection is shown using the blue–yellow–red scale indicated. For global institutions, the population-wide threshold is indicated using a dashed orange line.

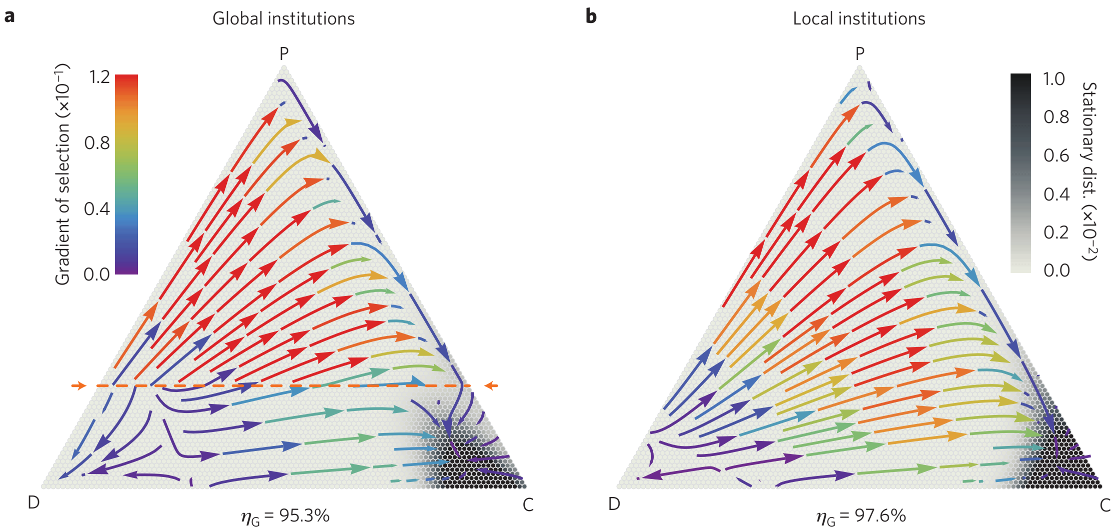

Figure 3 | Behaviour of the CRD in the full configuration space with three strategies: Cs, Ps and Ds. a, Global institutions. **b**, Local institutions. We use the same parameters as in Fig. 2, yet for high values of the perception of risk ($r = 0.5$, in this case), that is, a value of r above which most of the groups manage to coordinate their action, even in the absence of institutions (Fig. 1a).

of Cs, Ds and Ps sum up to one, whereas each vertex of the triangle is associated with monomorphic configurations. Arrows in each simplex represent the most probable direction of evolution, obtained from the computation of the two-dimensional gradient of selection (Methods). We used a continuous colour code associated with the likelihood of such transitions.

The two panels of Fig. 2 show representative examples of the behavioural dynamics of Cs, Ds and Ps under global institutions (Fig. 2a) and local institutions (Fig. 2b), for low values of the perception of risk ($r = 0.2$). For global institutions (Fig. 2a) and whenever the population starts below n_P (the punishment or institution threshold value indicated by a horizontal, orange dashed line), behavioural mutations allow the appearance of Ps in the population (Supplementary Information for further details), such that whenever the composition of the population lies above the threshold line, Ps rapidly outcompete Ds (see arrows), leading the population towards full cooperation, associated with the CP-edge of the simplex. Once in this situation, however, Ps will be outcompeted by Cs as now they contribute to support an institution that has become useless. Hence, the global institution becomes unstable, leading the population (slowly, as shown by the blue arrows along the whole path) to a configuration that falls below the threshold line again. Thus, for low perception of risk, a global institution cannot be maintained for long periods (Fig. 1c) and, as shown by the stationary distributions, the population will remain most of the time under widespread defection. This, in turn, leads to the small value of η_G reported in Fig. 2a.

For local institutions, however, the situation is quite different, as shown in Fig. 2b. Comparison between Fig. 2a and Fig. 2b shows that the role of the threshold line is not so pronounced in this case. Considering that we need the same fraction of Ps (compared to Fig. 2a) to make the institution efficient (25% in this example), but now at the level of the group (and no longer at the level of the population), it is possible that some (although not all) groups have enough Ps for sanctions to become effective. This leads to a marked increase of η_G, as in this case the population evolves towards regimes of widespread cooperation. This happens because the population will stabilize in configurations comprising a sizeable amount of Cs together with enough Ps to prevent Ds from invading. The fact that this happens for low values of risk r is important, given that, at present, the perception of risk regarding climate issues is low[3,7].

For high values of the perception of risk, shown in Fig. 3 ($r = 0.5$), both local and global institutions marginally enhance the positive prospects for cooperation already attained in the absence of any institution, as for high risk the dynamics occurs in the vicinity of the CD-edge of the simplex (Supplementary Information). Notwithstanding, and because local institutions are easier to emerge, they work as catalysers of collective action, while helping to prevent the invasion of Ds, as shown in Fig. 2. Neither local nor global institutions are robust to free-riding, a result that has been recently confirmed experimentally[28]. Finally, behavioural mutations enhance the prevalence of configurations in the inner part of the simplex, which in turn increases the chances of having enough Ps to establish institutions and cooperation, as previously shown in Fig. 1.

Our results support the conclusion that a decentralized, polycentric, bottom-up approach[10], involving multiple institutions instead of a single global one, provides better conditions both for cooperation to thrive and for ensuring the maintenance of such institutions. This result is particularly relevant whenever perception of risk of collective disaster, alone, is not enough to ensure global cooperation. In this case, local sanctioning institutions may provide an escape hatch to the tragedy of the commons humanity is facing. In this context, it is worth stressing that the mechanisms discussed here operate optimally whenever groups are small. Present-day local initiatives, such as the Western Climate Initiative[29], have started with a small group of US states. As time went by, the Western Climate Initiative group size has grown to include additional Canadian states and Mexican provinces. Although the reasons and motivations for such an evolution are comprehensible, one should not overlook that larger groups are more difficult to coordinate into widespread cooperation (Supplementary Information). Similar dynamics, in which cooperation nucleating in a small group expands into a larger and larger group, can be found in policies beyond climate governance with mixed results, from the major transitions in evolution[30] to the recent evolution of the European Union, stressing the common ground shared by governance and a variety of ecosystems[15]. In this context, it might be easier to seek a multi-scale (and multi-step) process, in which coordination is achieved in multiple small groups or climate blocks[12], before aiming, if needed, at agreements encompassing larger groups (or, alternatively, inter-group agreements). Hence, although most causes of climate change result from the combined action of all inhabitants of our planet, the solutions for such complex and global dilemma may be easier to achieve at a much smaller scale[10]. In light of our results, the widely repeated motto 'Think globally, act locally' would hardly seem more appropriate.

Methods

We consider a population of Z individuals, who set up groups of size N, in which they engage in the CRD public goods game[4,6], being capable of adopting one of the three strategies: C, P and D. Following the discussion in the main text, the payoff of an individual playing in a group in which there are j_C Cs, j_P Ps and $N - j_P - j_C$ Ds, can be written as $\Pi_C = -c + b\Theta(j_C + j_P - n_{pg}) + (1-r)b[1 - \Theta(j_C + j_P - n_{pg})]$, $\Pi_P = \Pi_C - \pi_t$ and $\Pi_D = \Pi_C + c - \Delta$ for Cs, Ps and Ds, respectively. In the equations above, $\Theta(k)$ is the Heaviside function (that is, $\Theta(k) = 1$ whenever $k \geq 0$, being zero otherwise), $0 < n_{pg} \leq N$ is a positive integer not greater than N, and r (the perception of risk) is a real parameter varying between 0 and 1; the parameters c, π_t and b are all real positive; Δ corresponds to the punishment function, which depends on whether the institution is global or local. For local institutions, punishment acts at the group level, and Δ yields $\Delta_{local} = \pi_f \Theta(j_P - n_P)$, which means that a punishment fine π_f is applied to each D in the group whenever $N \geq j_P \geq n_P \geq 0$. For global institutions, punishment acts at the population level; in a population with i_C Cs, i_P Ps and $Z - i_P - i_C$ Ds, the punishment function for global institutions can be written as $\Delta_{global} = \pi_f \Theta(i_P - n_P)$, applying a punishment fine π_f now to every D in the population, whenever $Z \geq i_P \geq n_P \geq 0$. Finally, the fitness f_X of an individual adopting a given strategy, X, will be associated with the average payoff of that strategy in the population. This can be computed for a given strategy in a configuration $i = \{i_C, i_P, i_D\}$ using a multivariate hypergeometric sampling (without replacement; Supplementary Information for details). The number of individuals adopting a given strategy will evolve in time according to a stochastic birth–death process combined with the pairwise comparison rule[21], which describes the social dynamics of Cs, Ps and Ds in a finite population. Under pairwise comparison, each individual of strategy X adopts the strategy Y of a randomly selected member of the population, with probability given by the Fermi function $(1 + e^{\beta(f_X - f_Y)})^{-1}$, where β controls the intensity of selection ($\beta = 5.0$ in all figures). In addition, we consider that, with a mutation probability μ, individuals adopt a randomly chosen strategy. As the evolution of the system depends only on its actual configuration, evolutionary dynamics can be described as a Markov process over a two-dimensional space. Its probability distribution function, $p_i(t)$, which provides information on the prevalence of each configuration at time t, obeys a master equation, a gain–loss equation involving the transition rates between all accessible configurations. The stationary distribution $\bar{p}_i$ is then obtained by reducing the master equation to an eigenvector search problem (Supplementary Information for details). Another central quantity, which portrays the overall evolutionary dynamics in the space of all possible configurations, is the gradient of selection Δ_i. For each configuration i, we compute the most likely path the population will follow, resorting to the probability to increase (decrease) by one the number of individuals adopting a strategy S_k, $T_i^{S_k+}$ ($T_i^{S_k-}$) in each time step. In addition, for each possible configuration i, we make use of multivariate hypergeometric sampling to compute both the (average) fraction of groups that reach n_{pg} contributors, that is, that successfully achieve the public good—which we designate by $a_G(i)$—and the (average) fraction of groups that reach n_p punishers (for local institutions) or whether for that configuration i a global institution will be formed—in both cases, we designate this quantity by $a_I(i)$. Average group achievement—η_G—and institution prevalence—η_I—are then computed averaging over all possible configurations i, each weighted with the corresponding stationary distribution: $\eta_G = \sum_i \bar{p}_i a_G(i)$ and $\eta_I = \sum_i \bar{p}_i a_I(i)$.

Received 9 January 2013; accepted 16 May 2013; published online 14 July 2013

References

1. Barrett, S. *Environment and Statecraft: The Strategy of Environmental Treaty-Making* (Oxford Univ. Press, 2005).
2. Milinski, M., Semmann, D., Krambeck, H. J. & Marotzke, J. Stabilizing the Earth's climate is not a losing game: Supporting evidence from public goods experiments. *Proc. Natl Acad. Sci. USA* **103,** 39 94–3998 (2006).
3. Barrett, S. *Why Cooperate? The Incentive to Supply Global Public Goods* (Oxford Univ. Press, 2007).
4. Milinski, M., Sommerfeld, R. D., Krambeck, H. J., Reed, F. A. & Marotzke, J. The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc. Natl Acad. Sci. USA* **105,** 2291–2294 (2008).
5. Tavoni, A., Dannenberg, A., Kallis, G. & Löschel, A. Inequality, communication and the avoidance of disastrous climate change in a public goods game. *Proc. Natl Acad. Sci. USA* **108,** 11825–11829 (2011).
6. Santos, F. C. & Pacheco, J. M. Risk of collective failure provides an escape from the tragedy of the commons. *Proc. Natl Acad. Sci. USA* **108,** 10421–10425 (2011).
7. Heal, G. & Kristrom, B. Uncertainty and climate change. *Environ. Resour. Econom.* **22,** 3–39 (2002).
8. Levin, S. A. Learning to live in a global commons: Socioeconomic challenges for a sustainable environment. *Ecol. Res.* **21,** 328–333 (2006).
9. Levin, S. A. The trouble of discounting tomorrow. *Solutions* **4,** 20–24 (2012).
10. Ostrom, E. Policy Research Working Paper, *Background Paper to the 2010 World Development Report, WPS5095*, The World Bank (2009).
11. Walker, B. *et al.* Looming global-scale failures and missing institutions. *Science* **325,** 1345–1346 (2009).
12. Poteete, A. R., Janssen, M. A. & Ostrom, E. *Working Together: Collective Action, the Commons, and Multiple Methods in Practice* (Princeton Univ. Press, 2010).
13. Young, O. R., King, L. A. & Schroeder, H. (eds) in *Institutions and Environmental Change* (MIT Press, 2008).
14. Pacheco, J. M., Santos, F. C., Souza, M. O. & Skyrms, B. Evolutionary dynamics of collective action in n-person stag hunt dilemmas. *Proc. R. Soc. B* **276,** 315–321 (2009).
15. May, R. M., Levin, S. A. & Sugihara, G. Ecology for bankers. *Nature* **451,** 893–895 (2008).
16. Rendell, L. *et al.* Why copy others? Insights from the social learning strategies tournament. *Science* **328,** 208–213 (2010).

17. Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **466,** 861–863 (2010).
18. Fowler, J. H. & Christakis, N. A. Cooperative behavior cascades in human social networks. *Proc. Natl Acad. Sci. USA* **107,** 5334–5338 (2010).
19. Carman, K. G. Discussion Paper 02–13 (Stanford Institute for Economic Policy Research, 2003).
20. Sigmund, K. *The Calculus of Selfishness* (Princeton Univ. Press, 2010).
21. Traulsen, A., Nowak, M. A. & Pacheco, J. M. Stochastic dynamics of invasion and fixation. *Phys. Rev. E* **74,** 011909 (2006).
22. Milinski, M., Röhl, T. & Marotzke, J. Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Climatic Change* 1–8 (2011).
23. Chakra, M. A. & Traulsen, A. Evolutionary dynamics of strategic behavior in a collective-risk dilemma. *PLoS Comput. Biol.* **8,** e1002652 (2012).
24. Van Segbroeck, S., Pacheco, J. M., Lenaerts, T. & Santos, F. C. Emergence of fairness in repeated group interactions. *Phys. Rev. Lett.* **108,** 158104 (2012).
25. Barrett, S. & Dannenberg, A. Climate negotiations under scientific uncertainty. *Proc. Natl Acad. Sci. USA* **109,** 17372–17376 (2012).
26. Traulsen, A., Hauert, C., De Silva, H., Nowak, M. A. & Sigmund, K. Exploration dynamics in evolutionary games. *Proc. Natl Acad. Sci. USA* **106,** 709–712 (2009).
27. Schroeder, H., Boykoff, M. T. & Spiers, L. Equity and state representations in climate negotiations. *Nature Clim. Change* **2,** 834–836 (2012).
28. Traulsen, A., Röhl, T. & Milinski, M. An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proc. R. Soc. B* **279,** 3716–3721 (2012).
29. Western Climate Initiative (http://www.westernclimateinitiative.org/).
30. Maynard-Smith, J. & Szathmáry, E. *The Major Transitions in Evolution* (Freeman, 1995).

Acknowledgements

The authors thank F. L. Pinheiro for useful discussions. This research was supported by FCT-Portugal through grants SFRH/BD/86465/2012, PTDC/FIS/101248/2008 and PTDC/MAT/122897/2010, by multi-annual funding of CMAF-UL and INESC-ID (under the project PEst-OE/EEI/LA0021/2013) provided by FCT-Portugal, and Fundação Calouste Gulbenkian through the Stimulus to Research programme for young researchers.

Author contributions

V.V.V., F.C.S. and J.M.P. have contributed equally to this work: they all designed and performed the research, analysed the data and wrote the paper.

Additional information

Supplementary information is available in the online version of the paper. Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to J.M.P.

Competing financial interests

The authors declare no competing financial interests.