Heinrich H. Nax & Matjaž Perc
Directional learning and the provisioning of public goods

Article (Published version)
(Refereed)

OPEN

Directional learning and the provisioning of public goods

Heinrich H. Nax[1] & Matjaž Perc[2,3]

[1]Department of Social Sciences, ETH Zürich, Clausiusstrasse 37-C3, 8092 Zurich, Switzerland, [2]Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, [3]Faculty of Natural Sciences and Mathematics, University of Maribor, Koroška cesta 160, SI-2000 Maribor, Slovenia.

We consider an environment where players are involved in a public goods game and must decide repeatedly whether to make an individual contribution or not. However, players lack strategically relevant information about the game and about the other players in the population. The resulting behavior of players is completely uncoupled from such information, and the individual strategy adjustment dynamics are driven only by reinforcement feedbacks from each player's own past. We show that the resulting "directional learning" is sufficient to explain cooperative deviations away from the Nash equilibrium. We introduce the concept of k–strong equilibria, which nest both the Nash equilibrium and the Aumann-strong equilibrium as two special cases, and we show that, together with the parameters of the learning model, the maximal k–strength of equilibrium determines the stationary distribution. The provisioning of public goods can be secured even under adverse conditions, as long as players are sufficiently responsive to the changes in their own payoffs and adjust their actions accordingly. Substantial levels of public cooperation can thus be explained without arguments involving selflessness or social preferences, solely on the basis of uncoordinated directional (mis)learning.

C ooperation in sizable groups has been identified as one of the pillars of our remarkable evolutionary success. While between-group conflicts and the necessity for alloparental care are often cited as the likely sources of the other-regarding abilities of the genus *Homo*[1,2], it is still debated what made us the "super-cooperators" that we are today[3,4]. Research in the realm of evolutionary game theory[5–10] has identified a number of different mechanisms by means of which cooperation might be promoted[11,12], ranging from different types of reciprocity and group selection to positive interactions[13], risk of collective failure[14], and static network structure[15,16].

The public goods game[17], in particular, is established as an archetypical context that succinctly captures the social dilemma that may result from a conflict between group interest and individual interests[18,19]. In its simplest form, the game requires that players decide whether to contribute to a common pool or not. Regardless of the chosen strategy by the player himself, he receives an equal share of the public good which results from total contributions being multiplied by a fixed rate of return. For typical rates of return it is the case that, while the individual temptation is to free-ride on the contributions of the other players, it is in the interest of the collective for everyone to contribute. Without additional mechanisms such as punishment[20], contribution decisions in such situations[18,19] approach the free-riding Nash equilibrium[21] over time and thus lead to a "tragedy of the commons"[22]. Nevertheless, there is rich experimental evidence that the contributions are sensitive to the rate of return[23] and positive interactions[13], and there is evidence in favor of the fact that social preferences and beliefs about other players' decisions are at the heart of individual decisions in public goods environments[24].

In this paper, however, we shall consider an environment where players have no strategically relevant information about the game and/or about other players, and hence explanations in terms of social preferences and beliefs are not germane. Instead, we shall propose a simple learning model, where players may mutually reinforce learning off the equilibrium path. As we will show, this phenomenon provides an alternative and simple explanation for why contributions rise with the rate of return, as well as why, even under adverse conditions, public cooperation may still prevail. Previous explanations of this experimental regularity[18] are based on individual-level costs of 'error'[25,26].

Suppose each player knows neither who the other players are, nor what they earn, nor how many there are, nor what they do, nor what they did, nor what the rate of return of the underlying public goods game is. Players do not even know whether the underlying rate of return stays constant over time (even though in reality it does) because their own payoffs are changing due to the strategy adjustments of other players, about which they have no

information. Without any such knowledge, players are unable to determine ex ante whether contributing or not contributing is the better strategy in any given period, i.e., players have no strategically relevant information about how to respond best. As a result, the behavior of players has to be *completely uncoupled*[27,28], and their strategy adjustment dynamics are likely to follow a form of *reinforcement*[29,30] feedback or, as we shall call it, *directional learning*[31,32]. We note that, in our model, due to the one-dimensionality of the strategy space, reinforcement and directional learning are both adequate terminologies for our learning model. Since reinforcement applies also to general strategy spaces and is therefore more general we will prefer the terminology of directional learning. Indeed, such directional learning behavior has been observed in recent public goods experiments[33,34]. The important question is how *well* will the population learn to play the public goods game despite the lack of strategically relevant information. Note that *well* here has two meanings due to the conflict between private and collective interests: on the one hand, how close will the population get to playing the Nash equilibrium, and, on the other hand, how close will the population get to playing the socially desirable outcome.

The learning model considered in this paper is based on a particularly simple "directional learning" algorithm which we shall now explain. Suppose each player plays both cooperation (contributing to the common pool) and defection (not contributing) with a mixed strategy and updates the weights for the two strategies based on their relative performances in previous rounds of the game. In particular, a player will increase its weight on contributing if a previous-round switch from not contributing to contributing led to a higher realized payoff or if a previous-round switch from contributing to not contributing led to a lower realized payoff. Similarly, a player will decrease its weight on contributing if a previous-round switch from contributing to not contributing led to a higher realized payoff or if a previous-round switch from not contributing to contributing led to a lower realized payoff. For simplicity, we assume that players make these adjustments at a fixed incremental step size δ, even though this could easily be generalized. In essence, each player adjusts its mixed strategy directionally depending on a Markovian performance assessment of whether a previous-round contribution increase/decrease led to a higher/lower payoff.

Since the mixed strategy weights represent a well-ordered strategy set, the resulting model is related to the directional learning/aspiration adjustment models[31,32,35], and similar models have previously been proposed for bid adjustments in assignment games[36], as well as in two-player games[37]. In ref. 36 the dynamic leads to stable cooperative outcomes that maximize total payoffs, while Nash equilibria are reached in ref. 37. The crucial difference between these previous studies and our present study is that our model involves more than two players in a voluntary contributions setting, and, as a result, that there can be interdependent directional adjustments of groups of players including more than one but not all the players. This can lead to uncoordinated (mis)-learning of subpopulations in the game.

Consider the following example. Suppose all players in a large standard public goods game do not contribute to start with. Then suppose that a number of players in a subpopulation uncoordinatedly but by chance simultaneously decide to contribute. If this group is sufficiently large (the size of which depends on the rate of return), then this will result in higher payoffs for all players including the contributors, despite the fact that not contributing is the dominant strategy in terms of unilateral replies. In our model, if indeed this generates higher payoffs for all players including the freshly-turned contributors, then the freshly-turned contributors would continue to increase their probability to contribute and thus increase the probability to trigger a form of stampede or herding effect, which may thus lead away from the Nash equilibrium and towards a socially more beneficial outcome.

Our model of uncoordinated but mutually reinforcing deviations away from Nash provides an alternative explanation for the following regularity that has been noted in experiments on public goods provision[18]. Namely, aggregate contribution levels are higher the higher the rate of return, despite the fact that the Nash equilibrium remains unchanged (at no-contribution). This regularity has previously been explained only at an individual level, namely that 'errors' are less costly – and therefore more likely – the higher the rate of return, following quantal-response equilibrium arguments[25,26]. By contrast, we provide a group-dynamic argument. Note that the alternative explanation in terms of individual costs is not germane in our setting, because we have assumed that players have no information to make such assessments. It is in this sense that our explanation complements the explanation in terms of costs by an argument based on group dynamics.

In what follows, we present the results, where we first set up the model and then deliver our main conclusions. We discuss the implications of our results in section 3. Further details about the applied methodology are provided in the Methods section.

Results

Public goods game with directional learning. In the public goods game, each player i in the population $N = 1, 2, …, n$ chooses whether to contribute ($c_i = 1$) or not to contribute ($c_i = 0$) to the common pool. Given a fixed *rate of return* $r > 0$, the resulting payoff of player i is then $u_i = (1 - c_i) + (r/n) * \sum_{j \in N} c_j$. We shall call r/n the game's *marginal per-capita rate of return* and denote it as R. Note that for simplicity, but without loss of generality, we have assumed that the group is the whole population. In the absence of restrictions on the interaction range of players[38], i.e., in well-mixed populations, the size of the groups and their formation can be shown to be of no relevance in our case, as long as R rather than r is considered as the effective rate of return.

The directional learning dynamics are implemented as follows. Suppose the above game is infinitely repeated at time steps $t = 0$, 1, 2, …, and suppose further that i, at time t, plays $c_i^t = 1$ with probability $p_i^t \in [\delta, 1 - \delta]$ and $c_i^t = 0$ with probability $(1 - p_i^t)$. Let the vector of contribution probabilities p^t describe the state of the game at time t. We initiate the game with all p_i^0 lying on the δ-grid between 0 and 1, while subsequently individual mixed strategies evolve randomly subject to the following three "directional bias" rules:

upward: if $u_i(c_i^t) > u_i(c_i^{t-1})$ and $c_i^t > c_i^{t-1}$, or if $u_i(c_i^t) < u_i(c_i^{t-1})$ and $c_i^t < c_i^{t-1}$, then $p_i^{t+1} = p_i^t + \delta$ if $p_i^t < 1$; otherwise, $p_i^{t+1} = p_i^t$.

neutral: if $u_i(c_i^t) = u_i(c_i^{t-1})$ and/or $c_i^t = c_i^{t-1}$, then $p_i^{t+1} = p_i^t$, $p_i^t + \delta$, or $p_i^t - \delta$ with equal probability if $0 < p_i^t < 1$; otherwise, $p_i^{t+1} = p_i^t$.

downward: if $u_i(c_i^t) > u_i(c_i^{t-1})$ and $c_i^t < c_i^{t-1}$, or if $u_i(c_i^t) < u_i(c_i^{t-1})$ and $c_i^t > c_i^{t-1}$, then $p_i^{t+1} = p_i^t - \delta$ if $p_i^t > 0$; otherwise, $p_i^{t+1} = p_i^t$.

Note that the second, neutral rule above allows random deviations from any intermediate probability $0 < p_i < 1$. However, $p_i = 0$ and $p_i = 1$ for all i are absorbing state candidates. We therefore introduce perturbations to this directional learning dynamics and study the resulting stationary states. In particular, we consider perturbations of order ϵ such that, with probability $1 - \epsilon$, the dynamics are governed by the original three "directional bias" rules. However, with probability ϵ, either $p_i^{t+1} = p_i^t$, $p_i^{t+1} = p_i^t - \delta$ or $p_i^{t+1} = p_i^t + \delta$ happens equally likely (with probability $\epsilon/3$) but of course obeying the $p_i^{t+1} \in [0,1]$ restriction.

Provisioning of public goods. We begin with a formal definition of the *k–strong equilibrium*. In particular, a pure strategy imputation $s*$ is a k-strong equilibrium of our (symmetric) public goods game if, for all $C \subseteq N$ with $|C| \le k$, $u_i\left(s_C^*; s_{N \setminus C}^*\right) \ge u_i\left(s_C'; s_{N \setminus C}^*\right)$ for all $i \in C$ for any alternative pure strategy set s_C' for C. As noted in the previous

section, this definition bridges, one the one hand, the concept of the Nash equilibrium in pure strategies[21] in the sense that any k–strong equilibrium with $k > 0$ is also a Nash equilibrium, and, on the other hand, that of the (Aumann-)strong equilibrium[39,40] in the sense that any k–strong equilibrium with $k = n$ is Aumann strong. Equilibria in between (for $1 < k < n$) are "more stable" than a Nash equilibrium, but "less stable" than an Aumann-strong equilibrium.

The maximal k-strengths of the equilibria in our public goods game as a function of r are depicted in Fig. 1 for $n = 16$. The cyan-shaded region indicates the "public bad game" region for $r < 1$ ($R < 1/n$), where the individual and the public motives in terms of the Nash equilibrium of the game are aligned towards defection. Here $c_i = 0$ for all i is the unique Aumann-strong equilibrium, or in terms of the definition of the k–strong equilibrium, $c_i = 0$ for all i is k–strong for all $k \in [1, n]$. The magenta-shaded region indicates the typical public goods game for $1 < r < n$ ($1/n < R < 1$), where individual and public motives are conflicting. Here there exists no Aumann-strong equilibria. The outcome $c_i = 0$ for all i is the unique Nash equilibrium, and that outcome is also k-strong equilibrium for some $k \in [1, n)$, where the size of k depends on r and n in that $\partial k/\partial r \leq 0$ while $\partial k/\partial n \geq 0$. Finally, the gray-shaded region indicates the unconflicted public goods game for $r > n$ ($R > 1$), where individual and public motives are again aligned, but this time towards cooperation. Here $c_i = 1$ for all i abruptly becomes the unique Nash and Aumann-strong equilibrium, or equivalently the unique k–strong equilibrium for all $k \in [1, n]$.

If we add perturbations of order ϵ to the unperturbed public goods game with directional learning that we have introduced in the previous section 2, there exist stationary distributions of p_i and the

following proposition can be proven. In the following, we denote by "k" the maximal k–strength of an equilibrium.

Proposition: As $t \to \infty$, starting at any p^0, the expectation with respect to the stationary distribution is $E[p^t] > 1/2$ if $R \geq 1$ and $E[p^t] < 1/2$ if $R < 1$. $\partial E[p^t]/\partial \epsilon < 0$ if $R \geq 1$, and $\partial E[p^t]/\partial \epsilon > 0$ if $R < 1$. Moreover, $\partial E[p^t]/\partial \delta > 0$, and $\partial E[p^t]/\partial \delta < 0$ if $R \geq 1$. Finally, $\partial E[p^t]/\partial k < 0$ if $R < 1$.

We begin the proof by noting that the perturbed process given by our dynamics results in an irreducible and aperiodic Markov chain, which has a unique stationary distribution. When $\epsilon = 0$, any absorbing state must have $p_i^t = 0$ or 1 for all players. This is clear from the positive probability paths to either extreme from intermediate states given by the unperturbed dynamics. We shall now analyze whether $p_i^t = 0$ or 1, given that $p_j^t = 0$ or 1 for all $j \neq i$, has a larger attraction given the model's underlying parameters.

If $R \geq 1$, the probability path for any player to move from $p_i^t = 0$ to $p_i^{t+T} = 1$ in some $T = 1/\delta$ steps requires a single perturbation for that player and is therefore of the order of a single ϵ. By contrast, the probability for any player to move from $p_i^t = 1$ to $p_i^{t+T} = 0$ in T steps is of the order ϵ^3, because at least two other players must increase their contribution in order for that player to experience a payoff increase from his non-contribution. Along any other path or if p^t is such that there are not two players j with $p_j^t = 0$ to make this move, then the probability for i to move from $p_i^t = 1$ to $p_i^{t+T} = 0$ in T steps requires even more perturbations and is of higher order. Notice that, for any one player to move from $p_i^t = 0$ to $p_i^{t+T} = 1$ we need at least two players to move away from $p_i^t = 0$ along the least-resistance paths. Because contributing 1 is a best reply for all $R \geq 1$, those two players will also continue to increase if continuing to contribute 1. Notice that the length of the path is $T = 1/\delta$ steps, and that the path requires no perturbations along the way, which is less likely the smaller δ.

If $R < 1$, the probability for any player to move from $p_i^t = 1$ to $p_i^{t+T} = 0$ in some $T = 1/\delta$ steps requires a single perturbation for that player and is therefore of the order of a single ϵ. By contrast, the probability for any player to move from $p_i^t = 0$ to $p_i^{t+T} = 1$ in some T steps is at least of the order ϵ^k, because at least k players (corresponding to the maximal k-strength of the equilibrium) must contribute in order for all of these players to experience a payoff increase. Notice that k decreases in R. Again, the length of the path is $T = 1/\delta$ steps, and that path requires no perturbations along the way, which is less likely the smaller δ. With this, we conclude the proof of the proposition. However, it is also worth noting a direct corollary of the proposition; namely, as $\epsilon \to 0$, $E[p^t] \to 1$ if $R \geq 1$, and $E[p^t] \to 0$ if $R < 1$.

Lastly, we simulate the perturbed public goods game with directional learning and determine the actual average contribution levels in the stationary state. Color encoded results in dependence on the normalized rate of return R and the responsiveness of players to the success of their past actions δ (alternatively, the sensitivity of the individual learning process) are presented in Fig. 2 for $\epsilon = 0.1$. Small values of δ lead to a close convergence to the respective Nash equilibrium of the game, regardless of the value of R. As the value of δ increases, the pure Nash equilibria erode and give way to a mixed outcome. It is important to emphasize that this is in agreement, or rather, this is in fact a consequence of the low k–strengths of the non-contribution pure equilibria (see Fig. 1). Within intermediate to large δ-values the Nash equilibria are implemented in a zonal rather than pinpoint way. When the Nash equilibrium is such that all players contribute ($R > 1$), then small values of δ lead to more efficient aggregate play (recall any such equilibrium is n–strong). Conversely, by the same logic, when the Nash equilibrium is characterized by universal free-riding, then larger values of δ lead to more efficient aggregate play. Moreover, the precision of implementation also depends on the rate of return in the sense that uncoordinated
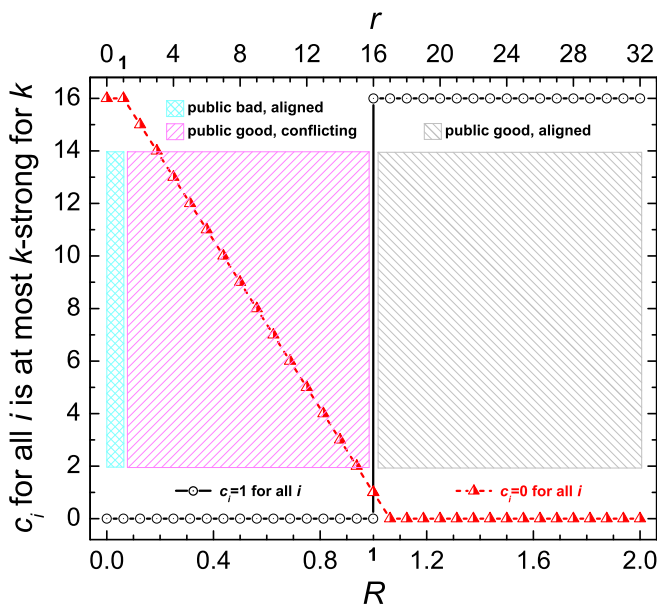

Figure 1 | The maximal k-strength of equilibria in the studied public goods game with directional learning. As an example, we consider the population size being $n = 16$. As the rate of return r increases above 1, the Aumann-strong (n–strong) $c_i = 0$ for all i (full defection) equilibrium looses strength. It is still the unique Nash equilibrium, but its maximal strength is bounded by $k = 17 - r$. As the rate of return r increases further above n ($R > 1$), the $c_i = 1$ for all i (full cooperation) equilibrium suddenly becomes Aumann-strong (n–strong). Shaded regions denote the public bad game ($r < 1$), and the public goods games with conflicting ($1 < r < n$) and aligned ($R > 1$) individual and public motives in terms of the Nash equilibrium of the game (see main text for details). We note that results for other population and/or group sizes are the same over R, while r and the slope of the red line of course scale accordingly.
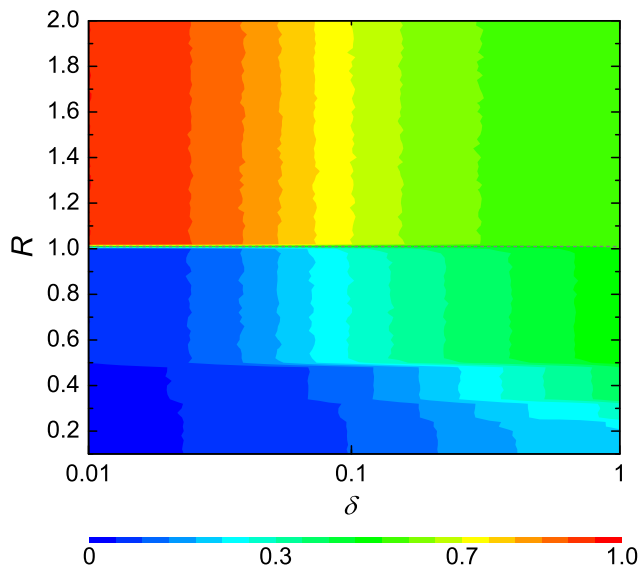
Figure 2 | Color-encoded average contribution levels in the unperturbed public goods game with directional learning. Simulations confirm that, with little directional learning sensitivity (i.e. when δ is zero or very small), for the marginal per-capita rate of return $R > 1$ the outcome $c_i = 1$ for all i is the unique Nash and Aumann-strong equilibrium. For $R = 1$ (dashed horizontal line), any outcome is a Nash equilibrium, but only $c_i = 1$ for all i is Aumann-strong while all other outcomes are only Nash equilibria. For $R < 1$, $c_i = 0$ for all i is the unique Nash equilibrium, and its maximal k–strength depends on the population size. This is in agreement with results presented in Fig. 1. Importantly, however, as the responsiveness of players increases, contributions to the common pool become significant even in the defection-prone $R < 1$–region. In effect, individuals' (mis)learn what is best for them and end up contributing even though this would not be a unilateral best reply. Similarly, in the $R > 1$ region free-riding starts to spread despite of the fact that it is obviously better to cooperate. For both these rather surprising and counterintuitive outcomes to emerge, the only thing needed is directional learning.

deviations of groups of players lead to more efficient outcomes the higher the rate of return. In other words, the free-riding problem is mitigated if group deviations lead to higher payoffs for every member of an uncoordinated deviation group, the minimum size of which (that in turn is related to the maximal k– strength of equilibrium) is decreasing with the rate of return.

Simulations also confirm that the evolutionary outcome is qualitatively invariant to: i) The value of ϵ as long as the latter is bounded away from zero, although longer convergence times are an inevitable consequence of very small ϵ values (see Fig. 3); ii) The replication of the population (i.e., making the whole population a group) and the random remixing between groups; and iii) The population size, although here again the convergence times are the shorter the smaller the population size. While both ii and iii are a direct consequence of the fact that we have considered the public goods game in a well-mixed rather than a structured population (where players would have a limited interaction range and where thus pattern formation could play a decisive role[38]), the qualitative invariance to the value of ϵ is elucidated further in Fig. 3. We would like to note that by "qualitative invariance" it is meant that, regardless of the value of $\epsilon > 0$, the population always diverges away from the Nash equilibrium towards a stable mixed stationary state. But as can be observed in Fig. 3, the average contribution level and its variance both increase slightly as ϵ increases. This is reasonable if one considers ϵ as an exploration or mutation rate. More precisely, it can be observed that, the lower the value of ϵ, the longer it takes for the population to move away from the Nash equilibrium where everybody contributes zero in the case
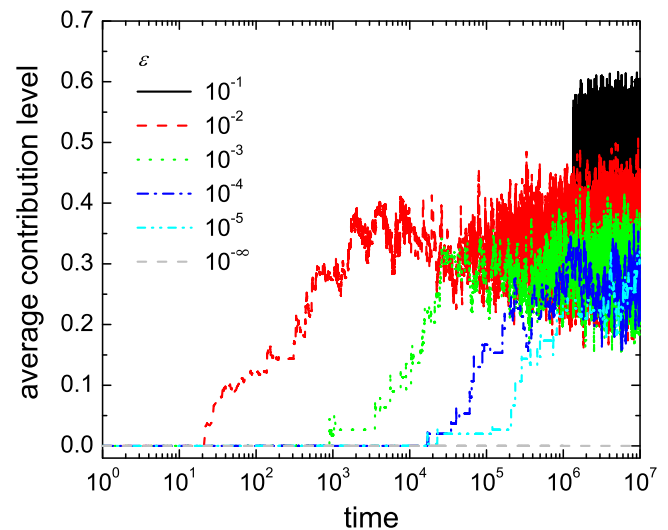

Figure 3 | Time evolution of average contribution levels, as obtained for $R = 0.7$, $\delta = 0.1$ and different values of ϵ (see legend). If only $\epsilon > 0$, the Nash equilibrium erodes to a stationary state where at least some members of the population always contribute to the common pool. There is a discontinuous transition to complete free-riding (defection) as $\epsilon \to 0$. Understandably, the lower the value of ϵ (the smaller the probability for the perturbation), the longer it may take for the drift to gain on momentum and for the initial deviation to evolve towards the mixed stationary state. Note that the time horizontally is in logarithmic scale.

that $1/n < R < 1$ (which was also the initial condition for clarity). However, as soon as initial deviations (from $p_i = 0$ in this case) emerge (with probability proportional to ϵ), the neutral rule in the original learning dynamics takes over, and this drives the population towards a stable mixed stationary state. Importantly, even if the value of ϵ is extremely small, the random drift sooner or later gains momentum and eventually yields similar contribution levels as those attainable with larger values of ϵ. Most importantly, note that there is a discontinuous jump towards staying in the Nash equilibrium, which occurs only if ϵ is exactly zero. If ϵ is bounded away from zero, then the free-riding Nash equilibrium erodes unless it is n–strong (for very low values of $R \leq 1/n$).

Discussion

We have introduced a public goods game with directional learning, and we have studied how the level of contributions to the common pool depends on the rate of return and the responsiveness of individuals to the successes and failures of their own past actions. We have shown that directional learning alone suffices to explain deviations from the Nash equilibrium in the stationary state of the public goods game. Even though players have no strategically relevant information about the game and/or about each others' actions, the population could still end up in a mixed stationary state where some players contributed at least part of the time although the Nash equilibrium would be full free-riding. Vice versa, defectors emerged where cooperation was clearly the best strategy to play. We have explained these evolutionary outcomes by introducing the concept of k–strong equilibria, which bridge the gap between Nash and Aumann-strong equilibria. We have demonstrated that the lower the maximal k–strength and the higher the responsiveness of individuals to the consequences of their own past strategy choices, the more likely it is for the population to (mis)learn what is the objectively optimal unilateral (Nash) play.

These results have some rather exciting implications. Foremost, the fact that the provisioning of public goods even under adverse conditions can be explained without any sophisticated and often

lengthy arguments involving selflessness or social preference holds promise of significant simplifications of the rationale behind seemingly irrational individual behavior in sizable groups. It is simply enough for a critical number (depending on the size of the group and the rate of return) of individuals to make a "wrong choice" at the same time once, and if only the learning process is sufficiently fast or naive, the whole subpopulation is likely to adopt this wrong choice as their own at least part of the time. In many real-world situations, where the rationality of decision making is often compromised due to stress, propaganda or peer pressure, such "wrong choices" are likely to proliferate. As we have shown in the context of public goods games, sometimes this means more prosocial behavior, but it can also mean more free-riding, depending only on the rate of return.

The power of directional (mis)learning to stabilize unilaterally suboptimal game play of course takes nothing away from the more traditional and established explanations, but it does bring to the table an interesting option that might be appealing in many real-life situations, also those that extend beyond the provisioning of public goods. Fashion trends or viral tweets and videos might all share a component of directional learning before acquiring mainstream success and recognition. We hope that our study will be inspirational for further research in this direction. The consideration of directional learning in structured populations[41,42], for example, appears to be a particularly exciting future venture.

Methods

For the characterization of the stationary states, we introduce the concept of k–strong equilibria, which nests both the Nash equilibrium[21] and the Aumann-strong equilibrium[39,40] as two special cases. While the Nash equilibrium describes the robustness of an outcome against unilateral (1-person) deviations, the Aumann-strong equilibrium describes the robustness of an outcome against the deviations of any subgroup of the population. An equilibrium is said to be (Aumann-)strong if it is robust against deviations of the whole population or indeed of any conceivable subgroup of the population, which is indeed rare. Our definition of the k–strong equilibrium bridges the two extreme cases, measuring the size of the group $k \geq 1$ (at or above Nash) and hence the degree to which an equilibrium is stable. We note that our concept is related to coalition-proof equilibrium[43,44]. In the public goods game, the free-riding Nash equilibrium is typically also more than 1–strong but never n–strong. As we will show, the maximal strength k of an equilibrium translates directly to the level of contributions in the stationary distribution of our process, which is additionally determined by the normalized rate of return R and the responsiveness of players to the success of their past actions δ, i.e., the sensitivity of the individual learning process.

1. Bowles, S. & Gintis, H. *A Cooperative Species: Human Reciprocity and Its Evolution* (Princeton University Press, Princeton, NJ, 2011).
2. Hrdy, S. B. *Mothers and Others: The Evolutionary Origins of Mutual Understanding* (Harvard University Press, Cambridge, MA, 2011).
3. Nowak, M. A. & Highfield, R. *SuperCooperators: Altruism, Evolution, and Why We Need Each Other to Succeed* (Free Press, New York, 2011).
4. Rand, D. A. & Nowak, M. A. Human cooperation. *Trends in Cognitive Sciences* **17**, 413–425 (2013).
5. Maynard Smith, J. *Evolution and the Theory of Games* (Cambridge University Press, Cambridge, U.K., 1982).
6. Weibull, J. W. *Evolutionary Game Theory* (MIT Press, Cambridge, MA, 1995).
7. Hofbauer, J. & Sigmund, K. *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, U.K., 1998).
8. Mesterton-Gibbons, M. *An Introduction to Game-Theoretic Modelling, 2nd Edition* (American Mathematical Society, Providence, RI, 2001).
9. Nowak, M. A. *Evolutionary Dynamics* (Harvard University Press, Cambridge, MA, 2006).
10. Myatt, D. P. & Wallace, C. When Does One Bad Apple Spoil the Barrel? An Evolutionary Analysis of Collective Action. *Rev. Econ. Studies* **75**, 499–527 (2008).
11. Mesterton-Gibbons, M. & Dugatkin, L. A. Cooperation among unrelated individuals: Evolutionary factors. *Q. Rev. Biol.* **67**, 267–281 (1992).
12. Nowak, M. A. Five rules for the evolution of cooperation. *Science* **314**, 1560–1563 (2006).
13. Rand, D. G., Dreber, A., Ellingsen, T., Fudenberg, D. & Nowak, M. A. Positive interactions promote public cooperation. *Science* **325**, 1272–1275 (2009).
14. 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).
15. Santos, F. C., Santos, M. D. & Pacheco, J. M. Social diversity promotes the emergence of cooperation in public goods games. *Nature* **454**, 213–216 (2008).
16. Rand, D. G., Nowak, M. A., Fowler, J. H. & Christakis, N. A. Static network structure can stabilize human cooperation. *Proc. Natl. Acad. Sci. USA* **111**, 17093–17098 (2014).
17. Isaac, M. R., McCue, K. F. & Plott, C. R. Public goods provision in an experimental environment. *J. Pub. Econ.* **26**, 51–74 (1985).
18. Ledyard, J. O. Public goods: A survey of experimental research. In The Handbook of Experimental Economics, Kagel, J. H. & Roth, A. E., editors, 111–194. Princeton University Press, Princeton, NJ(1997).
19. Chaudhuri, A. Sustaining cooperation in laboratory public goods experiments: a selective survey of the literature. *Exp. Econ.* **14**, 47–83 (2011).
20. Fehr, E. & Gächter, S. Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994 (2000).
21. Nash, J. Equilibrium points in n-person games. *Proc. Natl. Acad. Sci. USA* **36**, 48–49 (1950).
22. Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).
23. Fischbacher, U., Gächter, S. & Fehr, E. Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001).
24. Fischbacher, U. & Gächter, S. Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* **100**, 541–556 (2010).
25. Palfrey, T. R. & Prisbey, J. E. Anomalous Behavior in Public Goods Experiments: How Much and Why? *Am. Econ. Rev.* **87**, 829–846 (1997).
26. Goeree, J. K. & Holt, C. A. An Explanation of Anomalous Behavior in Models of Political Participation. *Am. Polit. Sci. Rev.* **99**, 201–213 (2005).
27. Foster, D. P. & Young, H. P. Regret testing: learning to play nash equilibrium without knowing you have an opponent. *Theor. Econ.* **1**, 341–367 (2006).
28. Young, H. P. Learning by trial and error. *Games Econ. Behav.* **65**, 626–643 (2009).
29. Roth, A. E. & Erev, I. Learning in extensive-form games: Experimental data and simple dynamic models in the intermediate term. *Games Econ. Behav.* **8**, 164–212 (1995).
30. Erev, I. & Roth, A. E. Predicting how people play games: Reinforcement learning in experimental games with unique, mixed strategy equilibria. *Am. Econ. Rev.* **88**, 848–881 (1998).
31. Selten, R. & Stoecker, R. End behavior in sequences of finite Prisoner's Dilemma supergames A learning theory approach. *J. Econ. Behav. Organ.* **7**, 47–70 (1986).
32. Selten, R. & Buchta, J. Experimental Sealed Bid First Price Auctions with Directly Observed Bid Functions. *Discuss. Pap. Ser. B* (1994).
33. Bayer, R.-C., Renner, E. & Sausgruber, R. Confusion and learning in the voluntary contributions game. *Exp. Econ.* **16**, 478–496 (2013).
34. Young, H. P., Nax, H., Burton-Chellew, M. & West, S. Learning in a Black Box. *Econ. Ser. Work. Pap.* (2013).
35. Sauermann, H. & Selten, R. Anspruchsanpassungstheorie der unternehmung. *J. Institut. Theor. Econ.* **118**, 577–597 (1962).
36. Nax, H. H., Pradelski, B. S. R. & Young, H. P. Decentralized dynamics to optimal and stable states in the assignment game. *Proc. IEEE* **52**, 2398–2405 (2013).
37. Laslier, J.-F. & Walliser, B. Stubborn learning. *Theor. Decis.* **forthcoming**, 1 (2014).
38. Perc, M., Gómez-Gardeñes, J., Szolnoki, A., Floría, Y. & Moreno, L. M. Evolutionary dynamics of group interactions on structured populations: a review. *J. R. Soc. Interface* **10**, 20120997 (2013).
39. Aumann, R. J. Subjectivity and correlation in randomized strategies. *J. Math. Econ.* **1**, 67–96 (1974).
40. Aumann, R. J. Correlated equilibrium as an expression of bayesian rationality. *Econometrica* **55**, 1–18 (1987).
41. Szabó, G. & Fáth, G. Evolutionary games on graphs. *Phys. Rep.* **446**, 97–216 (2007).
42. Perc, M. & Szolnoki, A. Coevolutionary games – a mini review. *BioSystems* **99**, 109–125 (2010).
43. Bernheim, B. D., Peleg, B. & Whinston, M. D. Coalition-proof equilibria i. concepts. *J. Econ. Theor.* **42**, 1–12 (1987).
44. Moreno, D. & Wooders, J. Coalition-proof equilibrium. *Games Econ. Behav.* **17**, 82–112 (1996).

Author contributions

H.H.N. and M.P. designed and performed the research as well as wrote the paper.

Additional information