Chapter 100
Evolution of Fairness and Conditional Cooperation in Public Goods Dilemmas

Sven Van Segbroeck, Jorge M. Pacheco, Tom Lenaerts, and Francisco C. Santos

Abstract Cooperation prevails in many collective endeavours. To ensure that cooperators are not exploited by free riders, mechanisms need to be put into place to protect them. Direct reciprocity, one of these mechanisms, relies on the facts that individuals often interact more than once, and that they are capable of retaliating when exploited. Yet in groups, strategies targeting retaliation against specific group members may be unfeasible, because individuals may not be able to identify clearly who contributed and who did not. Still, they may assess what constitutes a fair income from a collective endeavour. We discuss here how conditional cooperation in group interactions emerges naturally and how natural selection leads populations to evolve towards a specific level of fairness (Van Segbroeck et al., Phys. Rev. Lett., 108:158104, 2012), contingent on the nature and size of the collective dilemma faced by individuals.

Keywords Reciprocity · N-Player games · Repeated games · Fairness · Evolutionary game theory

Darwinian evolution dictates that cooperation, which is the act of helping someone at a personal cost, is not evolutionary viable as a behavioral strategy since others may profit directly from this act and are not required to behave in the same way.

S. Van Segbroeck (✉) · T. Lenaerts
MLG, Université Libre de Bruxelles, Brussels, Belgium

J.M. Pacheco
Departamento de Matemática e Aplicações, Universidade do Minho, Braga, Portugal

J.M. Pacheco · F.C. Santos
ATP-group, CMAF, Complexo Interdisciplinar, Lisbon, Portugal

T. Lenaerts
AI-lab, Vrije Universiteit Brussel, Brussels, Belgium

F.C. Santos
DEI & INESC-ID, Instituto Superior Técnico, TU Lisbon, Lisbon, Portugal

T. Gilbert et al. (eds.), *Proceedings of the European Conference on Complex Systems 2012*, Springer Proceedings in Complexity, DOI 10.1007/978-3-319-00395-5_100,

Yet, cooperation is prevalent at all biological levels and has been identified as an essential complexity inducing mechanism in different contexts. In light of this paradox, theoretical biologists have postulated a number of mechanisms that explain the evolution of cooperation in the competitive world defined by Darwin's natural selection [1].

One of the most prominent mechanisms is Robert Trivers' *direct reciprocity* [2], which became famous through the invention, by Anatol Rapaport, of the tit-for-tat strategy within the game theoretical tournaments setup by Robert Axelrod [3]. Direct reciprocity states that two players in a strategic interaction may prefer to cooperate if there is a high chance that they meet again, capturing the essence of "*If you scratch my back, I will scratch yours*". Trivers showed mathematically for pairs of players that if the probability of interacting again (w) is higher than the cost-to-benefit ratio of the game ($w > c/b$), then mutual cooperation would evolve.

Given that a lot of strategic situations involve more than two players, as for instance in Social Welfare or Climate Change related problems, one can wonder how these results extend to those situations. In that context, tit-for-tat can no longer be used as a direct retaliation strategy by each player since it is no longer obvious against whom to retaliate. In addition, retaliating blindly may send the wrong signal to the other participants, triggering them to defect in turn. In [4] we examined whether there exists a strategy like tit-for-tat that leads to cooperation in groups, providing at the same time a mechanism to respond towards good or bad "group behavior"? One solution could be to decide to cooperate if a sufficient number of members cooperated in the previous round, as is the case in the two-player situation. Yet, how many players should cooperate before one decides to do the same? In other words, what is acceptable, or even, fair?

To answer these questions we analyzed an evolutionary model (for both infinite and finite populations) in which N individuals interact in the context of a public goods dilemma, the (repeated) N-persons prisoner's dilemma (NPD) [1, 5]. In this game all players have the chance to contribute an amount c to the public good. The sum of the effective contributions is then multiplied by a factor F and this result is then shared equally among all the N group members, irrespective of their contribution. This entire process repeats itself with a probability w, resulting in an average number of $\langle r \rangle = (1-w)^{-1}$ rounds per group. The outcome of the game may differ from round to round, as individuals can base their decision to contribute on the result of the previous round. We distinguished N different types of reciprocators, encoded in terms of the strategies $R_M (M \in \{1, \ldots, N\})$, where R_M players always contribute in the first round and subsequently contribute to the public good in the next round only if at least M players did the same in the previous round. In addition to these N different reciprocator types, we also included the strategy always defect (AD) to account for unconditional defectors.

In infinitely large populations, using a replicator equation [6] to describe the evolutionary dynamics between a particular reciprocator type R_M and AD, our results identify that (when $F < N$) cooperation may be enforced by the reciprocators when a sufficient fraction, called the coordination equilibrium x_L, of these reciprocators is present. When the number of individuals belonging to the type R_M is lower than

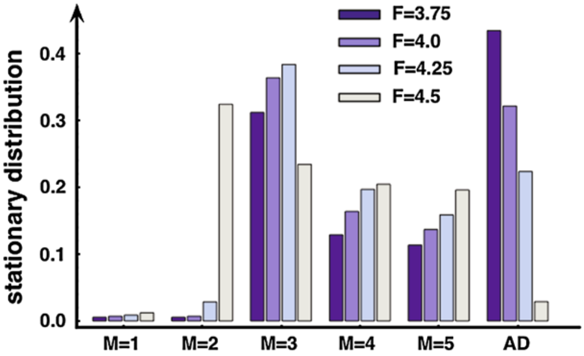

Fig. 100.1 The stationary distributions (prevalence in time of each strategy) for different values of the multiplication factor $F(w = 0.9, N = 5, Z = 100, C = 1)$. This figure was reproduced from [4]

this threshold x_L, R_M players are unsuccessful against the AD players since they receive insufficient benefits from cooperating always in the first round. Yet, even when the number of R_M individuals is higher than x_L, they will not take over the entire population. There is always a fraction of AD players remaining, defining a coexistence equilibrium x_R . Not only do the values of these two equilibria depend on the probability of repeating the game, they are also different for each value of M: the bigger M, i.e. the more group members are required to contribute to the public good, the less often the game needs to be repeated for cooperation to become viable in the population. Still for an insufficient number of repetitions, AD will dominate the population. In general, we show [4] how the presence of conditional cooperators transforms the nature of the dilemma from defection dominance, towards N-person coordination games [7, 8].

This first analysis only takes into account the selection dynamics between two types of players, i.e. a single type of reciprocators R_M and AD. There may be an evolutionary preference for a particular value M. As such, one needs to explore the viability of all the strategies together. Moreover, as populations are finite, certain stochastic effects influence the evolutionary equilibrium obtained in this game. Consequently, we examined analytically a stochastic, finite population analogue of the deterministic evolutionary dynamics defined above, in which strategies evolve according to a mutation-selection process defined in discrete time. We assumed furthermore a limit in which mutations are rare, allowing us to compute the stationary distributions of the six different strategies. We could also show through numerical simulations that our results hold for a wider interval of mutation regimes.

As shown in Fig. 100.1, there is a specific concept of fairness, associated with an aspiration level M^*, whose corresponding strategy is most favored by evolution for a given value of F, being most prevalent among all strategies. Moreover, several values of M, corresponding to more stringent requirements in terms of the number of contributions ($M > M^*$), may coexist in the population. Their abundance is determined by the harshness of the dilemma, which is in the NPD defined by the value of the multiplication factor F: As F decreases, the fractions of the other reciprocators decreases in favor of R_M and AD, until at a certain point when the current M is no longer viable and more stringent conditions (higher M) are required to enforce cooperation.

This evolutionary dynamics becomes clearer when one considers the different evolutionary flows among strategies. By adhering to conditional reciprocation towards groups, individuals find a way towards widespread cooperation. Yet, if by chance, the population ends up adopting less demanding conditions (low M), then defection may prosper again, as it increases the temptation for this behavior to spread. Hence, stochastic effects may lead to cyclic behaviors corresponding to the oscillations between cooperation and defection akin to those observed both in nature and human history.

In summary, we have shown in [4] that even in repeated group interactions cooperation triggered by reciprocal strategies becomes viable when the probability of repeating the game is sufficiently high. Besides, this probability is dependent on what individuals perceive as a fair collective effort M that defines each reciprocal strategy. Furthermore, we show that this process leads to the emergence particular levels of fairness, which depend on the dilemma at stake.

References

1. Sigmund K (2010) The calculus of selfishness. Princeton series in theoretical and computational biology. Princeton University Press, Princeton
2. Trivers RL (1971) The evolution of reciprocal altruism. Q Rev Biol 46:35–57
3. Axelrod R (1984) The evolution of cooperation. Basic Books, New York
4. Van Segbroeck S, Pacheco JM, Lenaerts T, Santos FC (2012) Emergence of fairness in repeated group interactions. Phys Rev Lett 108:158104
5. Hardin G (1968) The tragedy of the commons. Science 162:1243–1248
6. Hofbauer J, Sigmund K (1998) Evolutionary games and population dynamics. Cambridge University Press, Cambridge
7. Pacheco JM, Santos FC, Souza MO, Skyrms B (2009) Evolutionary dynamics of collective action in N-person stag hunt dilemmas. Proc - Royal Soc, Biol Sci 276(1655):315–321
8. 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