Supplementary Information: Emergence of responsible sanctions without second order free riders, antisocial punishment or spite

Christian Hilbe[1,*] & Arne Traulsen[1]

[1] Max Planck Institute for Evolutionary Biology, D-24306 Plön, Germany

Our mathematical model is an application of stochastic evolutionary game dynamics in finite populations. In section 1, we compute the payoffs for each strategy and determine conditions for the stability of responsible sanctions. In section 2 we specify the evolutionary game dynamics which can be interpreted as a social learning process and derive analytical approximations for the case of rare exploration. In the last section, we assess the robustness of our results by analyzing the impact of parameter changes, as well as the consequences of counter-punishment. Additionally, we show that our results are robust with respect to errors in the perception of the co-players' reputation, and to extensions of the strategy space. Moreover, we demonstrate that our qualitative results do not rely on the assumption of pairwise interactions; in fact, our conclusions can be easily transferred to the case of social dilemmas between more than two players.

1 Payoffs

Let us first calculate the payoff for each pairwise interaction. If, for example, an opportunistic O_C-donor encounters a non-punishing N-recipient, then the opportunist knows with probability λ that it is safe to refuse cooperation, leading to zero payoff for both. With probability $\bar{\lambda} = 1 - \lambda$, no such information about the recipient is available and the opportunistic donor will choose his default action, cooperation. In total, this results in an average payoff of $-\bar{\lambda}c$ for the O_C-donor and $\bar{\lambda}b$ for the N-recipient. Repeating this computation for all other strategy pairs yields a bimatrix $(\mathcal{A}, \mathcal{B})$. In this bimatrix, the first entry denotes the payoff of the donor whereas the second entry denotes the corresponding payoff of the recipient:

$$
\begin{array}{c|cccc}
 & \mathbf{R} & \mathbf{N} & \mathbf{A} & \mathbf{S} \\
\hline
\mathbf{C} & \big(-c, b\big) & \big(-c, b\big) & \big(-c-\beta, b-\gamma\big) & \big(-c-\beta, b-\gamma\big) \\
\mathbf{O_C} & \big(-c, b\big) & \big(-\bar{\lambda}c, \bar{\lambda}b\big) & \big(-\bar{\lambda}(c+\beta), \bar{\lambda}(b-\gamma)\big) & \big(-\bar{\lambda}c-\beta, \bar{\lambda}b-\gamma\big) \\
\mathbf{O_D} & \big(-\lambda c-\bar{\lambda}\beta, \lambda b-\bar{\lambda}\gamma\big) & \big(0, 0\big) & \big(0, 0\big) & \big(-\beta, -\gamma\big) \\
\mathbf{D} & \big(-\beta, -\gamma\big) & \big(0, 0\big) & \big(0, 0\big) & \big(-\beta, -\gamma\big),
\end{array}
\tag{2}
$$

where the recipient strategies are responsible sanctioners R, non-punishers N, antisocial punishers A, and spiteful punishers S. On the donor side, the strategies are always cooperate (C), opportunistic cooperation (O_C), opportunistic defection (O_D), and always defect (D).

To study the social learning process, we consider a finite population of size n. Each individual of the population acts according to its strategy $[i, j]$, where $i \in \{C, O_C, O_D, D\}$ is the indivdual's action in the role of a donor and $j \in \{R, N, A, S\}$ describes how to react as a recipient. Thus, we have in total 16 strategies, but we will always consider the two strategies in the different roles separately. We denote the number of $[i, j]$-players with n_{ij}, whereas n_C, n_{O_C}, n_{O_D} and n_D gives the total number of unconditional cooperators, opportunistic cooperators, opportunistic defectors and unconditional defectors, respectively. Similarly, we introduce the variables n_R, n_N, n_A and n_S to denote the total number of responsible sanctioners, non-punishers, antisocial punishers and spiteful punishers, respectively. Because players have an equal chance to be the donor or the recipient in a given interaction, and since self-interactions are excluded, the average payoff of an $[i, j]$-player is given by

$$\pi_{ij} = \frac{1}{n-1}\left(\sum_{k\in\{R,N,A,S\}} \frac{\mathcal{A}_{ik}\cdot n_k}{2} \quad + \sum_{l\in\{C,O_C,O_D,D\}} \frac{\mathcal{B}_{lj}\cdot n_l}{2} \quad - \quad \frac{\mathcal{A}_{ij}+\mathcal{B}_{ij}}{2}\right). \quad (3)$$

We can derive several conclusions from the payoff formula (3):

1. **Componentwise stability.** Because the payoff of strategy $[i,j]$ is a linear combination of the payoff as a donor and the payoff as a recipient, it follows that a homogeneous $[i,j]$-population is evolutionarily stable if and only if it is componentwise stable (that is, neither an $[i,l]$-mutant nor a $[k,j]$-mutant can invade).

2. **Stability of responsible sanctions.** Responsible sanctions can only deter players from non-contributing, if punishment fines are sufficiently high. In fact, in a homogeneous population of cooperative responsible sanctioners, $[C,R]$, a single non-punishing defector has a lower payoff than the residents only if

$$\pi_{DN} - \pi_{CR} = \frac{1}{2}(b-\beta) - \frac{1}{2}\left(\frac{n-2}{n-1}b - c - \frac{\gamma}{n-1}\right) < 0, \quad (4)$$

   This condition is equivalent to

$$\beta > \frac{b+\gamma}{n-1} + c, \quad (5)$$

   which simplifies to $\beta > c$ in the case of large populations. Only if punishment fines are above this threshold, sanctions can potentially stabilize cooperation, and in the following, we will therefore always assume that condition (5) is met.

3. **Conditional behaviour is beneficial.** Opportunism is beneficial in the sense that an opportunistic player never yields a lower payoff than a player with the corresponding unconditional strategy. To see this, consider an arbitrary strategy

$j \in \{R, N, A, S\}$ for the role as a recipient and compute the payoff difference between the unconditional and the respective opportunistic strategy,

$$\pi_{Cj} - \pi_{O_Cj} = \frac{1}{n-1}\left(\sum_{k\in\{R,N,A,S\}} \frac{\mathcal{A}_{Ck} - \mathcal{A}_{O_ck}}{2} \cdot (n_k - \delta_{kj}) \qquad - \qquad \frac{\mathcal{B}_{Cj} - \mathcal{B}_{O_cj}}{2}\right), \tag{6}$$

where δ_{jk} is one if $j = k$ and equal to zero otherwise. Since it follows from payoff table (2) that $\mathcal{A}_{Ck} \leq \mathcal{A}_{O_Ck}$ for all k and that $\mathcal{B}_{Cj} \geq \mathcal{B}_{O_Cj}$ for all j, we may thus conclude that $\pi_{Cj} \leq \pi_{O_Cj}$. A similar computation verifies that opportunistic defectors always get at least the payoff of the respective unconditional strategy, $\pi_{Dj} \leq \pi_{O_Dj}$ for all recipient's actions j. Intuitively, if information about the co-players' previous actions is available, it is always advantageous to consider this information when deciding whether to cooperate or not.

4. **Emergence of cooperation in a population of defectors.** Once the population only consists of non-punishing defectors $[D, N]$, then the three strategies $[D, A]$, $[O_D, N]$ and $[O_D, A]$ can invade through neutral drift for all parameter combinations. If punishment is sufficiently costly, $\gamma/\beta > 1/(n-1)$, however, there is no other strategy $[i, j]$ that can invade a homogeneous $[D, N]$-population. Indeed, if we compute the payoff of a single $[i, j]$-invader, then we find that

$$\pi_{ij} - \pi_{DN} = \begin{cases} 0 & \text{if } \ i \in \{D, O_D\} \ \text{ and } \ j \in \{N, A\} \\ -\gamma + \beta/(n-1) & \text{if } \ i \in \{D, O_D\} \ \text{ and } \ j \in \{R, S\} \\ -c - b/(n-1) & \text{if } \ i = C \ \text{ and } \ j = N \\ -\bar{\lambda}c - \bar{\lambda}b/(n-1) & \text{if } \ i = O_C \ \text{ and } \ j = N \end{cases} \tag{7}$$

A similar argument holds for antisocial-punishing defectors $[D, A]$, which can only be invaded through neutral drift by $[D, N]$, $[O_D, N]$ and $[O_D, A]$. However, once the whole population uses an opportunistic strategy, $[O_D, N]$ or $[O_D, A]$, the use of responsible sanctions becomes beneficial, provided that the reputation level λ is sufficiently high: Indeed, if threshold (1) from the main text is met, that is if

$$\lambda > \frac{(n-1)\gamma - \beta}{(n-1)(\gamma + b) + c - \beta}, \tag{8}$$

then a single $[O_D, R]$-invader has always a higher payoff than the residents. The strategy $[O_D, R]$, in turn, can easily be invaded by the more cooperative strategies $[C, R]$ and $[O_C, R]$.

5. **Breakdown of cooperation.** For moderate punishment fines, i.e. $(b+\gamma)/(n-1)+c < \beta < \gamma(n-1)$, a homogeneous $[O_C, R]$-population can only be invaded through neutral drift by $[C, R]$. Once a homogeneous $[C, R]$-population is reached, neutral drift may either lead back to $[O_C, R]$, or it may lead to a non-punishing $[C, N]$-population. A $[C, N]$-population, in turn, is highly unstable as it can be invaded by all other non-punishing strategies, $[O_C, N]$, $[O_D, N]$ and $[D, N]$. Overall, cooperation is thus most stable in a population of opportunistic social sanctioners: When the whole population makes use of $[O_C, R]$, it takes two neutral transitions (from $[O_C, R]$ to $[C, R]$ and from there to $[C, N]$) to reach a state that is susceptible for invasion by defectors. Thus, the evolution of cooperation in our model is more likely than the breakdown of cooperation, leading to a mutation-selection equilibrium that favours cooperation.

2 Evolutionary dynamics

To model the dynamics of strategy adaptation in the population, we apply a pairwise comparison process[1,2,3], which is closely related to the frequency-dependent Moran process[4]. That is, we assume that in each time-step, subjects interact with all other members of the population, such that their payoffs are given by Eq. (3). Thereafter, one individual is randomly selected to imitate the strategy of a peer, whereby strategies of peers with high payoffs are more likely to be adopted. In particular, if a focal individual with strategy $[i, j]$ selects a role model with strategy $[k, l]$, then the probability of adopting the role model's strategy is given by the so-called Fermi-rule [1,2,3]:

$$p_{[i,j]\to[k,l]} = \frac{1}{1 + \exp\left[-s(\pi_{kl} - \pi_{ij})\right]} \tag{9}$$

The parameter $s \geq 0$ denotes the imitation strength: For small s, a coin toss essentially decides whether or not to imitate the role model. In the other limit $s \to \infty$, the focal player only imitates co-players that have a higher payoff. These two limits are usually referred to as the case of weak and of strong selection, respectively. Additionally, we allow for random exploration of strategies. In each time step, the focal individual switches to another random strategy with probability $\mu > 0$. Each of the other 15 strategies has an equal chance to be selected.

For the simulations, we focus on two exploration scenarios: In the case of frequent exploration, the exploration rate is set to $\mu = 0.1$. For sufficiently large populations, frequent exploration thus implies that typically all 16 strategies are present in the population. In the other case of rare exploration, we used an exploration rate of $\mu = 0.0001$. Since there are no stable coexistences, this choice implies that a sufficiently small population is typically in a monomorphic state [5].

2.1 Analytical approximations for the case of rare exploration

In finite populations with small exploration rates, the population spends almost all of its time in a homogeneous state. When one player mutates to a different strategy, then this newly introduced strategy either dies out or goes to fixation before the next mutation occurs [5]. We can therefore assemble a transition matrix between homogeneous states of the system. The transition probability from state $[i, j]$ to state $[k, l]$ is the product of the probability $\mu/15$ of a mutant type $[k, l]$ arising and the probability $\rho_{ij,kl}$ that this mutant reaches fixation[6]. The fixation probability can be calculated for any birth death process and for any intensity of selection[7]; in the case of updating rule (9) it is given by[3]

$$\rho_{ij,kl} = \frac{1}{1 + \sum_{m=1}^{n-1} \prod_{n_{kl}=1}^{m} \exp\big(- s(\pi_{kl} - \pi_{ij})\big)} \tag{10}$$

The 16×16 transition matrix that describes the probabilities to move from one homogeneous population to another is thus defined as

$$\begin{pmatrix} 1 - \sum_{k,l} \frac{\mu\, \rho_{CR,kl}}{15} & \frac{\mu\, \rho_{CR,CN}}{15} & \frac{\mu\, \rho_{CR,CA}}{15} & \cdots & \frac{\mu\, \rho_{CR,DS}}{15} \\ \frac{\mu\, \rho_{CN,CR}}{15} & 1 - \sum_{k,l} \frac{\mu\, \rho_{CN,kl}}{15} & \frac{\mu\, \rho_{CN,CA}}{15} & \cdots & \frac{\mu\, \rho_{CN,DS}}{15} \\ \frac{\mu\, \rho_{CA,CR}}{15} & \frac{\mu\, \rho_{CA,CN}}{15} & 1 - \sum_{k,l} \frac{\mu\, \rho_{CA,kl}}{15} & \cdots & \frac{\mu\, \rho_{CA,DS}}{15} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \frac{\mu\, \rho_{DS,CR}}{15} & \frac{\mu\, \rho_{DS,CN}}{15} & \frac{\mu\, \rho_{DS,CA}}{15} & \cdots & 1 - \sum_{k,l} \frac{\mu\, \rho_{DS,kl}}{15} \end{pmatrix} \tag{11}$$

From this transition matrix, the steady state distribution $x = (x_{CR}, \ldots, x_{DS})$ of the stochastic process can be calculated by solving the corresponding eigenvector problem. Note that the exploration rate μ drops out in this calculation. Thus, we consider in the following the transition matrix T given that an exploration step occurred. This matrix follows from Eq. 11 simply from dropping the exploration parameter μ, the steady state is the solution of $xT = x$. The entries x_{ij} of the steady state distribution may be interpreted as the frequency of finding the population in state $[i, j]$ after a sufficiently long time. Since for non-weak selection, the transition probabilities $\rho_{ij,kl}/15$ involve the payoffs in a highly non-linear way, we have calculated the steady state distribution x numerically (which can be done with arbitrarily high precision).

	CR	CN	CA	CS	O_CR	O_CN	O_CA	O_CS	O_DR	O_DN	O_DA	O_DS	DR	DN	DA	DS
CR	$\frac{15n-2}{15n}$	$\frac{1}{15n}$	0	0	$\frac{1}{15n}$	0	0	0	0	0	0	0	0	0	0	0
CN	$\frac{1}{15n}$	$\frac{7n-1}{15n}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
CA	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{n-1}{15n}$	$\frac{1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
CS	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$	$\frac{n-1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
O_CR	$\frac{1}{15n}$	0	0	0	$\frac{15n-1}{15n}$	0	0	0	0	0	0	0	0	0	0	0
O_CN	$\frac{1}{15}$	0	0	0	$\frac{1}{15}$	$\frac{8}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0
O_CA	$\frac{1}{15}$	0	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
O_CS	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
O_DR	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	0	0	0	$\frac{10}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	0	0	0	0	0
O_DN	0	0	0	0	0	0	0	0	0	$\frac{5n-1}{5n}$	$\frac{1}{15n}$	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0
O_DA	0	0	0	0	0	0	0	0	0	$\frac{1}{15n}$	$\frac{5n-1}{5n}$	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0
O_DS	0	0	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{7n-1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$
DR	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	0	$\frac{6n-1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$
DN	0	0	0	0	0	0	0	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0	0	$\frac{5n-1}{5n}$	$\frac{1}{15n}$	0
DA	0	0	0	0	0	0	0	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0	0	$\frac{1}{15n}$	$\frac{5n-1}{5n}$	0
DS	0	0	0	0	0	$\frac{1}{15}$	0	0	$\frac{1}{30}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$	$\frac{1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{19n-4}{30n}$

Table 1: Transition matrix $T = (\tau_{ij})$ for strong selection in the case that the information level λ is close to, but below the critical threshold given by Eq. (1). The entries τ_{ij} give the probability that a mutant with strategy j occurs and reaches fixation in a homogeneous population with strategy i. Note that once the population is in one of the four states $[O_D, N]$, $[O_D, A]$, $[D, N]$ or $[D, A]$, there is no other strategy that could take over.

2.2 Exact results for the limit of strong selection

While the previous section allows a numerical calculation of the fixation probabilities for any strength of selection, the fixation probabilities take a particularly simple form when selection is strong, that is when $s \to \infty$. In this case, the fixation probabilities $\rho_{ij,kl}$ are given by 0, $1/n$, or 1, depending on whether mutants have a lower, equal, or higher payoff than the residents, respectively.

Table 1 gives the transition matrix for the case of a low information level and moderate punishment fines β, that is, λ does not fulfill condition (1) from the main text and $(b+\gamma)/(n-1)+c < \beta < \gamma(n-1)$. As can be seen, the four non-cooperative states $[O_D, N]$, $[O_D, A]$, $[D, N]$ or $[D, A]$ (marked in blue) form an evolutionary trap, in the sense that once one of these four states is reached, there is no other strategy that is able to take over.

	CR	CN	CA	CS	O_CR	O_CN	O_CA	O_CS	O_DR	O_DN	O_DA	O_DS	DR	DN	DA	DS
CR	$\frac{15n-2}{15n}$	$\frac{1}{15n}$	0	0	$\frac{1}{15n}$	0	0	0	0	0	0	0	0	0	0	0
CN	$\frac{1}{15n}$	$\frac{7n-1}{15n}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
CA	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{n-1}{15n}$	$\frac{1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
CS	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$	$\frac{n-1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
O_CR	$\frac{1}{15n}$	0	0	0	$\frac{15n-1}{15n}$	0	0	0	0	0	0	0	0	0	0	0
O_CN	$\frac{1}{15}$	0	0	0	$\frac{1}{15}$	$\frac{8}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0
O_CA	$\frac{1}{15}$	0	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{4}{15}$	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
O_CS	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{2}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$
O_DR	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	0	0	0	$\frac{12}{15}$	0	0	0	0	0	0	0
O_DN	0	0	0	0	0	0	0	0	$\mathbf{\frac{1}{15}}$	$\frac{14n-3}{15n}$	$\frac{1}{15n}$	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0
O_DA	0	0	0	0	0	0	0	0	$\mathbf{\frac{1}{15}}$	$\frac{1}{15n}$	$\frac{14n-3}{15n}$	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0
O_DS	0	0	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{7n-1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$
DR	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	0	0	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15}$	0	$\frac{6n-1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$
DN	0	0	0	0	0	0	0	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0	0	$\frac{5n-1}{5n}$	$\frac{1}{15n}$	0
DA	0	0	0	0	0	0	0	0	0	$\frac{1}{15n}$	$\frac{1}{15n}$	0	0	$\frac{1}{15n}$	$\frac{5n-1}{5n}$	0
DS	0	0	0	0	0	$\frac{1}{15}$	0	0	$\frac{1}{30}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{1}{15n}$	$\frac{1}{15n}$	$\frac{1}{15}$	$\frac{1}{15}$	$\frac{19n-4}{30n}$

Table 2: Transition matrix T for strong selection in the case that the information level λ is close to, but above the critical threshold $\hat{\lambda}$ given by Eq. (1). The four states $[O_D, N]$, $[O_D, A]$, $[D, N]$ or $[D, A]$ are no longer absorbing; instead, responsible sanctioners can invade via the two opportunistic states $[O_D, N]$ and $[O_D, A]$. These transitions are marked in red.

Indeed, by the previous section, any different mutant strategy obtains a payoff that is lower than resident's payoff, and thus, by the assumption of strong selection, this mutant strategy goes extinct. As a consequence, in the steady state of the evolutionary process, only these four non-cooperative strategies are played with positive probability. In fact, the invariant distribution x fulfills $x_{O_DN} = x_{O_DA} = x_{DN} = x_{DA} = 1/4$, whereas $x_{ij} = 0$ for all other strategies $[i, j]$.

The steady state of the evolutionary process changes drastically when the information level exceeds the critical threshold given by Eq. 1 in the main text. In this case, as shown in Table 2, reputation opens an exit path that leads out of the non-cooperative trap formed by the four strategies $[O_D, N]$, $[O_D, A]$, $[D, N]$ and $[D, A]$. As the population reaches one of the opportunistic states, $[O_D, N]$ or $[O_D, A]$, responsible sanctioners $[O_D, R]$ can easily invade (marked in red) and take over. Once the population has moved to the state

$[O_D, R]$, however, more cooperative strategies such as $[C, R]$ and $[O_C, R]$ become beneficial. Especially the opportunistic state $[O_C, R]$ is relatively stable, as only its unconditional counterpart $[C, R]$ can invade through neutral drift. These findings are also reflected in the steady state of the process: For example, for a population size $n = 80$, the steady state x for the transition matrix in Table 2 fulfills $x_{CR} \approx 0.35$ and $x_{O_C R} \approx 0.55$, whereas the population is in a non-cooperative state in only one of ten cases.

Therefore, once the information level exceeds the critical threshold, the population moves from a fully non-cooperative regime to a highly cooperative state, which is stabilized by responsible sanctions. While these result were derived in the limit of strong selection and small exploration rates, simulations (Figs. 1–3) illustrate that also for finite selection pressure and higher exploration rates, the threshold Eq. 1 is a reasonable approximation for the critical information level that needs to be met for cooperation to evolve.

3 Robustness

3.1 Robustness of the results with respect to parameter changes

The impact of the game parameters b, c, γ, β, as well as the impact of population size n can be investigated by analyzing the parameters' influence on the critical information threshold $\left((n-1)\gamma - \beta\right) / \left((n-1)(\gamma + b) + c - \beta\right)$. For example, a simple calculation verifies that this threshold is strictly increasing in population size n. Thus, cooperation requires higher information levels in large populations. However, even in infinitely large populations, the critical information level never exceeds $\gamma/(b + \gamma)$. As a consequence, cooperation is particularly likely to evolve if the benefit of cooperation b is sufficiently high compared to the costs of sanctions γ. Intuitively, the higher the benefit b, the more it pays off to invest an amount γ in order to gain a strict reputation that helps to ensure future cooperation.

Punishment fines β have, especially in large populations, a negligible impact on the critical information threshold. However, if punishment fines are too low, $\beta < (b+\gamma)/(n-1)+c$, then sanctions do not act as a deterrent and unconditional defection dominates all other donor's strategies. On the other hand, if fines are too high and $\beta > \gamma(n-1)$, then spite, instead of responsible sanctions, evolves. Therefore, spite requires small population sizes n and cheap punishment γ in order to emerge. Furthermore, a straightforward calculation shows that a spiteful mutant can only invade a homogeneous $[O_C, R]$-population if

$$\lambda < \frac{\beta - (n-1)\gamma}{(n-1)b + c + \beta - (n-1)\gamma}.$$

Thus, in opportunistic populations, spite additionally requires a high degree of anonymity.

In order to investigate how the strength of selection affects the resulting dynamics, Figure S1 shows the steady state distribution as a function of the selection parameter s. We can roughly distinguish between two different scenarios:

1. **Strong selection.** If selection is sufficiently strong ($s \gg 0.1$), we find that recipients mostly rely on responsible sanctions. Donors, on the other hand, mostly cooperate, with a notable trend towards opportunistic cooperation (which is especially pronounced under frequent exploration).

2. **Weak selection.** If selection is weak ($s \ll 0.1$) and game payoffs play a subordinate role on the strategies that are played, cooperation clearly falls behind. This happens due to a representation effect: In the case of weak selection, all strategies are played with almost equal shares. However, only one out of four of the recipients' strategies supports cooperation (namely R), whereas the other three actions N, A and S implicitly promote defection. Thus the choice of the strategy space, together with the assumption of weak selection, leads to a bias towards less cooperation.

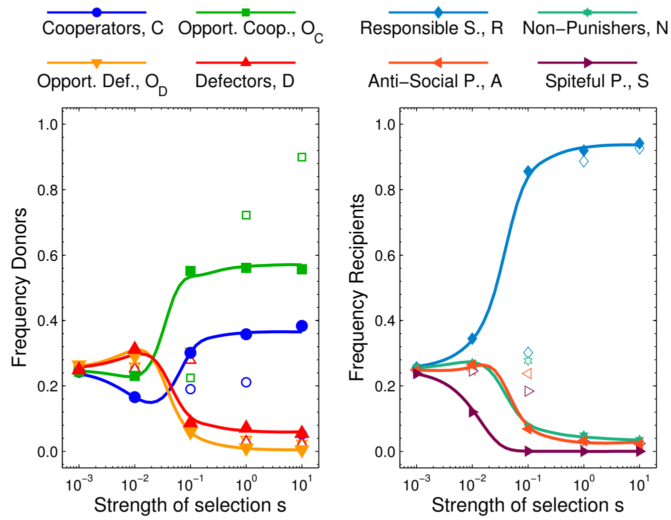

Figure S1: Impact of selection strength on the co-evolution of cooperation and responsible sanctions. The graph shows the steady-state frequencies for the strategies of donors (left graph) and recipients (right graph), respectively. Solid lines indicate exact results for the limiting case of rare exploration, whereas dots represent simulation results for exploration rates $\mu = 0.0001$ and $\mu = 0.1$. There are two major regimes: Cooperative strategies are clearly predominant under strong selection, whereas for weak selection, there is a slight bias towards defection. As in the previous figures, parameter values are $n = 80$, $b = 4$, $\beta = 3$, $c = \gamma = 1$ and $\lambda = 30\%$. Simulations were run over a period of 10^{10} time steps starting from a single random initial condition (i.e., each individual was allowed to implement more than 10^8 strategy changes.)

3.2 The effect of counter-punishment

Several studies suggest that subjects may use punishment for retaliation, i.e. as a response to being punished previously[8,9]. Obviously, such retaliatory punishment threatens the co-evolution of responsible sanctions and cooperation, because it increases the costs of punishment and thus may prevent social sanctioners from punishing defectors. To investigate the effect of counter-punishment, we have thus considered a scenario where the costs of punishment are as high as the costs of being punished, that is, we have considered a scenario where $\gamma = \beta$. This can be interpreted as a situation in which counter-punishment is a sure event.

Surprisingly, we find that while counter-punishment prevents the evolution of spite, it still allows for the evolution of responsible sanctions. Indeed, as Figure S2 shows,

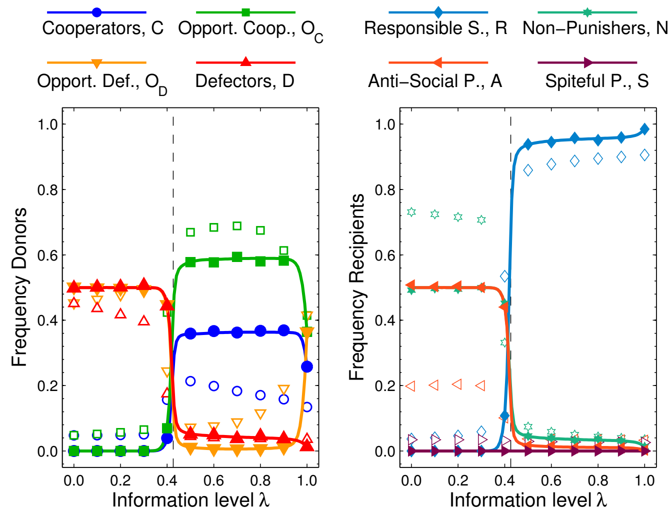

Figure S2: Counter-punishment and the evolution of responsible sanctions. The graph corresponds to Figure 1 for the case that punishment is equally costly for the punisher and for the target of punishment. The possibility of counter-punishment has increased the critical information-threshold from roughly 20% (as in Figure 1) to 40%. However, above this threshold, we still find that cooperation and responsible sanctions co-evolve. Parameter values are $n = 80$, $b = 4$, $\beta = \gamma = 3$, $c = 1$ and $s = 0.5$.

counter-punishment leads to an increase of the critical information threshold. However, above this threshold, subjects still learn to behave opportunistically, and to use sanctions against non-cooperators. In contrast, spite does not evolve for any parameter values, as the necessary condition for spite, $\beta > (n-1)\gamma$, is no longer feasible. Intuitively, spiteful punishment can only prevail if it leads to a relative payoff advantage for the punisher. However, if counter-punishment is a sure event, then sanctions are equally costly for both parties and thus spite cannot gain a foothold in the population.

3.3 Extension of the strategy space

So far we have only considered a restricted strategy space; donors could either optimally adapt to the co-player's reputation (O_C and O_D), or they could not react on the co-player's reputation at all (C and D). This approach entails the risk of leaving out other relevant strategies[10]. It is thus the aim of this section to show that our results can be transferred

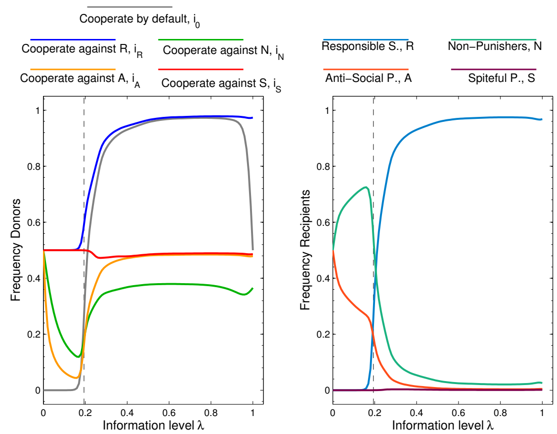

Figure S3: Impact of reputation in case of the full strategy set. The graph shows the exact value of the invariant distribution in the limit of rare exploration. As before, responsible sanctioners take over if λ exceeds the critical information threshold, in which case donors learn to cooperate by default. Parameters were set to the corresponding values in Figure 1: $b = 4$, $\beta = 3$, $c = \gamma = 1$, population size $n = 80$, strength of selection $s = 0.5$.

to the case where donors can adopt any possible strategy they want.

In order to model the full strategy space, we encode the strategy of a player as a 6-tuple $(i_0, i_R, i_N, i_A, i_S; j)$. Here, the first five variables refer to the player's strategy in the role of the donor: $i_0 \in \{C, D\}$ gives the player's action if the co-player's reputation is unknown. The other four variables i_R, i_N, i_A, i_S correspond to the player's action if the co-player is known as a responsible sanctioner, a non-punisher, an antisocial punisher, or a spiteful punisher, respectively. The last variable $j \in \{R, N, A, S\}$ encodes the player's strategy in the role of the recipient. Therefore, there are $2^5 \cdot 4 = 128$ different strategies, including the previous 16 strategies (for example, $O_C R$ is $[CCDDD; R]$).

By calculating the invariant distribution in the limit of rare exploration, we confirm that the qualitative features of the dynamics are unchanged (Fig. S3): As λ exceeds the critical threshold, recipients use responsible sanctions to deter co-players from defection. This means of deterrence, in turn, proves successful: Above the critical information threshold, almost all players cooperate if the other's reputation is unknown, or if the co-player is known to be a responsible punisher. The propensity to cooperate against other recipients is considerably lower, but due to neutral drift still relatively high (between 40-50 %). On the level of individual strategies, DN (i.e. $[D, D, D, D, D; N]$) and $O_D N$ (i.e. $[D, C, D, D, D; N]$) are the most abundant strategies for $\lambda < 20\%$. Note, however, that the corresponding strategies that cooperate against spiteful punishers (i.e. $[D, D, D, D, C; N]$ and $[D, C, D, D, C; N]$) are equally abundant, because the evolutionary process does not give rise to spiteful individuals. For $\lambda > 25\%$, the strategy $O_C R$ is most abundant, together with the corresponding strategy that cooperates against spiteful subjects, $[C, C, D, D, C; R]$.

3.4 Errors in perception

In the previous analysis we have relied on the assumption that a player's knowledge about the co-player's reputation is always correct, while everyday experience suggests that information gained from gossip or other sources may be error-prone. Such errors in the perception of the co-player's reputation have two effects: First, opportunistic donors bear the risk of choosing a wrong best reply to the co-player's strategy; and second, perception errors diminish the incentive for recipients to use responsible sanctions as a signal to bystanders. Both effects endanger the co-evolution of cooperation and responsible sanctions, thereby calling the robustness of our results into question.

Let us therefore assume that a player's commonly known reputation is wrong with probability ε. That is, with probability ε, a recipient with punishment strategy $i \in \{R, N, A, S\}$ is perceived as a player with strategy $j \neq i$ (where all $j \neq i$ have equal probability to be the recipient's wrong reputation). In a given game there are therefore three possible scenarios:

1. With probability $1 - \lambda$ the donor does not know the recipient's strategy, in which case donors use their default strategy.

2. With probability $\lambda(1 - \varepsilon)$, the donor knows the recipient's true strategy and oppor-

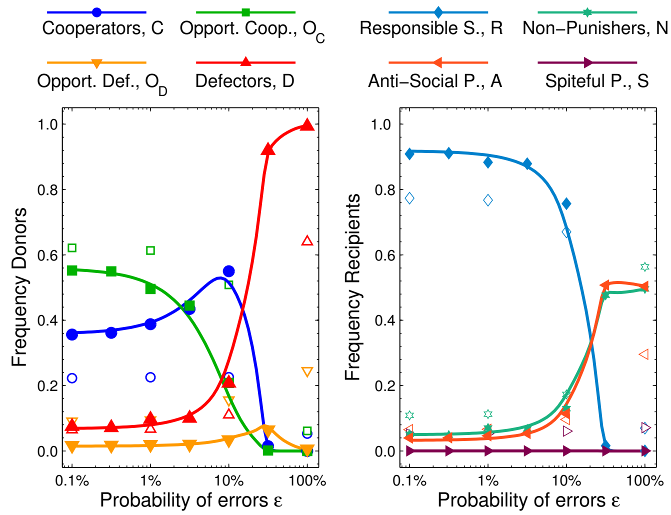

Figure S4: Impact of perception errors on the co-evolution of responsible sanctions and cooperation. As long as perception errors are sufficiently rare ($\varepsilon < 10\%$), responsible punishment is predominant among the recipients, thereby allowing the evolution of cooperation among the donors. Only if perception errors increase further, defectors take over. Parameter values are $b = 4$, $\beta = 3$, $c = \gamma = 1$, population size $n = 80$, strength of selection $s = 0.5$, and $\lambda = 30\%$.

tunistic donors adapt their action accordingly.

3. With the remaining probability $\lambda\varepsilon$ a recipient's publicly known reputation is wrong, which may lead opportunists to use an inappropriate strategy.

Figure S4 illustrates the consequences of perception errors on the stability of cooperation and responsible sanctions. As one may expect, frequent perception errors make a player's reputation an incredible signal; therefore unconditional defection, combined with no punishment (or anti-social punishment), evolves for extremely high values of ε. But reputation allows the evolution of responsible sanctions not only without errors, but also for moderate error rates such as $\epsilon = 10\%$. In this case, however, a majority of donors cooperates unconditionally, rather than as an opportunistic response to the recipient's reputation. As the error rate decreases, opportunistic cooperators take over. Note that for Figure S4, the information level λ was set relatively low ($\lambda = 30\%$), and that higher information levels have a positive influence on achieved cooperation.

3.5 Games in groups

While our previous analysis has assumed pairwise games, many real world social dilemmas, such as the management of common resources[11], take place in groups of more than two individuals. It is therefore the aim of this section to extend our results to the more general case of games between $m > 2$ players. To this end, we study the co-evolution of cooperation and punishment in public good games (PGG), the most commonly applied metaphor for social dilemmas among groups of individuals.

Let us therefore consider the following standard scheme for PGG: A group of m individuals must decide whether to make a contribution c to a public pool, knowing that this public pool leads to a return rc per contribution, which is divided equally among all subjects of the group. As we assume that $1 < r < m$, the social optimum is attained if all subjects contribute, while the individual optimum is to withhold all contributions. After observing the others' contributions, individuals are allowed to punish others based on the co-players' contribution behaviour. Punishment leads to a cost β for the punished, and to a cost γ for the punisher. We assume that the relation $\beta > c$ holds, which ensures that it becomes beneficial to cooperate if threatened by punishment.

As before, the players can choose among four possible strategies in the punishment stage: They can punish all defectors (R), all cooperators (A), everyone (S) or no one (N). For the contribution stage, we assume that with probability λ, individuals can correctly anticipate the punishment behaviour of all their co-players. Cooperators (C) always contribute to the public pool, whereas defectors (D) never contribute. Opportunistic cooperators (O_C) contribute, unless they know that it is beneficial to defect (which is the case if they know that the number of social sanctioners R in the group is below or equal to the number of antisocial punishers A). Similarly, opportunistic defectors (O_D) usually withhold contributions, unless they know that the number of social sanctioners R in the group exceeds the number of antisocial punishers A.

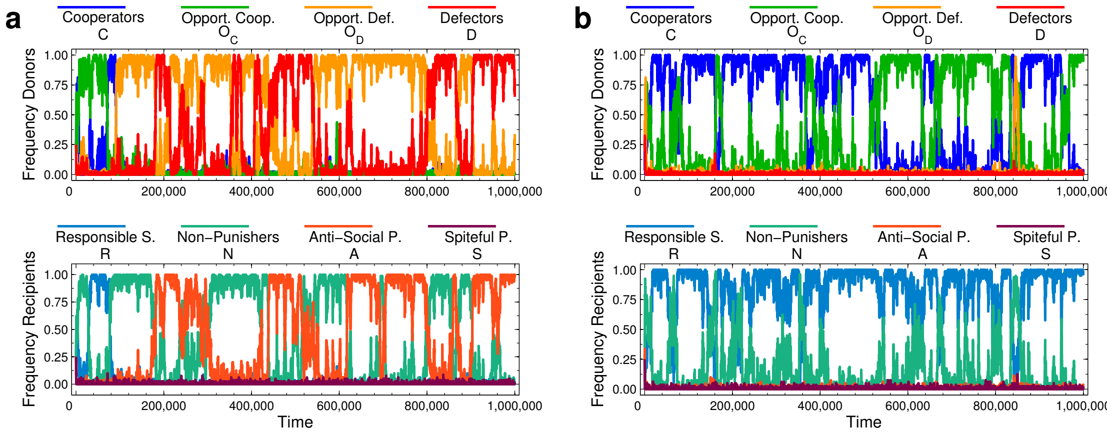

Figure S5: Evolutionary dynamics of cooperation and punishment in public good games. (a) If the probability to know all other group members' reputation is below the critical information threshold (12), then neither responsible sanctions nor cooperation can evolve. (b) However, above this threshold, individuals make use of responsible sanctions, which in turn promotes cooperation. Both figures show the first 1,000,000 iterations of a typical simulation run for (a) $\lambda = 0.3$ and (b) $\lambda = 0.6$. The other parameters were set to: Population size $n = 80$, group size $m = 5$, contribution costs $c = 1$, punishment fine $\beta = 3/2$, punishment costs $\gamma = 1/2$, multiplication factor $r = 3$, strength of selection $s = 0.5$ and exploration rate $\mu = 0.01$. Note that for these parameter values, the critical information threshold (12) becomes $\lambda > 5/11 \approx 0.45$.

The evolutionary dynamics of the system is modeled as in the previous case of two-player interactions: We consider a population of n players. Individuals are then randomly assigned to groups of m players who interact in the previously described PGG. Given the state of the current population, this allows us to compute the expected payoff π_{ij} for each of the 16 possible strategies, with $i \in \{C, O_C, O_D, D\}$ and $j \in \{R, N, A, S\}$. After these interactions, one randomly chosen player is given the opportunity to update the strategy by comparing the own payoff with a random co-player's payoff. The updating probability to switch to the role model's strategy is again specified by the Fermi rule (9). For the simulations, we estimated the expected payoff π_{ij} of a strategy $[i, j]$ by considering 100 randomly chosen groups containing an $[i, j]$-player.

Analogously to condition (1) in the main text, we can calculate a critical information threshold for λ that needs to be met for responsible sanctions to originate in a popu-

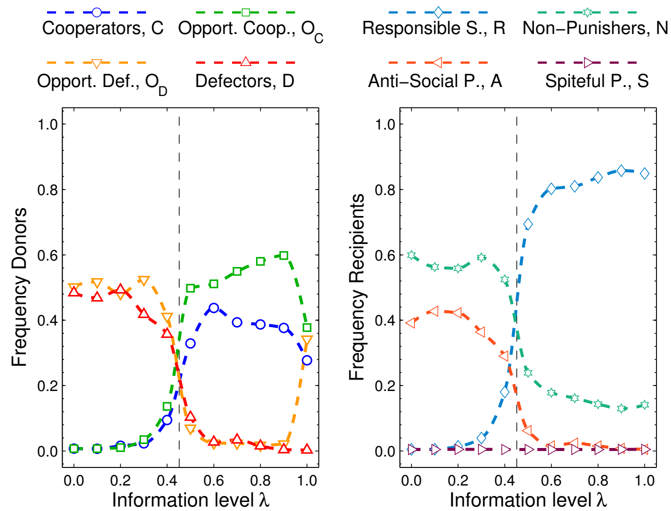

Figure S6: The impact of the information level on the co-evolution of cooperation and responsible punishment in public good games. The graph shows time-averaged frequencies for the strategies in the contribution stage (left graph) and in the punishment stage (right graph), respectively. Open symbols represent simulation results for intermediate exploration rates ($\mu = 0.01$), the colored dashed lines serve as a guide to the eye. The black dashed line represents the critical information level given by Eq. (12). Above this information level, a substantial part of the individuals makes use of responsible sanctions to deter opportunists from defection. Parameter values were chosen as in Figure S5. Simulations were run over a period of 10^7 time steps.

lation of non-punishing opportunists $[O_D, N]$. The critical information threshold takes a particularly simple form for large population sizes n: In this case, the payoff of the resident $[O_D, N]$ population is zero, whereas a single $[O_D, R]$-invader yields on average $\pi_{ij} = -(1-\lambda)(m-1)\cdot\gamma + \lambda(m-1)\cdot rc/m$. Thus, the condition for the emergence of responsible sanctions is given by

$$\lambda > \frac{m\gamma}{rc + m\gamma}. \tag{12}$$

The predictive value of this information threshold is confirmed by simulations (see Figures S5 and S6): If individuals have no sufficient opportunity to build up a strict reputation, then the population is dominated by non-cooperating strategies. However, as λ exceeds the critical threshold, responsible punishment is clearly the most abundant strategy in the punishment stage, which in turn allows cooperative strategies to evolve in the contribution stage. Interestingly, threshold (12) is formally similar to the corresponding threshold in the case of two-player interactions $\lambda > \gamma/(b+\gamma)$. However, it is noteworthy that for PGG, the critical information threshold (12) increases with group size m, while it is realistic to assume that the probability to know the co-players' punishment reputation λ is a decreasing function of group size m. Thus, there is a critical group size m^* such that groups of smaller size are able to establish a cooperative regime, whereas bigger groups fail to maintain cooperation. This observation suggests that peer punishment is a very effective mechanism in relatively small groups, while it may fail in larger collective actions. This might explain why large societies rather rely on centralized punishment institutions than on self-governance[12].

References

[1] Blume, L. E. The statistical mechanics of strategic interaction. *Games and Economic Behavior* **5**, 387–424 (1993).

[2] Szabó, G. & Tőke, C. Evolutionary Prisoner's Dilemma game on a square lattice. *Phys. Rev. E* **58**, 69 (1998).

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

[4] Nowak, M. A., Sasaki, A., Taylor, C. & Fudenberg, D. Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004).

[5] Wu, B., Gokhale, C., Wang, L. & Traulsen, A. How small are small mutation rates? *Journal of Mathematical Biology* 1–25 (2011).

[6] Fudenberg, D. & Imhof, L. A. Imitation process with small mutations. *J. Econ. Theory* **131**, 251–262 (2006).

[7] Nowak, M. A. *Evolutionary Dynamics* (Harvard University Press, Cambridge, MA, 2006).

[8] Herrmann, B., Thöni, C. & Gächter, S. Antisocial punishment across societies. *Science* **319**, 1362–1367 (2008).

[9] Nikiforakis, N. Punishment and counter-punishment in public good games: Can we really govern ourselves? *Journal of Public Economics* **92**, 91–112 (2008).

[10] Rand, D. G. & Nowak, M. A. The evolution of antisocial punishment in optional public goods games. *Nature Communications* **2**, http://dx.doi.org/10.1038/ncomms1442 (2011).

[11] Hardin, G. The tragedy of the commons. *Science* **162**, 1243–1248 (1968).

[12] Sigmund, K., De Silva, H., Traulsen, A. & Hauert, C. Social learning promotes institutions for governing the commons. *Nature* **466**, 861–863 (2010).