

Supporting Information

Vasconcelos et al. 10.1073/pnas.1323479111

SI Text

Theoretical and Experimental Studies of Behavior Regarding Climate Change

Several experiments have been performed to understand human behavior in dealing with global warming (1–4). Social dilemmas involving collective action were set up in a repeated game framework, where a given threshold had to be surpassed—otherwise, there was a variable risk (externally defined) of everyone losing all their endowments. The first results have shown that, most of the time, disaster was not avoided (1), risk being an important factor in promoting disaster avoidance. Later on the possibility to make pledges was introduced in the same repeated threshold public goods game, showing that pledges led to an increase of cooperation (despite the possibility of acting differently from what was pledged) (2). However, when the same treatments were run using players with different wealth, this improvement was demoted (2). Using the same game settings, two time horizons were introduced into the dilemma (3), showing that fixing intermediate goals in climate agreements is beneficial, although the final target is reached less often than the intermediate target. More recently, a nonrepeated threshold public goods game experiment was performed in which the effects of both impact uncertainty (where individuals could lose a random amount of money if the threshold was not met) and threshold uncertainty (where the threshold was a random value) were investigated (4). The authors set up the experiment based on results from a theoretical analysis of such a game (assuming fully rational individuals engaging in a one-shot threshold public goods game) that predicted the existence of a critical value for the threshold uncertainty above which cooperation would collapse. Experiments fully confirmed this catastrophic prediction.

Despite the limited number of scenarios realizable in the laboratory, data stemming from behavioral experiments have provided crucial insights, not only because they unravel human behavior when confronted with climate change issues, but also because they provide important guidelines toward developing theoretical models (5–9). Indeed, a dynamical approach to the problem of cooperating to tame the planet's climate was developed, inspired by the intriguing results that experiments were revealing (5–7), allowing one not only to generalize the experimental settings to scenarios that are more difficult to realize in the laboratory, but also to predict what the impact of different approaches to the solution of the climate change problem may bring. Namely, the effect of risk perception and the disruptive power of uncertainty have been captured in the models (*SI Text, Threshold Uncertainty*). Theoretical models also extended the experimental insights by predicting the importance of small groups and stringent requirements in improving cooperation, as well as the role and scale of sanctioning institutions in supervising agreements. In keeping with this discussion, the present model brings additional information to this important subject.

In particular, although conventional wisdom would lead one to believe that wealth inequality and homophily would constitute important obstacles regarding overall cooperation in climate change negotiations, the present model predicts that, as long as (i) risk perception is high; (ii) climate negotiations are partitioned into smaller groups agreeing on local, short-term targets; and (iii) individuals are influenced by their more successful peers, whom they imitate—irrespective of their wealth class—and making errors while doing so, the prospects are not that grim. On the contrary we find that, under such conditions, cooperation may outcompete defection, benefiting from wealth inequality. On the other hand, and to the extent that agreements aim at short-term targets in-

volving smaller groups, it may also become easier to narrow down threshold uncertainties that, if large, do haunt overall cooperation (4) (*SI Text, Threshold Uncertainty*).

Evolutionary Dynamics in Finite Populations Under Wealth Inequality, Uncertainty, and Homophily

Let us consider a population of Z individuals. As stated in the main text, each individual adopts one of the two possible strategies $X \in \{C, D\}$ and belongs to one of two possible wealth classes $k \in \{R, P\}$. Let us assume there are Z_R rich (with initial endowment b_R) and Z_P poor individuals (with an initial endowment b_P). These numbers will remain fixed. Individuals are given an initial endowment (with $b_P < b_R$) and play the climate threshold Public Goods Game (PGG) (1, 5), engaging in groups of size N . Following the discussion in the main text, and given that rich C s contribute $c_R = cb_R$ whereas poor C s contribute $c_P = cb_P$, the payoff of an individual playing in a group in which there are j_R rich C s, j_P poor C s, and $N - j_R - j_P$ D s can be written as

$$\Pi_{R/P}^D(j_R, j_P) = b_{R/P} \{ \Theta(c_R j_R + c_P j_P - M \bar{b}) + (1 - r) [1 - \Theta(c_R j_R + c_P j_P - M \bar{b})] \}$$

and $\Pi_{R/P}^C(j_R, j_P) = \Pi_{R/P}^D(j_R, j_P) - c_{R/P}$, for rich/poor C s and D s, respectively. In the equations above, $\Theta(k)$ is the Heaviside function [that is, $\Theta(k) = 1$ whenever $k \geq 0$, being zero otherwise], $0 < M \leq N$ is a positive integer, \bar{b} is the average endowment ($Z \bar{b} = Z_R b_R + Z_P b_P$), and r (the perception of risk) is a real parameter varying between 0 and 1; the parameters $0 < c < 1$, \bar{b} , b_R , and b_P are all real positive. Finally, the fitness f_k^X of an individual adopting a given strategy X in a population of wealth class k will be associated with the average payoff of that strategy in the entire population. This can be computed for a given configuration of strategies and wealth classes specified by $\mathbf{i} = \{i_R, i_P\}$, using a multivariate hypergeometric sampling (without replacement):

$$f_R^C(\mathbf{i}) = \binom{Z-1}{N-1}^{-1} \sum_{j_R=0}^{N-1} \sum_{j_P=0}^{N-1-j_R} \binom{i_R-1}{j_R} \binom{i_P}{j_P} \times \binom{Z-i_R-i_P}{N-1-j_R-j_P} \Pi_R^C(j_R+1, j_P) \quad [\text{S1a}]$$

$$f_R^D(\mathbf{i}) = \binom{Z-1}{N-1}^{-1} \sum_{j_R=0}^{N-1} \sum_{j_P=0}^{N-1-j_R} \binom{i_R}{j_R} \binom{i_P}{j_P} \times \binom{Z-1-i_R-i_P}{N-1-j_R-j_P} \Pi_R^D(j_R, j_P) \quad [\text{S1b}]$$

$$f_P^C(\mathbf{i}) = \binom{Z-1}{N-1}^{-1} \sum_{j_R=0}^{N-1} \sum_{j_P=0}^{N-1-j_R} \binom{i_R}{j_R} \binom{i_P-1}{j_P} \times \binom{Z-i_R-i_P}{N-1-j_R-j_P} \Pi_P^C(j_R, j_P+1) \quad [\text{S1c}]$$

$$f_P^D(\mathbf{i}) = \binom{Z-1}{N-1}^{-1} \sum_{j_R=0}^{N-1} \sum_{j_P=0}^{N-1-j_R} \binom{i_R}{j_R} \binom{i_P}{j_P} \times \binom{Z-1-i_R-i_P}{N-1-j_R-j_P} \Pi_P^D(j_R, j_P). \quad [\text{S1d}]$$

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 (10, 11), which describes the social dynamics of rich Cs, poor Cs, rich Ds, and poor 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 $(1 + e^{\beta(f_k^X - f_l^Y)})^{-1}$, for any wealth class $\{k, l\} \in \{R, P\}$, where β controls the intensity of selection. Additionally we consider that, with a mutation probability μ , individuals adopt a randomly chosen different strategy, in such a way that when $\mu = 1$, the individual does change 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 of the form

$$p_i(t + \tau) - p_i(t) = \sum_{i'} \{T_{ii'} p_{i'}(t) - T_{i'i} p_i(t)\}, \quad [\text{S2}]$$

a gain–loss equation that allows one to compute the evolution of $p_i(t)$ given the transition probabilities per unit time τ from the configuration \mathbf{i} to \mathbf{i}' , $T_{i'i}$ (11–13). The stationary distribution \bar{p}_i analyzed in the main text is obtained by making the left-hand side equal to zero, which transforms Eq. S2 into an eigenvector search problem (12), namely, the eigenvector associated with the eigenvalue 1 of the transition matrix W (12) whose matrix elements W_{qp} are built from the transition probabilities per unit time $T_{i'i}$ in the following way: Let us enumerate each of all possible configurations $\mathbf{i} = \{i_R, i_P\}$ of the population by an integer number—we do so by defining a bijective function V such that $p = V(\mathbf{i})$ and $q = V(\mathbf{i}')$ and, therefore, $\mathbf{i} = V^{-1}(p)$ and $\mathbf{i}' = V^{-1}(q)$. Then, we may write $W_{qp} = T_{i'i}$. The transition probabilities $T_{i'i}$ can all be written in terms of the following expression, which gives the probability that an individual with strategy $X \in \{C, D\}$ in the subpopulation $k \in \{R, P\}$ changes to a different strategy $Y \in \{C, D\}$, both from the same subpopulation k and from the other population l (that is, $l = P$ if $k = R$, and $l = R$ if $k = P$):

$$T_k^{X \rightarrow Y} = \frac{i_k^X}{Z} \left((1 - \mu) \left[\frac{i_l^Y}{Z_k - 1 + (1 - h)Z_l} (1 + e^{\beta(f_k^X - f_l^Y)})^{-1} + \frac{(1 - h)i_l^Y}{Z_k - 1 + (1 - h)Z_l} (1 + e^{\beta(f_k^X - f_l^Y)})^{-1} \right] + \mu \right).$$

Thus, if the homophily is maximum ($h = 1$), the imitation occurs only between individuals of the same wealth class (rich or poor), whereas $h = 0$ means that everyone influences and may be influenced by anyone else.

Another central quantity—which portrays the overall evolutionary dynamics in the space of all possible configurations—is the gradient of selection ∇_i (GoS). For each configuration $\mathbf{i} = \{i_R, i_P\}$, we compute the most likely path each subpopulation $k \in \{R, P\}$ will follow, resorting to the probability to increase (decrease) by one, in each time step, the number of cooperators for that configuration \mathbf{i} of the population, which we denote by $T_{i,k}^+$ ($T_{i,k}^-$), such that

$$\nabla_i = \{T_{i,R}^+ - T_{i,R}^-, T_{i,P}^+ - T_{i,P}^-\}.$$

Finally, for each possible configuration \mathbf{i} , we make use of multivariate hypergeometric sampling (Eq. S1) to compute the (average) fraction of groups that reach a total of $M\bar{c}\bar{b}$ in contributions, that is,

that successfully achieve the public good—which we designate by $a_G(\mathbf{i})$. Average group achievement— η_G —is then computed by averaging over all possible configurations \mathbf{i} , each weighted with the corresponding stationary distribution $\eta_G = \sum_i \bar{p}_i a_G(\mathbf{i})$.

Timescale Separation: Games Among the Rich and Among the Poor

To assess what games the rich play in the presence of the poor (Cs and Ds) and the poor play in the presence of the rich (Cs and Ds), we let each subpopulation evolve assuming that the rate of evolution of the other subpopulation is zero. The results are shown in Fig. S1, where we compute the gradient of selection (∇) that governs the evolutionary dynamics of the rich in the presence of frozen, mixed configurations of the poor (Fig. S1 A and B) and vice versa (Fig. S1 C and D). We consider the cases in which the population is subdivided into subpopulations of equal size ($Z_P = Z_R$, Fig. S1 A and C) or not ($Z_P = 4Z_R$, as in the main text, Fig. S1 B and D).

The results in Fig. S1 show that the rich tend to be more cooperative as the difference between the endowments of both classes increases, whereas cooperation among the poor largely remains unaffected. Moreover, whereas the poor engage in a coexistence game in which overall cooperation decreases as cooperation among rich decreases, the dynamics of the rich are influenced by the relative size of the poor subpopulation. In general, the rich engage in an N -player stag-hunt game (14) with different degrees of coordination and coexistence, depending on the (fixed) fraction of poor cooperators. As a result, different combinations of parameters may transform the original N -player stag-hunt dilemma (14)—characterized by two internal roots—into a pure coordination or coexistence dilemma or even a defection dominance dilemma (Fig. S1 A and C). However, as discussed before (5, 6) and illustrated in Fig. 2 of the main text, the unstable fixed point can be overcome by stochastic effects—such as errors in imitation and random exploration of the strategy space—such that the population spends most of its time in the vicinity of the coexistence points. Thus, the prevalent levels of cooperation among the rich will be ultimately defined by the size of the cooperative basin attraction and the position of

the respective coexistence root, both influenced by the dynamics occurring among the poor. It is also noteworthy that the gradients of the rich are 10 times smaller than those of the poor. This means the rate of response of the rich to changes is (on average) slower than that of the poor. In practice, the poor will adjust their behavior more rapidly to changes of the configuration of the rich, thus quickly shifting between the corresponding levels of coexistence between poor Cs and Ds.

Evolutionary Dynamics for the Same Amounts of Rich and Poor

In all experimental settings carried out to date, the fraction of rich and poor in each group was kept equal. Here we compute the analog situation in our model; that is, we compute the stationary distribution in the case when $Z_P = Z_R$. The results are shown in Fig. S2.

Comparison with Fig. 2 in the main text shows unequivocally that, for low risk ($r = 0.2$), rich and poor populations of the same

size lead to more pessimistic prospects concerning overall cooperation (compare the values of η_G below each panel). For a corresponding increase of risk perception (Fig. 2), we observe that, overall, cooperation remains below that observed for asymmetric subpopulation sizes. In particular, the rich cooperate less and are no longer able to compensate for the collapse of cooperation among the poor, a feature that becomes more pronounced with increasing homophily.

Threshold Uncertainty

In the absence of wealth inequality, all individuals are equivalent. This has been, to date, the most studied situation in the laboratory. In particular, in ref. 4 it has been demonstrated how, in a situation that all individuals in the group are equivalent, threshold uncertainty has a disruptive effect on the overall chances of cooperation. Below we show that this is also the case in our model.

Therefore, we modify the individual payoffs so that the games played have a random threshold, with a value drawn from a uniform probability distribution in the interval $[Mc\bar{b} - \delta, Mc\bar{b} + \delta]$. The larger the value of δ is, the larger the uncertainty associated with the threshold. Fig. S3A shows how this uncertainty induces a regime shift in the overall behavior of the population, from an N -player coordination game (5, 14, 15) where cooperators do have a chance toward a defection dominance dilemma. This shift leads, in turn, to a radical change in the profile of the stationary distribution, also shown in Fig. S3B. The impact of this threshold uncertainty on group achievement, $\eta_G(r)$, is shown in Fig. S4, corroborating the results obtained in ref. 4. The study of the effects of threshold uncertainty in the presence of wealth inequality, which is more complex given the problem that the threshold does not affect in the same way the rich and the poor, will be deferred to a future study.

Evolutionary Dynamics in the Presence of Obstinate Cooperators and Defectors

Here we investigate in more detail the role played by obstinate individual behavior in the population. We assume that a fixed fraction of individuals in the population exhibits obstinate behavior;

that is, these individuals are not susceptible to changing their behavior in time. We compute, for all possible combinations, the overall group achievement η_G (*Methods* and *SI Text, Evolutionary Dynamics in Finite Populations Under Wealth Inequality, Uncertainty, and Homophily*) shown in Fig. S5 for three different values of risk (0.2, 0.3, and 0.4, each associated with a different line color) and for the fractions of obstinate individuals indicated in Fig. S5 A–G, *Insets*.

We carry out this analysis as a function of the homophily parameter h . The results corroborate the idea that obstinate Cs generally lead to positive effects in what concerns overall group achievement, whereas obstinate Ds lead to negative effects. Among these, obstinate poor Cs play a crucial role in sustaining cooperation, mostly when homophily is high ($h \sim 1$), whereas obstinate Ds are generally detrimental to overall cooperation.

Robustness of Results as a Function of N

In the following we investigate the dependence of our model results when we change group size and group threshold. To this end we compute, as a function of risk, the same curves that we plot in Fig. 1 of main text, for two group sizes ($N = 6$ and $N = 12$) and for several combinations of M and N , leading to six different scenarios. We use the same parameters as those used in Fig. 1; namely, we split the population of $Z = 200$ individuals into $Z_R = 40$ rich (20%) and $Z_P = 160$ poor (80%); initial endowments are $b_R = 2.5$ and $b_P = 0.625$, ensuring that the average endowment \bar{b} remains $\bar{b} = 1$ (used to generate the gray line in Fig. 1 and Fig. S6); the cost of cooperation also remains, on average, $0.1\bar{b}$, which means $c_R = 0.1b_R$ and $c_P = 0.1b_P$. The results are shown in Fig. S6. Clearly, group size constitutes a very important parameter, because smaller groups lead to higher chances of success (5). Nonetheless, what we observe, in all cases, is that the message contained in Fig. 1 remains valid for all combinations of parameters shown: Wealth inequality without homophily (blue line) systematically fosters overall cooperation for lower values of risk than what is observed under wealth equality (gray line). Finally, homophilic and wealth-unequal subpopulations lead to the grimmest prospects for overall cooperation.

- Milinski M, Sommerfeld RD, Krambeck HJ, Reed FA, Marotzke J (2008) The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc Natl Acad Sci USA* 105(7):2291–2294.
- Tavoni A, Dannenberg A, Kallis G, Löschel A (2011) Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc Natl Acad Sci USA* 108(29):11825–11829.
- Milinski M, Röhl T, Marotzke J (2011) Cooperative interaction of rich and poor can be catalyzed by intermediate climate targets. *Clim Change* 109:807–814.
- Barrett S, Dannenberg A (2012) Climate negotiations under scientific uncertainty. *Proc Natl Acad Sci USA* 109(43):17372–17376.
- Santos FC, Pacheco JM (2011) Risk of collective failure provides an escape from the tragedy of the commons. *Proc Natl Acad Sci USA* 108(26):10421–10425.
- Santos FC, Vasconcelos VV, Santos MD, Neves P, Pacheco JM (2012) Evolutionary dynamics of climate change under collective-risk dilemmas. *Math Models Methods Appl Sci* 22(Suppl 01):1140004.
- Vasconcelos VV, Santos FC, Pacheco JM (2013) A bottom-up institutional approach to cooperative governance of risky commons. *Nature Clim Change* 3(9):797–801.
- Abou Chakra M, Traulsen A (2012) Evolutionary dynamics of strategic behavior in a collective-risk dilemma. *PLoS Comput Biol* 8(8):e1002652.
- Hilbe C, Abou Chakra M, Altrock PM, Traulsen A (2013) The evolution of strategic timing in collective-risk dilemmas. *PLoS ONE* 8(6):e66490.
- Szabó G, Toke C (1998) Evolutionary prisoner's dilemma game on a square lattice. *Phys Rev E Stat Phys Plasmas Fluids Relat Interdiscip Topics* 58(1):69–73.
- Traulsen A, Nowak MA, Pacheco JM (2006) Stochastic dynamics of invasion and fixation. *Phys Rev E Stat Nonlin Soft Matter Phys* 74(1 Pt 1):011909.
- Kampen NV (2007) *Stochastic Processes in Physics and Chemistry* (North-Holland, Amsterdam).
- Imhof LA, Fudenberg D, Nowak MA (2005) Evolutionary cycles of cooperation and defection. *Proc Natl Acad Sci USA* 102(31):10797–10800.
- Pacheco JM, Santos FC, Souza MO, Skyrms B (2009) Evolutionary dynamics of collective action in N-person stag hunt dilemmas. *Proc Biol Sci* 276(1655):315–321.
- Souza MO, Pacheco JM, Santos FC (2009) Evolution of cooperation under N-person snowdrift games. *J Theor Biol* 260(4):581–588.

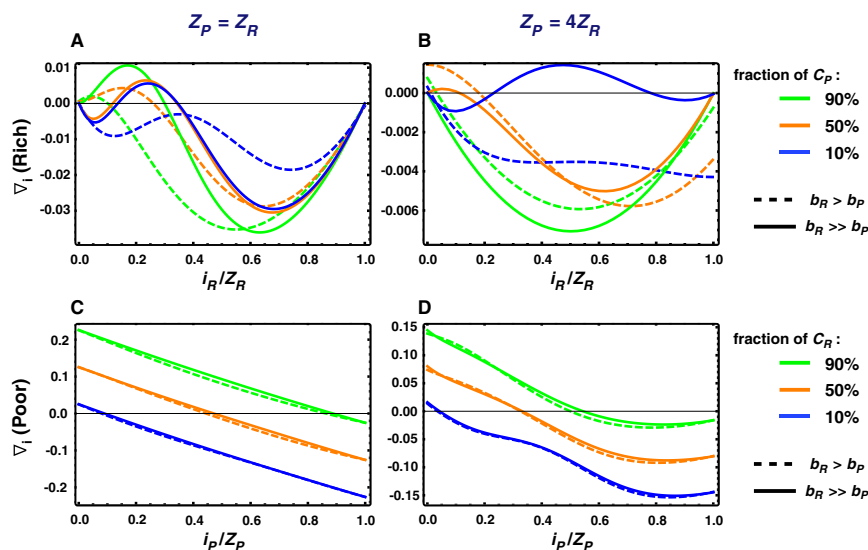


Fig. S1. Rich and poor play different games. (A–D) The gradient of selection (V , *Methods*) in the case when one subpopulation (rich or poor) evolves while the rate of evolution of the other subpopulation is zero, in this way assessing which effective game each subpopulation engages in. (A and B) Number of poor (Z_P) and the fraction of Cs (i_P/Z_P) in the poor population are kept constant (at three different values, legend at *Right*) while the fraction of Cs (i_R/Z_R) and D_S in the rich population is allowed to evolve. (C and D) Here we reverse the timescales and evolve the poor subpopulation, freezing the composition of the rich subpopulation. For each case we consider two regimes associated with different ratios for the initial endowments of rich and poor: $b_R > b_P$ ($b_R = 1.35$) and $b_R < b_P$ ($b_R = 1.75$); imposing that $(b_R Z_R + b_P Z_P)/Z = \bar{b} = 1$ fixes the values of b_P for each case. (A and C) $Z_P = Z_R$. (B and D) $Z_P = 4Z_R$. Other model parameters: $Z = 200$, $N = 10$, $M = 3$, $\beta = 10$, $h = 0$, $r = 0.3$, $c_R = 0.1b_R$, and $c_P = 0.1b_P$.

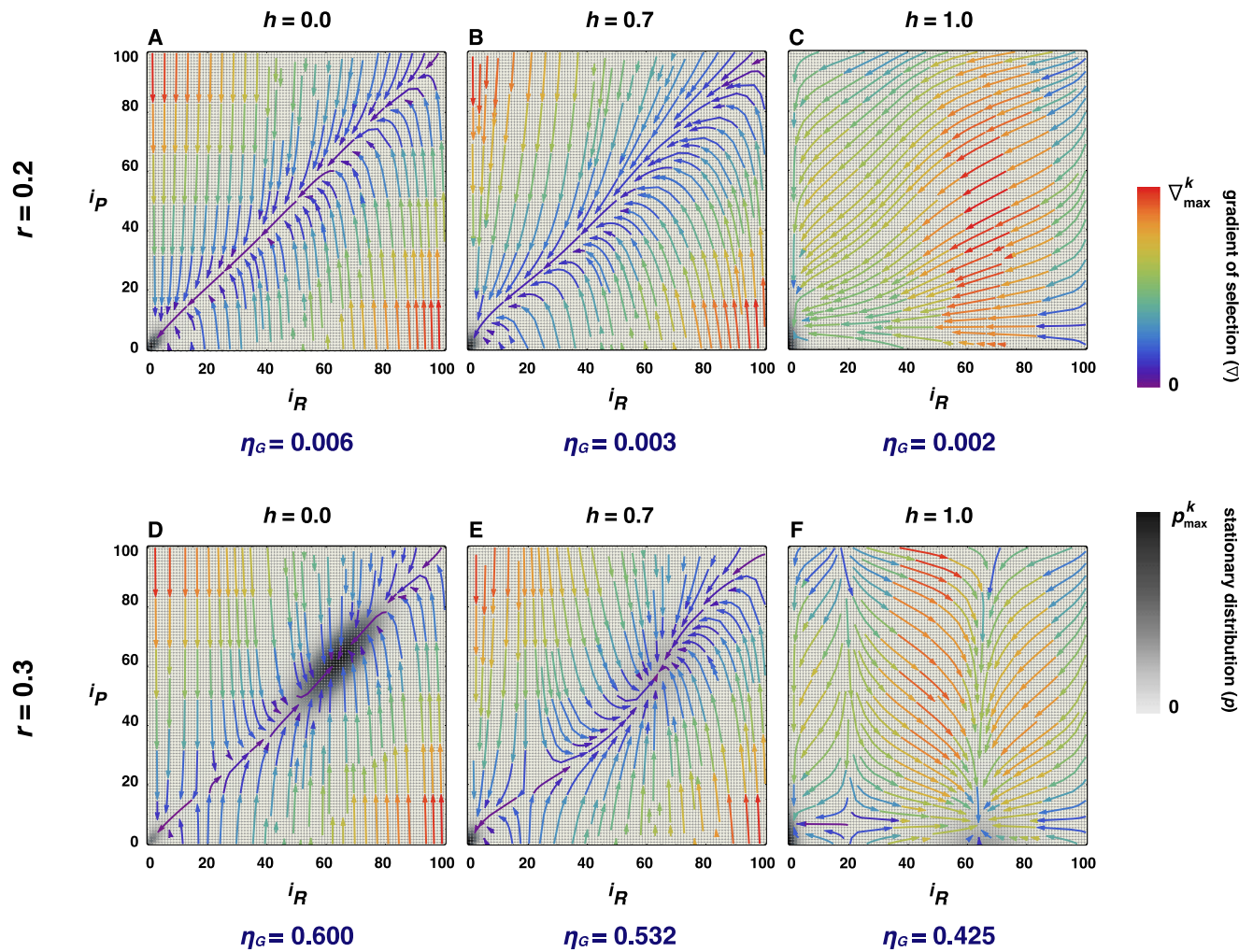


Fig. S2. Evolutionary dynamics for equal fractions of rich and poor. (A–F) The gradient of selection (∇ , *Methods*) in the case when each subpopulation (rich or poor) evolves with the same number of individuals (compare with Fig. 2). Other model parameters: $Z = 200$, $Z_P = Z_R$, $N = 6$, $M = 3c\bar{b}$, $\beta = 5$, $\mu = 1/Z$, $c_R = 0.1b_R$, $c_P = 0.1b_P$, $b_R = 1.7$, and $b_P = 0.3$, ensuring that $(b_R Z_R + b_P Z_P)/Z = \bar{b} = 1$; $\rho_{\max}^{k=A..F} = \{4.2, 5.3, 3.4, 0.2, 0.7, 1.6\} \times 10^{-2}$; and $\nabla_{\max}^{k=A..F} = \{0.25, 0.12, 0.02, 0.25, 0.12, 0.02\}$.

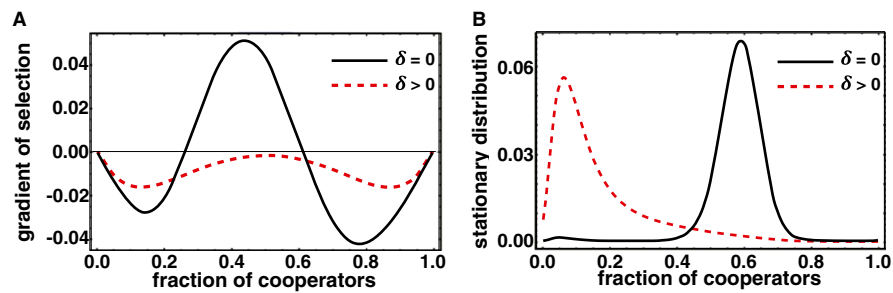


Fig. S3. Threshold uncertainty effect. (A) The gradient of selection; (B) the stationary distribution, that is, the fraction of time the population spends in each population composition specified by x . The black lines provide results for no threshold uncertainty ($\delta = 0$) whereas the red lines show results for $\delta = 2.75$. Other parameters are $Z = 200$, $N = 8$, $M = 4$, $c = 0.1$, $\bar{b} = 1.0$, $\beta = 6.0$, $\mu = 1/Z$, and $r = 0.6$.

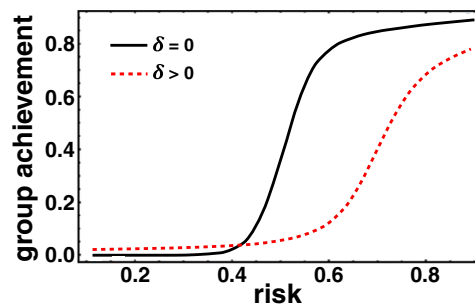


Fig. S4. Group achievement under threshold uncertainty. Results for the group achievement η_G as a function of the risk r are shown. Same parameters as in Fig. S3 are used. Clearly, there is no chance of cooperation before the risk promotes cooperation in an otherwise defection-dominant dilemma imposed by threshold uncertainty.

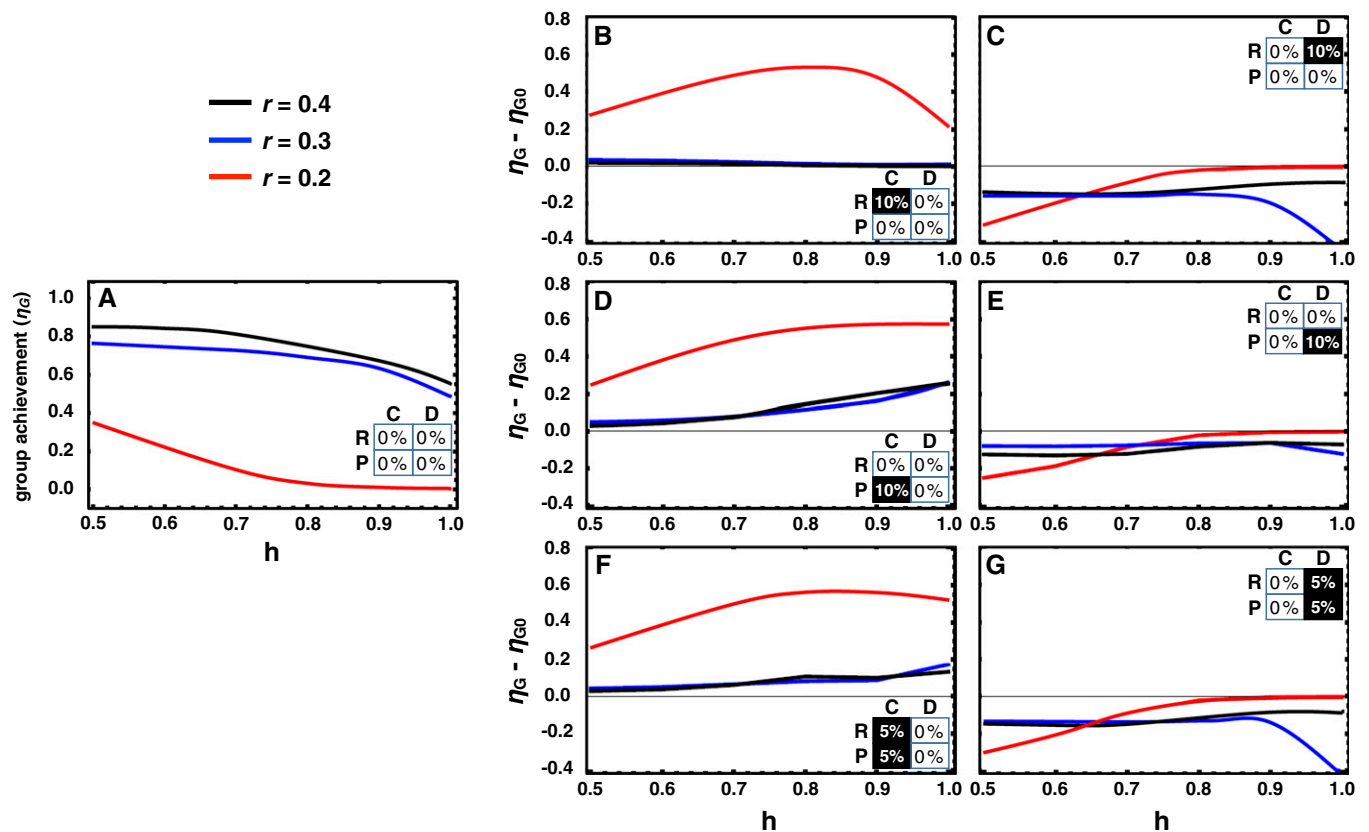


Fig. S5. (A–G) Average group achievement in the presence of obstinate players. *B–G* show the difference between the average group achievement (η_{IG} , *Methods* and *SI Text*, *Evolutionary Dynamics in Finite Populations Under Wealth Inequality, Uncertainty, and Homophily*) in the presence of players whose (obstinate) behavior remains unchanged throughout the evolution of the population and the average group achievement (η_{IG0}) computed in the absence of these types of individuals (*A*). Results are shown for the different values of risk indicated. Sizeable positive effects (in particular for low risk) are obtained whenever obstinate behavior occurs among cooperators. Other parameters: $Z = 200$, $Z_P = 4Z_{R_0}$, $N = 6$, $M = 3$, $\beta = 10$, $\mu = 1/Z$, $c_R = 0.1b_{R_0}$, $c_P = 0.1b_{P_0}$, $b_R = 2.5$, and $b_P = 0.625$.

