Methods

The methods are based on a straightforward application of evolutionary game dynamics for finite populations. First, we discuss the dynamics based on social learning together with analytical approximations and implementations of individual based simulations. In section 1 we describe what happens in the absence of punishment, i.e. for three strategies: X-players participate, and contribute an amount c to the public goods (PG) game; Y-players participate, but do not contribute; and Z-players do not participate. With X, Y and Z, we also denote the number of players using the corresponding strategy (and $M = X + Y + Z$ is the total population size, which we assume to be constant). In sections 2 and 3, we additionally consider V-players, who contribute to the PG as well as to a punishment pool, with and without second-order punishment. In sections 4 and 5 this is repeated for W-players, who contribute to the PG game and then peer-punish (with or without second order punishment). Finally, in section 6, we address the competition of peer and pool punishment (i.e., $M = X + Y + Z + V + W$). In each section, we compute the average payoff values, and analyze a limiting case ('strong imitation').

Social learning

We assume that two players i and j are randomly chosen. Their expected payoff values P_i and P_j depend on the strategies of the two players and on the frequencies $X, Y, ...$ of the strategies. There are many ways to model social learning. We shall assume that player i adopts the strategy of player j with a probability which is an increasing function of the payoff difference $P_j - P_i$. A frequently used choice for this probability is

$$\frac{1}{1 + \exp\left[-s(P_j - P_i)\right]}, \tag{1}$$

where the 'imitation strength' $s \geq 0$ measures how strongly the players are basing their decisions on payoff comparisons[31,32,33,34,35]. For $s \to 0$ (or for $P_i = P_j$), a coin toss decides whether to imitate or not. Small values of s correspond to a regime we call 'weak imitation'. In this case, imitation is basically random, but more successful players are imitated slightly more often. For $s \to +\infty$, i.e., 'strong imitation', a more successful player is always imitated, a less successful never. The homogeneous populations correspond to absorbing states of the stochastic process: once such a state is reached, imitation cannot produce any change. Thus we shall assume that additionally, with a certain probability $\mu > 0$ (the exploration rate), a player switches randomly to another strategy without imitating another player. The resulting Markov chain has a stationary distribution which, if the population size M is large and there are more than two strategies, requires considerable efforts to compute numerically. In addition to individual-based computer simulations, we shall consider the limiting case of very small exploration rates, the so-called 'adiabatic' case. In that case, if in a homogeneous population a single dissident arises, then its fate (elimination or fixation) will be settled through the imitation process before the next exploration step occurs.

More precisely, let us assume that there are d strategies 1,...,d. By X_k we denote the number of players using strategy k ($\Sigma X_k = M$). The homogeneous population with $X_k = M$ will be denoted by All_k. With probability $\mu/(d-1)$, a single individual switches from k to $l \neq k$. The probability that subsequently, imitation leads to the fixation of the dissident strategy l is denoted by ρ_{kl}. The fixation probability can be computed by the formulas known from the theory of birth-death processes[34,35,36],

$$\rho_{kl} = \frac{1}{1 + \sum_{q=1}^{M-1} \prod_{X_l=1}^{q} \frac{T_{l\to k}(X_l)}{T_{k\to l}(X_l)}}.$$

In our case, the probability that one out of X_l players with strategy l is chosen as a focal player and imitates one of the $X_k = M - X_l$ players with strategy k is given by

$$T_{l\to k}(X_l) = \frac{X_l}{M}\frac{M - X_l}{M}\frac{1}{1 + \exp\left[-s(P_k - P_l)\right]},$$

where payoffs P_l and P_k depend on the number of l and k players, i.e., on X_l and $X_k = M - X_l$. The fixation probability ρ_{kl} simplifies to

$$\rho_{kl} = \frac{1}{1 + \sum_{q=1}^{M-1} \exp\left[s \sum_{X_l=1}^{q} (P_k - P_l)\right]} \qquad (2)$$

This form makes it easy to address the limit of strong imitation, $s \to +\infty$.

The probability of a transition from All_k to All_l is $\mu\rho_{kl}/(d-1)$. If the $d \times d$ transition matrix is mixing, it has a unique normalized left eigenvector to the eigenvalue 1, and this is the stationary distribution which describes the percentage of time (in the long run) spent by the state of the population in the vicinity of the homogeneous state All_k. One can show[37,38] that the stationary distribution of the full system converges for $\mu \to 0$ to the stationary distribution of this 'embedded' Markov chain on the homogeneous states whose transition probabilities from All_k to All_l (for $k \neq l$) are given by $\rho_{kl}/(d-1)$ (μ cancels out). For numerical confirmation, we refer to Fig. 1 (in the Supplementary Information).

Simulations and numerical solutions

The individual based simulations mimic the social learning dynamics outlined above for arbitrary exploration rates, μ. Each individual achieves an average payoff based on random sampling of the interaction groups. This reflects a situation where individuals interact often and only occasionally reassess and update their strategies. With probability μ, players randomly adopt any other available strategy, and with probability $1-\mu$, they update according to Eq. (1). The long-run mean frequency of each strategy is determined by averaging over $T > 10^7$ updates per player.

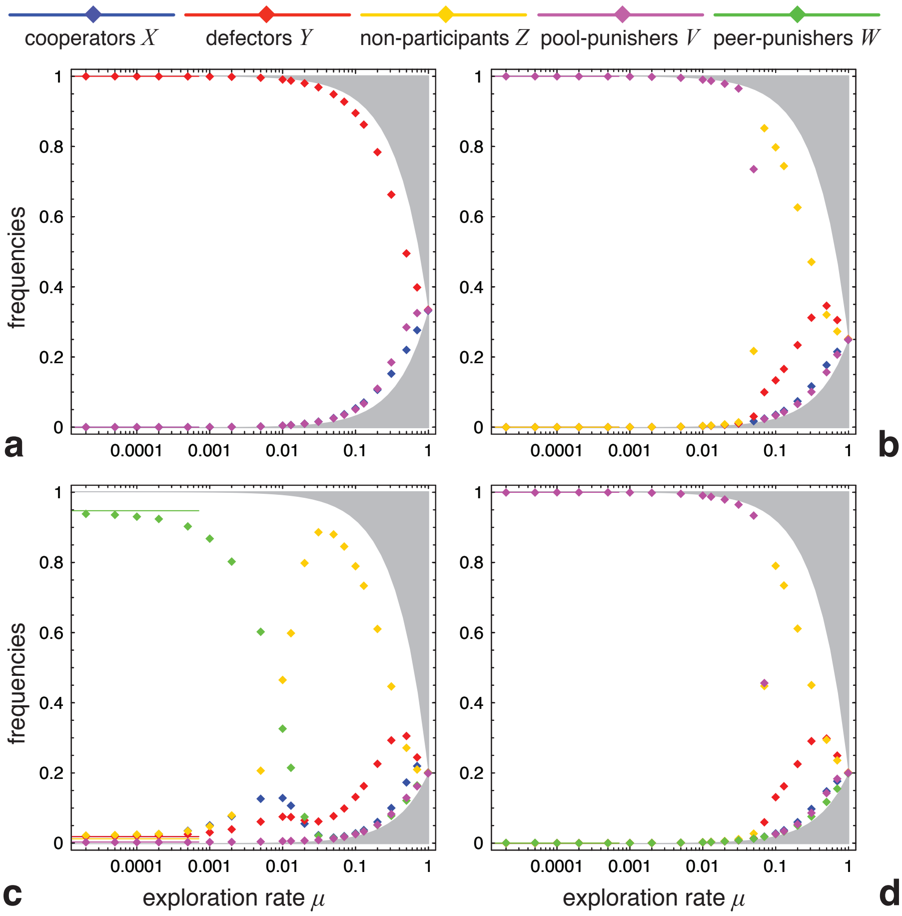

Figure 1: Scenarios of sanctioning in public goods games for variable exploration rates μ. For large μ, random exploration dominates, which results in roughly equal average frequencies of all available strategies. All strategies tend to be present in the population at all times. Because a fraction μ of the population always mutates, the minimum frequency of each strategy is μ/d (for d strategies) and the grey-shaded areas are inaccessible to the process. For smaller μ, the population spends increasing amounts of time in homogeneous states between subsequent mutations. (**a**) even though sufficiently large μ can push the population from the region of attraction of *AllY* to *AllV*, the population is unable to remain near this cooperative state and defection dominates. (**b**) in voluntary public goods games, pool punishers prevail except for large μ, where risk-averse non-participants take over. (**c, d**) whether peer-punishers or pool-punishers prevail in voluntary public goods games depends on second-order punishment (c.f. Fig. 3). Without second-order punishment, peer-punishers prevail, **c**, but exploration rates μ of order 0.01 are large enough to destroy cooperation, so that non-participants prevail. With second-order punishment, pool-punishers dominate, (**d**), and since *AllV* is strongly attracting, much larger μ-values are required before cooperation is destroyed. The data points, obtained by having each player update 10^7 times, are supported by analytical approximations (solid lines) for very small values of μ. Parameters: same as in Fig. 2 (main text), but with fixed imitation strength $s = 10$ and variable exploration rate μ.

Numerical computations of the stationary distribution (for small μ) based on the fixation probabilities in Eq. (2) and individual-based simulations show that the results hold not only for the limiting case, but for a large set of plausible values for the parameters μ, s, c, r, γ, β, M, G and B, see[37,38]. For online experimentation, we refer to http://www.hanneloredesilva.at/sanctions and the *VirtualLabs* at http://www.univie.ac.at/virtuallabs.

1 No punishment

In a population consisting of X contributors and $Y = M - X$ defectors, random samples of N individuals play the PG game. A co-operator obtains on average

$$\sum_{k=0}^{N-1} \frac{\binom{X-1}{k}\binom{M-X}{N-1-k}}{\binom{M-1}{N-1}}(rc\frac{k}{N-1} - c)$$

$$= rc\frac{X-1}{M-1} - c$$

(the summation variable k represents the number of other contributors, sampling is done without replacement, probabilities obey the hypergeometric distribution). Defectors obtain from the public good on average

$$\sum_{k=0}^{N-1} \frac{\binom{X}{k}\binom{M-1-X}{N-1-k}}{\binom{M-1}{N-1}} rc\frac{k}{N-1} = rc\frac{X}{M-1}.$$

Let us now assume that the population consists of X contributors, Y defectors and Z non-participants. The probability that the other $N-1$ players of a sample are unwilling to participate is

$$\frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}.$$

Hence the average payoff for defectors is

$$P_Y = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}\sigma + (1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}})rc\frac{M-Z-Y}{M-Z-1}, \quad (3)$$

that for contributors

$$P_X = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}\sigma + (1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}})c(r\frac{M-Z-Y-1}{M-Z-1} - 1), \quad (4)$$

and of course $P_Z = \sigma$ (cf.[39]). The three strategies form a Rock-Paper-Scissors cycle. More precisely, if $Z = 0$, defectors do always better than contributors ($P_Y > P_X$); but in the absence of contributors ($X = 0$), non-participants do better than defectors ($P_Z \geq P_Y$, with equality if and only if $Y = 1$); and in the absence of defectors ($Y = 0$), contributors do better than non-participants ($P_X \geq P_Z$, with equality if and only if $X = 1$).

The resulting stochastic process exhibits cycling behavior. It is clear that if most players use strategy X, then Y-players do better, and if most players use strategy Y, the Z-players do better. It is less obvious to see why, in a population where most players use Z, X players do best. But if most players are non-participants, PG groups are small. In that case, random fluctuations can lead to groups with mostly X-players, who do well, so that many imitate them. This relates to Simpson's paradox[39]

For small exploration rates, the embedded Markov chain describing the transitions between $AllX$, $AllY$ and $AllZ$ is given by

$$\begin{pmatrix} 1-\frac{1}{2}\rho_{XY}-\frac{1}{2}\rho_{XZ} & \frac{1}{2}\rho_{XY} & \frac{1}{2}\rho_{XZ} \\ \frac{1}{2}\rho_{YX} & 1-\frac{1}{2}\rho_{YX}-\frac{1}{2}\rho_{YZ} & \frac{1}{2}\rho_{XY} \\ \frac{1}{2}\rho_{ZX} & \frac{1}{2}\rho_{ZY} & 1-\frac{1}{2}\rho_{ZX}-\frac{1}{2}\rho_{ZY} \end{pmatrix}. \tag{5}$$

The normalized left eigenvector to the eigenvalue 1 gives the stationary distribution, which by Eq. (2) can be evaluated numerically as a function of the imitation strength s. This is the basis of the analytical approximation under weak mutation shown in Figs. 2 and 3 (main text) and Fig. 1 (SI). In the limiting case of strong imitation, $s \to +\infty$, the embedded Markov chain simplifies significantly; for instance, $\rho_{XZ} = 0$ and $\rho_{XY} = 1$. Interestingly, $\rho_{ZX} = 1/2$. The reason is that a single X-mutant in a population of Z-players cannot participate in any game, and has payoff σ like the other non-participants. The next change obtained through the imitation process is equally likely to revert the X-player to the fold or to produce a second X-player. From then on, X-players have an expected payoff larger than σ and will increase to fixation. The transition matrix between $AllX$, $AllY$ and $AllZ$ reduces for $s \to \infty$ to

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & \frac{1}{2} & \frac{1}{2} \\ \frac{1}{4} & 0 & \frac{3}{4} \end{pmatrix} \tag{6}$$

and the stationary distribution (the left eigenvector to the eigenvalue 1) is given by $(\frac{1}{4}, \frac{1}{4}, \frac{1}{2})$. The same argument used to simplify the dynamics for small μ and $s \to \infty$ is used in the discussion about punishment below.

2 Pool Punishment

Let us now assume in addition that V of the M players engage in pool punishment. This means that when N_v of them find themselves in a PG game, they not only contribute c to the public good, but pay an extra fee G towards the punishment pool. The fine of each exploiter will be proportional to the number of punishers, and hence of the form $N_v B$, for some $B > 0$.

For the moment, we neglect the possibility of second-order punishment (i.e. the punishment of non-punishers). The payoffs for non-participants and contributors are therefore unaffected. The payoff for pool-punishers satisfies

$$P_V = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}\sigma + (1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}})[c(r\frac{M-Z-Y-1}{M-Z-1} - 1) - G]. \tag{7}$$

Indeed, punishers only pay a fee into the pool if another player is willing to participate. The payoff for defectors is

$$P_Y = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}\sigma + (1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}})cr\frac{M-Z-Y}{M-Z-1} - \frac{B(N-1)V}{M-1}. \tag{8}$$

(If there is at least one punisher among the $N - 1$ co-players in the sample, the PG game is played.)

If we assume that a population of pool-punishers does better than the non-participants, i.e., that

$$\sigma < (r-1)c - G, \tag{9}$$

then we obtain, in the case of strong imitation, the following embedded Markov chain for the transitions between the states $AllX$, $AllY$, $AllZ$ and $AllV$:

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \end{pmatrix}. \tag{10}$$

The explanation for the $\frac{1}{6}$ terms is the same as that for the $\frac{1}{4}$ in (6). If, in an $AllZ$-population, a mutation produces a single X-player, this player finds no partners for the PG game and obtains the same payoff as the non-participants. The next change obtained through the imitation process is equally likely to revert the X-player to the fold or to produce a second X-player. From then on, X-players have an expected payoff larger than σ and will increase to fixation.

The unique stationary distribution is given by $\frac{1}{7}(2,2,2,1)$. This corresponds to two rock-paper-scissors cycles, one from $AllY$ to $AllZ$ to $AllX$ and back to $AllY$ again, the other (four-membered) from $AllY$ to $AllZ$ to $AllV$ to $AllX$ and back to $AllY$. Computer simulations confirm that the four homogeneous states supersede each other.

If the game is compulsory, i.e., if there are no Z-players, then the transitions between the states $AllX$, $AllY$ and $AllV$ are given by

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 \\ 0 & 1 & 0 \\ \frac{1}{2} & 0 & \frac{1}{2} \end{pmatrix} \tag{11}$$

and the stationary distribution is $(0,1,0)$. Free-riders take over.

3 Second-order pool-punishers

Let us now assume that the second-order exploiters, i.e., the X-players, are also punished. Thus their payoff is given by

$$P_X = \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}}\sigma + (1 - \frac{\binom{Z}{N-1}}{\binom{M-1}{N-1}})c(r\frac{M-Z-Y-1}{M-Z-1} - 1) - \frac{B(N-1)V}{M-1}. \tag{12}$$

The other payoff values remain unchanged.

If pool-punishers can invade non-participants, i.e., (9) holds, the embedded Markov chain is given by in the case of strong imitation by

$$\begin{pmatrix} \frac{2}{3} & \frac{1}{3} & 0 & 0 \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{13}$$

The unique stationary distribution is $(0, 0, 0, 1)$, which means that punishers prevail.

4 Peer Punishment

Let us now assume instead that W players in the population engage in peer punishment. Each peer-punisher imposes a fine β on each defector in his or her sample, at a cost γ. Thus if there are N_y defectors and N_w peer-punishers in the sample, each defector pays a total fine $N_w\beta$, and each punisher incurs a cost $N_y\gamma$. We first omit second-order punishment.

In the absence of pool punishment, i.e. if $M = X + Y + Z + W$, the average payoff for punishers is

$$P_W = P_X - \frac{(N-1)Y}{M-1}\gamma \tag{14}$$

where P_X is given by (4), and the defectors' payoff is given by the expression in (3), reduced by

$$\frac{(N-1)W}{M-1}\beta. \tag{15}$$

For strong imitation, the embedded Markov chain on the states $AllX$, $AllY$, $AllZ$ and $AllW$ has the transition matrix

$$\begin{pmatrix} \frac{2}{3} - \frac{1}{3M} & \frac{1}{3} & 0 & \frac{1}{3M} \\ 0 & \frac{2}{3} & \frac{1}{3} & 0 \\ \frac{1}{6} & 0 & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{3M} & 0 & 0 & 1 - \frac{1}{3M} \end{pmatrix}. \tag{16}$$

If, for instance, W-dissidents arise in an X-population, they do as well as the residents (all contribute, no one punishes), and the fixation probability in this 'neutral case' is $1/M$. It is easy to see that this Markov chain has a unique stationary distribution, given by

$$\frac{1}{M+8}(2, 2, 2, M+2). \tag{17}$$

For instance, if the populations size is $M = 92$, then for 94 percent of the time, the population is dominated by peer punishers.

5 Second order peer-punishers

Let us now assume that peer-punishers engage in second-order punishment: thus they impose fines β on the contributors too, at a cost γ to themselves.

If $M = X + Y + Z + W$, the average payoff P_X for contributors is given by (4), reduced by the average fine

$$\frac{(N-1)W}{M-1}\beta(1 - \frac{\binom{M-Y-2}{N-2}}{\binom{M-2}{N-2}}) \quad (18)$$

and the peer-punishers' payoff by (14), reduced by the average cost

$$\frac{(N-1)X}{M-1}\gamma(1 - \frac{\binom{M-Y-2}{N-2}}{\binom{M-2}{N-2}}) \quad (19)$$

for meting out extra punishment. The term $(1 - ...)$ corresponds to having at least one defector in the sample (otherwise a punisher cannot be aware that the contributor does not punish).

In the limiting case of strong imitation, the Markov chain is exactly as before. Indeed, during the imitation process the population never consists of more than two types. Hence second-order punishment (which requires that W-players see that X-players fail to punish Y-players) will never occur.

6 The competition of pool- and peer-punishers

The outcome is: without second order punishment, pool-punishers lose and peer-punishers predominate in the long run. With second-order punishment, it is just the reverse. (We assume that pool-punishers punish peer-punisher, since these do not contribute to the punishment pool. It seems less likely that peer-punishers will punish pool-punishers, and we shall not assume it here. However, we stress that this assumption does not really matter. The reason: in a population with peer- and pool-punishers only, peer-punishment is not used and the pool-punishers do not reveal that they do not engage in it.)

Let us first consider the case without second-order punishment. For strong imitation, the embedded Markov chain describing the transitions matrix between $AllX$, $AllY$, $AllZ$, $AllV$

and $AllW$ is

$$\begin{pmatrix} \frac{3}{4}-\frac{1}{4M} & \frac{1}{4} & 0 & 0 & \frac{1}{4M} \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ \frac{1}{4} & 0 & 0 & \frac{1}{2} & \frac{1}{4} \\ \frac{1}{4M} & 0 & 0 & 0 & 1-\frac{1}{4M} \end{pmatrix}. \tag{20}$$

The unique stationary distribution is $\frac{1}{3M+23}(6, 6, 4, 1, 3M + 6)$. This means that the majority consists of peer-punishers. In the case with second-order punishment, the matrix is

$$\begin{pmatrix} \frac{3}{4} & \frac{1}{4} & 0 & 0 & 0 \\ 0 & \frac{3}{4} & \frac{1}{4} & 0 & 0 \\ \frac{1}{8} & 0 & \frac{5}{8} & \frac{1}{8} & \frac{1}{8} \\ 0 & 0 & 0 & 1 & 0 \\ \frac{1}{4M} & 0 & 0 & 0 & 1-\frac{1}{4M} \end{pmatrix}. \tag{21}$$

In this case, the stationary distribution is $(0, 0, 0, 1, 0)$. Pool-punishers win.

References

31. McFadden, D. Structural Discrete Probability Models Derived from Theories of Choice, in *Structural Analysis of Discrete Data and Econometric Applications*, Manski C.F. & McFadden D.L. (Eds), Cambridge, MIT Press (1981)

32. Blume, L.E. The statistical mechanics of strategic interaction *Games Econ. Behav.***5**,387-424 (1993)

33. Szabó, G. & Toke, C. Evolutionary prisoner's dilemma game on a square lattice, *Phys. Rev. E* **58**, 69-73 (1998)

34. Traulsen, A., Nowak, M.A. & Pacheco, J.M. Stochastic dynamics of invasion and fixation, *Phys. Rev. E* **74**, 011909 (2006)

35. Nowak, M.A. *Evolutionary Dynamics*, Cambridge MA, Harvard UP (2006)

36. Karlin, S., & Taylor, H.E. *A First Course in Stochastic Processes*, 2nd ed., Academic Press, London

37. Fudenberg, D. & Imhof, L. Imitation Processes with Small Mutations, *Journal of Economic Theory* **131**, 251-262 (2005)

38. Antal, T. & Scheuring, I. Fixation of strategies for an evolutionary game in finite populations, *Bull. Math. Biol.***68**, 1923- (2006)

39. De Silva, H., Hauert, C., Traulsen, A. & Sigmund, K. Freedom, enforcement, and the social dilemma of strong altruism, *J. Evolutionary Economics* **20** 203-217 (2010)