How Social Inequality Can Promote Cooperation

Thomas Chadefaux[1*], Dirk Helbing[2]
1 Department of Political Science, University of Rochester, Rochester, NY 14627, USA
2 ETH Zurich, CLU E1, Chair of Sociology, in particular of Modeling and Simulation, Clausiusstrasse 50, 8092 Zurich, Switzerland
∗ E-mail: Corresponding thomas.chadefaux@rochester.edu

Abstract

Does inequality hinder or promote cooperation? To answer this question, we study a Prisoner's Dilemma with local adaptation, to which we add heterogeneity in payoffs. In our model, agents vary in their wealth, and this inequality affects their potential gains and losses. We find that, in such a world in which the rich can get richer, cooperation defeats exploitation under a wide range of conditions. This is in stark contrast to the traditional evolutionary prisoner's dilemma, in which cooperation rarely survives, and almost never thrives. Here on the contrary, cooperators do better than defectors, and this even without any strategic behavior or exogenously imposed strategies. We show how different types of inequality increase cooperation, but also how this effect is marginally decreasing. These results have important consequences for our understanding of the type of social and economic arrangements that are optimal and efficient.

Introduction

Explaining the emergence and stability of cooperation has been a central challenge in biology, economics and sociology [1]. Unfortunately, the mechanisms known to promote cooperation either require elaborate strategies [2], or hold only under restrictive conditions [3]. Even when these conditions are met, cooperation typically merely survives but does not thrive.

Here, we show instead that cooperators can dominate exploiters even without complex strategies and for a wide range of parameters. We obtain this result by studying heterogeneity among individuals, but not in the number of their interaction partners or the relative sizes of the payoff parameters. Rather, people rather vary in their wealth, and this inequality affects their potential gains and losses, so that the rich can get richer over time by accumulating their gains [4,5]. This emergent heterogeneity in which present gains affect future ones is more realistic than the conventional game-theoretical assumption of equal payoffs since, in the real world, the rich can typically engage in deals that generate more benefits. While this is not the first study to emphasize the importance of diversity on cooperation, we adopt a different approach in which diversity is neither predetermined nor exogenous. In particular, previous publications have assumed fixed and exogenously imposed heterogeneity in the payoffs [6] or strategies [7]. Here, on the contrary, we allow a more dynamic and endogenous "rich-get-richer effect," and we do not impose any strategy on the players.

Our study follows the standard literature in analyzing the problem of cooperation by means of 'games'—simplified mathematical representations of social or strategic dilemmas. In these games, people interact in a pairwise fashion with members of their local network, and can take one of two actions: 'cooperate' or 'defect' (i.e., exploit the other). If both cooperate, they each receive R; if both defect, they receive P; finally, if only one defects and the other cooperates, the exploiter receives T, whereas the 'sucker' receives the worst possible payoff, S. The well-known prisoner's dilemma, for example, is defined by $T > R > P > S$. These interactions are repeated over time, and individuals imitate the strategy of the best-performing member of their interaction network (without forecasting).

To this typical setup, we added inequality by varying gains and losses across agents in the following manner. In each interaction, individuals invest some fraction $\alpha \leq 1$ of their wealth (to keep the model simple, we assume a common α for the entire population), and the return on this investment is then determined by the outcome of the game. Assume for example that somebody with wealth 100 interacts with someone else of wealth 2. Then, the smaller budget (here: 2) determines the size of the deal and α the proportion of wealth actually devoted to it, so that both are assumed to invest 2α. This is intuitive: middle-income individuals cannot enter into multi-million deals, and individuals usually do not invest their entire wealth into a single risky deal (hence $\alpha \leq 1$). Note that these specific modeling choices are made out of concern for realism—not because they are needed for our results.

In turn, the payoffs are logically determined by the size of the deal. That is, each player receives αx, where $x \in \{T, R, P, S\}$ denotes the possible payoffs. To return to our earlier numerical example, if one player cooperates and the other decides to exploit, the cooperator receives a gain of $2\alpha S$, whereas the defector receives $2\alpha T$. These gains are added to the individual's existing wealth, which in turn affects her future gains. Thus, the nature of the game (be it a prisoner's dilemma, a stag-hunt or a snowdrift game) is preserved, but gains and losses are endogenous, in the sense that they are a function of past performance. This is the main theoretical innovation of this paper: payoffs are not exogenously defined, but rather endogenously determined as a function of the players' past actions. In addition, we also allow for another form of inequality by randomly varying individuals' initial wealth endowment. This initial inequality is unnecessary for our main result, but leads to stronger results (mainly by increasing the probability with which cooperation emerges).

Model and Results

Setup

We analyze a spatial game with only two types of behaviors: cooperation (C) and defection (D). n^2 players are randomly assigned an initial strategy (C or D) and placed on the sites (cells) of a two-dimensional $n \times n$ square lattice with periodic boundaries (a torus). Time increases discretely (i.e., we use the standard parallel update, though the results are robust to continuous updating). In every round, every player interacts with each of its four neighbors (denoted by $j \in J$) in a pairwise fashion (self-interactions are excluded). Thus, each individual plays four games in each round and her score for the round is the sum of her payoffs against each of its four neighbor (but the fundamental result does not change if we use larger 'Moore' neighborhoods instead). Player i's payoff in her pairwise interaction with player j at time t is defined by the matrix

$$A_t^{ij} = \alpha \times \min(\pi_t^i, \pi_t^j) \times \begin{pmatrix} R & S \\ T & P \end{pmatrix}, \tag{1}$$

where π_t^i denotes i's cumulative payoff at the beginning of round t. Note that, if individuals i and j invested different fractions α_i and α_j of their wealth, $\alpha \times \min(\pi_t^i, \pi_t^j)$ would just have to be replaced by $\min(\alpha_i \pi_t^i, \alpha_j \pi_t^j)$, but this is not a relevant issue in our model. To avoid division by zero and the intricacies of negative wealth ('debt'), we assume that players have a minimal cumulative payoff of 1 (again, this is no crucial model ingredient. Alternatively, all payoffs could be shifted by a constant amount towards positive values.) So in any round t, we have

$$\pi_{t+1}^i = \max\left(1, \sum_{s=0}^{t} \sum_{j \in J} A_s^{ij}\right). \tag{2}$$

Note that any other convex combination of π_i and π_j can support cooperation as well (albeit to different degrees). However, it appears reasonable to assume that the magnitude of the gains (or losses) two players can obtain is limited by the wealth of the weaker player. When a rich person meets a poor one, the stakes in the game they might play are small in absolute terms.

At the end of each step, agents update their payoffs and switch their strategy to the one of their most successful neighbor, but do not switch if they were the most successful in that round. By 'successful', we mean here the amount of gains obtained during that round (but similar results hold if we use total, cumulative payoffs over time instead).

Unless otherwise specified, the simulations used to generate the graphs are based on the following setup: 40,000 agents are placed on a 200 × 200 torus, each simulation is run for 1,000 steps, and the results are averaged over 100 different runs for each set of parameters. The default set of parameters is $\alpha = 1$ (except for figure 5, which shows the impact of various levels of α), $R = 1$, $T = 2$, $S = 0$, and $P = 0$ (but $P > S$ does not change our conclusions). These values are standard in the literature on games. Agents start with an initial wealth uniformly distributed between 0 and 10 (again, figure 5 shows the impact of changes in this initial distribution). The interaction network in each simulation is of degree $k = 4$. We start with 50% cooperators and 50% defectors uniformly distributed in space. The updating rule is synchronous, but results are robust to an asynchronous one. The asymptotic proportion is determined by averaging over the last 100 rounds. In none of the cases is the standard deviation in these last steps large enough to suggest any instability that would warrant a different approach. In other words, the level of cooperation converges to a fixed, parameter-dependent value.

Results

Extending the classical spatial prisoner's dilemma [8] by wealth accumulation and social inequality yields striking results. Intuitively, we would expect the 'rich get richer' effect to undermine cooperation and, initially, defectors do indeed obtain a high score, while co-operators perform poorly. This is because defectors exploit cooperators, thereby securing an initial level of wealth that allows them to do well. As a result, most players imitate these successful defectors, and cooperation almost disappears from the world (Fig. 1B). However, those who do survive are those who were initially isolated from defectors by a cooperative network, and have thereby accumulated a substantial cumulative wealth. They have been at the center of a cluster of cooperators, and hence have been able to accumulate a sizeable wealth in their first rounds. This wealth then makes them 'competitive' against defectors (Fig. 1C-D). More precisely, their wealth enables them to secure large gains with their peers, but only small losses when interacting with defectors (since defectors, who perform poorly in the defective environment that follows the first rounds, have little to invest, and hence do not pose a great threat to cooperators). Therefore, after the initial turmoil, cooperators rapidly take over the entire lattice and defectors vanish almost completely.

These results are in marked contrast to the classical spatial prisoner's dilemma, in which defectors tend to spread and cooperators do not thrive. Moreover, our results are robust to a large number of modeling choices such as the size of the world, the updating rule, the level of initial inequality, and are robust to random strategy switches ("mutations") [9].

The differences between the cooperation level under the traditional prisoner's dilemma setup and those that result from the present model are striking. There are three ways to appreciate these differences. First, in an equal world (the classical model) in which payoffs and initial wealth are homogenous across agents, the final proportion of cooperators is low for any $T > 1$. That is, it is very rare and difficult to obtain cooperation and even more difficult to sustain it without the rich-get-richer dynamics and social inequality considered here. In our setup, however, cooperation is stable for a much larger range of parameters

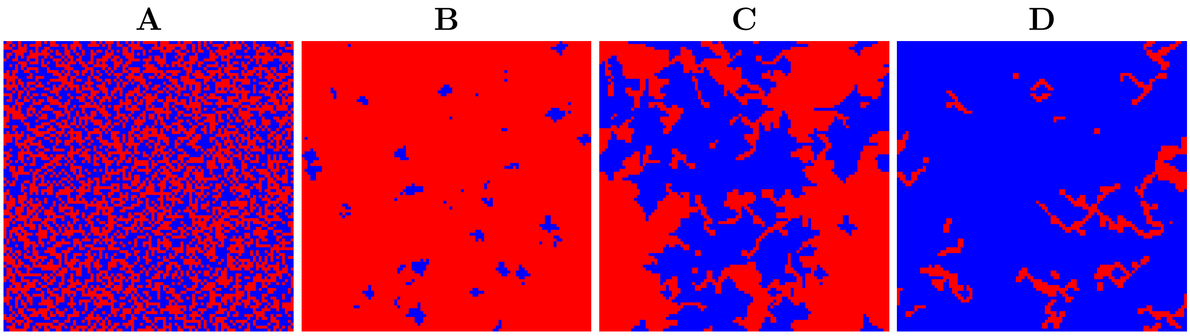

Figure 1. Evolution of cooperation over time. Snapshots of the lattice of a typical run (red denotes defectors, blue cooperators). (A) We start at $t = 0$ with 50% cooperators. (B) After only a few steps ($t = 4$), the number of cooperators has dramatically decreased, and only a few cooperative clusters survive. (C) Those who do survive, however, expand ($t = 50$) and (D) ultimately take over the entire lattice ($t = 400$).

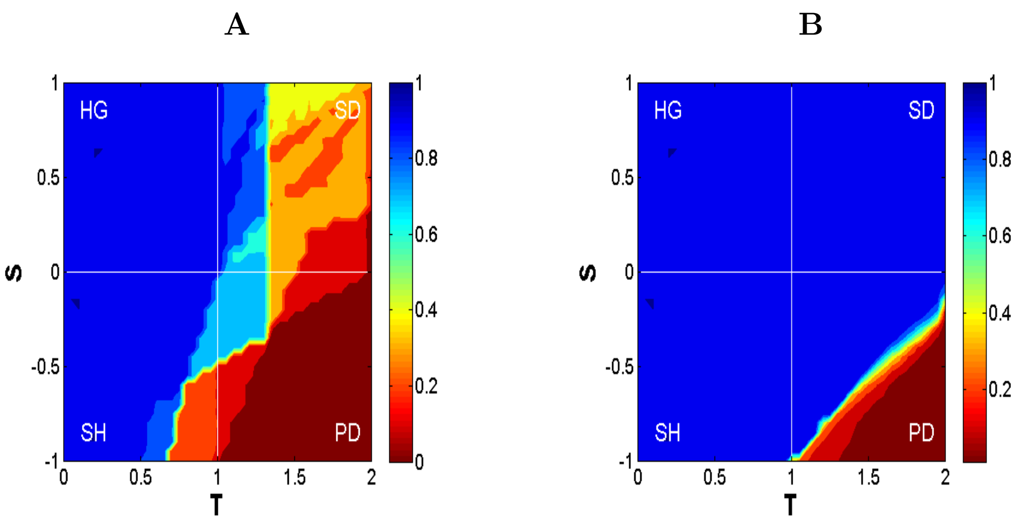

Figure 2. Final proportion of cooperators. Asymptotic proportion of cooperators in equal (A) and unequal (B) worlds as a function of the payoff parameters T and S. (A) In an equal world, in which payoffs and initial wealth are homogenous across agents, the proportion of cooperators is low for any $T > 1$. (B) In an unequal environment, in which the rich get richer and initial wealth is heterogeneous, cooperation is stable for a much larger range of payoff parameters. Initial inequality is not necessary for the result, but promotes greater differences (see also Fig. 5). The top-left quadrant corresponds to the harmony game (HG); the bottom-left ($T \leq 1$ and $S < 0$) to the stag-hunt (or 'assurance') game (SH); the upper-right quadrant ($S \geq 0$, $T > 1$) to the snow-drift (or 'chicken') game (SD); and the lower-right quadrant ($S < 0$ and $T > 1$) corresponds to the prisoner's dilemma (PD) [10].

(Fig. 2). Second, and even more striking, we find that cooperators *dominate* defectors for a far wider range of parameters than under the classical prisoner's dilemma. That is,

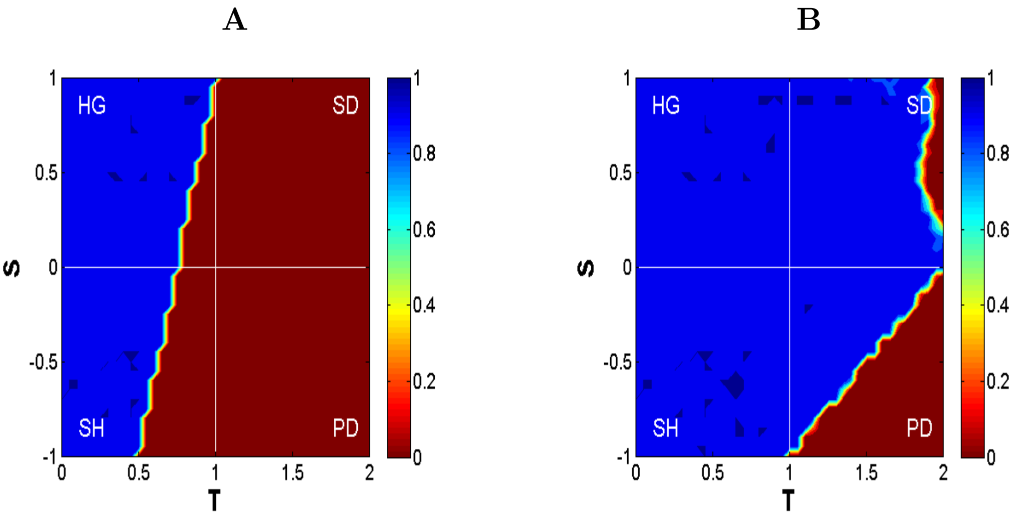

Figure 3. **Domination of cooperation in equal (A) and unequal environments (B).** The contour plot shows the percentage of runs that end with more than 99% cooperators, as a function of the two payoff parameters S and T. In an equal world in which wealth and gains are the same for all individuals and wealth is not accumulated (A), the range of parameters for which cooperation can take over is very limited (blue area in A). When the rich get richer (B), however, the range of parameters is far larger (blue area in B).

not only can cooperation survive more often, but it thrives and dominates far more than in an equal world (Fig. 3). Finally, cooperation *survives* more often in our setup. In fact, cooperation fails to survive only very rarely and only for the most extreme parameters (Fig. 4).

We also investigated the effect of two types of inequality on the cooperation level. First, we varied the initial inequality level (heterogeneity). That is, we changed the range from which initial wealth levels were drawn. We find that a more unequal initial world leads to increased levels of cooperation, holding other parameters constant (Fig. 5). However, beyond a certain level of inequality, a further increase in social inequality does not enhance the level of cooperation significantly anymore.

We also varied what we call the proportion of wealth that individuals invest (α), which determines how much past gains affect present benefits, i.e. the rate of wealth accumulation. $\alpha = 1$, for example, means that present payoffs depend on the full amount of past benefits. $\alpha = 0.01$, on the other hand, means that only one percent of the accumulated wealth affects the present payoffs and, correspondingly, the rich-get-richer effect is small. We find that the larger the "rich-get-richer effect" (Fig 5), the more likely cooperation is to prevail. As for the level of initial inequality, however, there is no noteworthy increase in the level of cooperation beyond a value of $\alpha = 0.7$, i.e. more inequality does not seem to be beneficial.

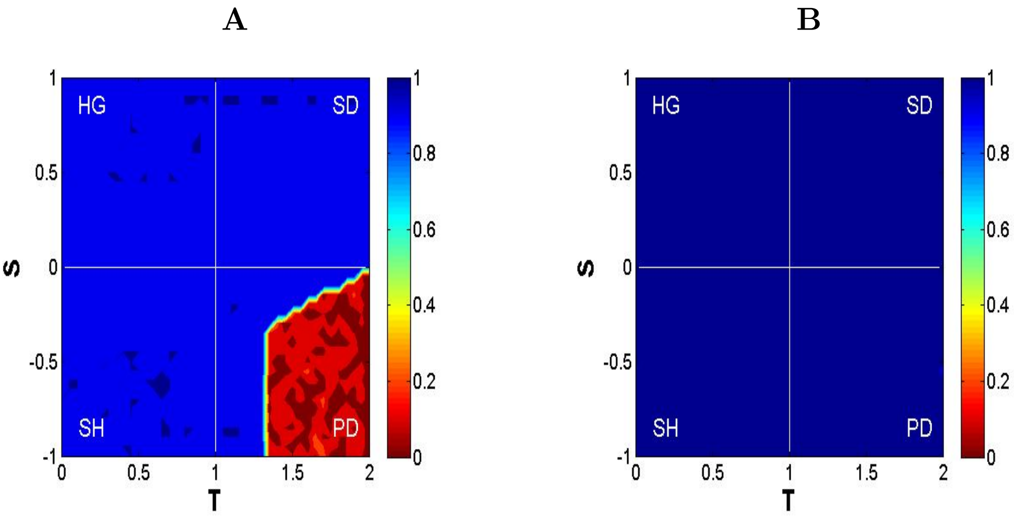

Figure 4. When Does Cooperation Survive? Percentage of runs in which at least 1% of cooperators survive after 1000 steps (A) in an equal world and (B) in an unequal world. Note that cooperation can survive in much more hostile conditions ($T > 1.5$ and $S < 0$) when payoffs are unequal than when they are not. In particular, for extreme values (T close to 2 and S close to -1), cooperation never survives in an equal world, but can survive in the unequal one.

Finally, we investigated the robustness of our finding to changes in the number of individuals in the world, and found a strong positive correlation with the likelihood that cooperation emerges (Fig. 6). That is, for any payoff parameters, adding individuals on the lattice increases the probability that cooperation will survive and dominate. The logic behind this result is that the survival of cooperators during the first steps of the game is critical. Large worlds—those with many individuals—are likely to have at least one cluster of cooperators of sufficient size to survive the initial turmoil. Since one such cluster is sufficient to foster and promote the eventual spread of cooperation, a large world also increases the chances that cooperation spreads eventually.

Conclusion

Overall, our results highlight that, besides social diversity [6, 11], emergent heterogeneity through the rich-get-richer effect supports the welfare of society. Social inequality can dramatically and unexpectedly promote cooperation, which has broad theoretical and practical implications. The recent financial crisis, for example, has raised fundamental questions regarding appropriate incentive structures and income distributions. Our results suggest that inequality is favorable for social cooperation, although it does not have to be extreme to be effective (Fig. 5).

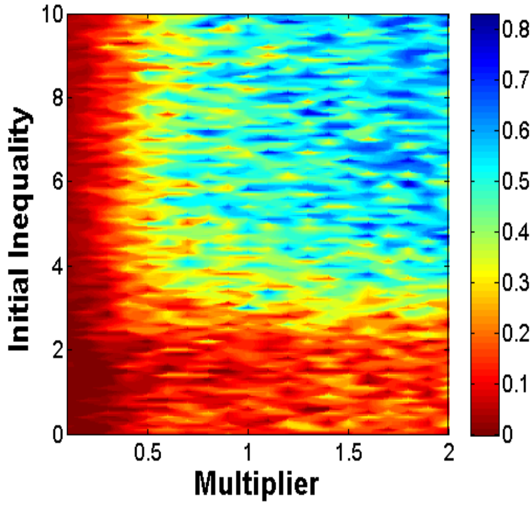

Figure 5. **Effect of inequality on cooperation.** The contour plot shows the asymptotic proportion of cooperators for a given set of parameters, as a function of the initial inequality and the multiplier α, which reflects the extent to which the past affects the present (we assume the same α for each individual). A low multiplier means that the past has little effect, whereas a large one means that the rich get a lot richer. "Initial inequality" defines the variation in the initial wealth. A large number means that the world starts with very unequal conditions, whereas a low one means a more equal initial distribution.

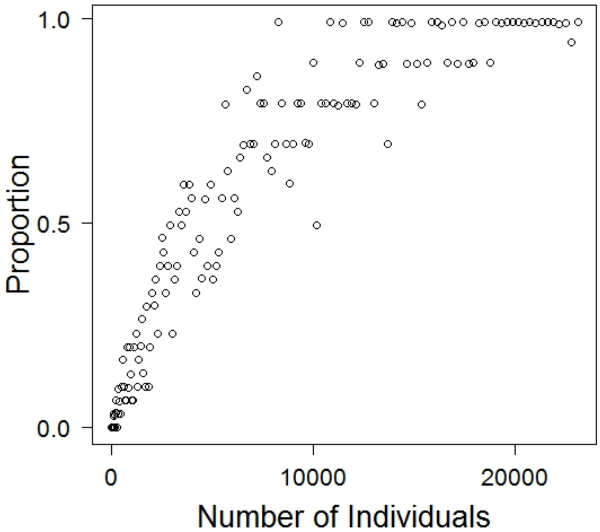

Figure 6. Cooperation as a function of the number of individuals. Asymptotic proportion of cooperators as a function of the number of individuals. A larger number of individuals increases the likelihood that an initial supercritical cluster of cooperators will survive, and hence that at least one cooperator is protected sufficiently long to accumulate enough wealth to out compete defectors. Hence, for any given set of parameters, a large number of individuals increases the probability with which cooperation will prevail.

Acknowledgments

This work was partially supported by the Future and Emerging Technologies programme FP7-COSI-ICT of the European Commission through the project QLectives (grant no.: 231200).

References

1. Nowak M (2006) Five rules for the evolution of cooperation. Science 314: 1560–1563.

2. Helbing D, Yu W (2009) The outbreak of cooperation among success-driven individuals under noisy conditions. PNAS 106: 3680.

3. Hauert C, Doebeli M (2004) Spatial structure often inhibits the evolution of cooperation in the snowdrift game. Nature 428: 643–646.

4. Merton RK (1968) The matthew effect in science: The reward and communication systems of science are considered. Science 159: 56–63.

5. Albert R, Barabasi A (1999) Emergence of scaling in random networks. Science 286: 509–512.

6. Perc M, Szolnoki A (2008) Social diversity and promotion of cooperation in the spatial prisoners dilemma game. Physical Review E 77: 11904.

7. McNamara JM, Barta Z, Houston AI (2004) Variation in behaviour promotes cooperation in the prisoner's dilemma game. Nature 428: 745–748.

8. Nowak M, May R (1992) Evolutionary games and spatial chaos. Nature 359: 826–829.

9. Traulsen A, Semmann D, Sommerfeld RD, Krambeck H, Milinski M (2010) Human strategy updating in evolutionary games. Proceedings of the National Academy of Sciences 107: 2962–2966.

10. Roca CP, Cuesta JA, Sanchez A (2009) Effect of spatial structure on the evolution of cooperation. Physical Review E 80: 046106.

11. Santos F, Santos M, Pacheco J (2008) Social diversity promotes the emergence of cooperation in public goods games. Nature 454: 213–216.