Adaptive Dynamics of Extortion and Compliance

Christian Hilbe[1]*, Martin A. Nowak[2], Arne Traulsen[1]

1 Evolutionary Theory Group, Max-Planck Institute for Evolutionary Biology, Plön, Germany, 2 Program for Evolutionary Dynamics, Harvard University, Cambridge, Massachusetts, United States of America

Abstract

Direct reciprocity is a mechanism for the evolution of cooperation. For the iterated prisoner's dilemma, a new class of strategies has recently been described, the so-called zero-determinant strategies. Using such a strategy, a player can unilaterally enforce a linear relationship between his own payoff and the co-player's payoff. In particular the player may act in such a way that it becomes optimal for the co-player to cooperate unconditionally. In this way, a player can manipulate and extort his co-player, thereby ensuring that the own payoff never falls below the co-player's payoff. However, using a compliant strategy instead, a player can also ensure that his own payoff never exceeds the co-player's payoff. Here, we use adaptive dynamics to study when evolution leads to extortion and when it leads to compliance. We find a remarkable cyclic dynamics: in sufficiently large populations, extortioners play a transient role, helping the population to move from selfish strategies to compliance. Compliant strategies, however, can be subverted by altruists, which in turn give rise to selfish strategies. Whether cooperative strategies are favored in the long run critically depends on the size of the population; we show that cooperation is most abundant in large populations, in which case average payoffs approach the social optimum. Our results are not restricted to the case of the prisoners dilemma, but can be extended to other social dilemmas, such as the snowdrift game. Iterated social dilemmas in large populations do not lead to the evolution of strategies that aim to dominate their co-player. Instead, generosity succeeds.

Introduction

Repeated games are among the best-studied objects in game theory, and the iterated prisoner's dilemma has stimulated research on the evolution of cooperation for more than five decades [1–5]. The prisoner's dilemma describes a social dilemma between two players, each having the choice whether to cooperate or to defect. When both cooperate, they each receive a mutual reward R, which exceeds their payoff for mutual defection, P. But if one player cooperates and the other defects, then the defector gets the highest payoff T, whereas the cooperator ends up with the lowest payoff S. Thus, if the game is played only once (or for a known finite number of rounds), then mutual defection is the only equilibrium. However, when players cannot anticipate how often the game will be played, cooperative solutions become feasible [3,5,6].

Researchers from diverse disciplines have used the iterated prisoner's dilemma to discuss the potential of direct reciprocity for the evolution of cooperation [7–19]. However, recently Press and Dyson [20] discovered that the infinitely repeated prisoner's dilemma also contains strategies that allow the manipulation and extortion of opponents [21–25]. To show this, they first proved that there are simple strategies, which only depend on the outcome of the previous round, such that each side can enforce a linear relationship between the payoffs of the two players. More precisely, suppose player 1 applies a memory-one strategy $p=(p_R, p_S, p_T, p_P)$, where p_i is the probability to cooperate after yielding a payoff $i \in \{R, S, T, P\}$ in the previous round (additionally, such a strategy needs to specify a move for the first round.

However, for infinitely iterated games, the first round can often be neglected). Moreover, assume that there are three constants α, β, γ such that p can be written as

$$\begin{aligned}
p_R &= \alpha R + \beta R + \gamma + 1 \\
p_S &= \alpha S + \beta T + \gamma + 1 \\
p_T &= \alpha T + \beta S + \gamma \\
p_P &= \alpha P + \beta P + \gamma.
\end{aligned} \qquad (1)$$

Press and Dyson [20] showed that when player 1 applies such a strategy against an opponent with arbitrary strategy q, then the player's payoff $A(p,q)$ and the opponent's payoff $A(q,p)$ fulfill the linear relation

$$\alpha A(p,q) + \beta A(q,p) + \gamma = 0. \qquad (2)$$

Since their proof required certain determinants to vanish, Press and Dyson called such strategies p *zero-determinant strategies*. At first sight, zero-determinant strategies might seem as a mere mathematical curiosity [26]. However, their existence has several surprising consequences. Press and Dyson [20] discovered that certain zero-determinant strategies can guarantee that a player always yields at least the opponent's payoff. They showed that by setting $\gamma = -(\alpha + \beta)P$, a zero-determinant strategist can enforce the relation

$$A(p,q) - P = \chi(A(q,p) - P), \qquad (3)$$

where $\chi = -\beta/\alpha \geq 1$ is called the extortion factor [20,23]. Such *extortioner* strategies p guarantee that the player's own surplus (over the maximin value P) exceed's the co-player's surplus by a fixed percentage. In particular, when the the typical payoff relations $P < (T+S)/2 < R$ hold, the payoff of an extortioner is never below the payoff of its co-player, suggesting that extortioners would dominate any evolutionary opponent [20].

On the other hand, Stewart and Plotkin [21,25] considered a generous counterpart to extortioners. Starting from $\gamma = -(\alpha + \beta)R$, they investigated zero-determinant strategists that enforce the relation

$$R - A(p,q) = \chi(R - A(q,p)), \qquad (4)$$

where again $\chi = -\beta/\alpha \geq 1$. With such a generous strategy, a player can ensure that her payoff is never above the opponent's payoff. In [23] such players are called *compliers*. Although compliant strategies seem to be too generous to succeed in competitive environments, Stewart and Plotkin [21] showed that compliers do surprisingly well in round robin tournaments, in which the compliant strategy was outperforming all other strategies (including the most prominent strategies All D, Tit for Tat, Win-Stay Lose-Shift, and an extortioner strategy). Moreover, as shown in [25], a large fraction of compliant strategies is "evolutionary robust", meaning that no mutant with another strategy can have a selective advantage over a resident population of compliers.

Zero-determinant strategies thus have remarkable conceptual properties, but comparably little is known which of these strategies would evolve in a natural setup. It has recently been argued that extortioners are evolutionarily unstable [22]: since extortioners demand an extortionate share from any surplus, two interacting extortioners would end up with a surplus of zero. Moreover, numerical simulations indicate that zero-determinant strategies in general are disfavored by selection in sufficiently large populations [23]. However, this does not preclude certain zero-determinant strategies, such as compliers, to play an important role, as recently demonstrated by [21,25]. To identify such important strategies, researchers have focused on particular limiting cases of zero-determinant strategies, such as extortioners, equalizers, and compliers. Moreover, to investigate the dynamics of these strategies, previous studies either had to resort to individual-based simulations, or they needed to restrict attention to a finite subset of representative strategies [22,23,25].

Instead, it is the aim of this study to provide an analytical framework that allows to study the evolutionary dynamics of *all* zero-determinant strategies. Constructing an analytical model for the evolutionary dynamics of the iterated prisoner's dilemma is not straightforward. Already for simple memory-one strategies, a calculation of the resulting payoffs may become prohibitively laborious (for an example see [22]). To derive an analytical model of the dynamics, we will thus focus on an appropriate super-set of zero-determinant strategies: the set of all memory-one strategies that enforce a linear relation of the form (2), as in [25]. We show that if all players apply such strategies then the payoffs and the resulting adaptive dynamics take a remarkably simple form. In particular, we find that populations either move to the edge of compliers, or they move towards a neighborhood of unconditional defectors *AllD*. In this process, extortioners play an important role, as they can neutrally invade unconditional defectors, thereby

promoting the emergence of compliance. On the other hand, altruistic strategies (such as unconditional cooperators) have the opposite effect: they can subvert a population of compliers, giving rise to the evolution of selfish strategies. Which of these strategies gets the upper hand in the long run, critically depends on the population size. While small populations favor the emergence of selfish strategies, compliance succeeds as populations become sufficiently large.

Results

In the following, let us focus on the set of all memory-one strategies that enforce a linear relation between the payoffs of the two players. As players cannot set their own score [20], it is reasonable to consider only those strategies fulfilling Eq. (2) for which $\beta \neq 0$ (formally this means that we exclude the strategy $repeat = (1,1,0,0)$ from the set of zero-determinant strategies, which is fully dependent on the initial condition). In the appendix we show that this subset of strategies is then identical to the set

$$\mathcal{LR} = \{p \in [0,1]^4 | \exists l, s \in \mathbb{R}, \forall q : s(A(p,q) - l) = A(q,p) - l\}. \quad (5)$$

Instead of the three parameters $\alpha, \beta \neq 0$ and γ, this specification only requires two free parameters, l and s. Both parameters allow an intuitive interpretation (see Figure 1). The parameter s gives the correlation between both players' payoffs. A factor $s > 0$ means that a player enforces a positive linear relation between the payoffs, whereas for $s < 0$, the payoffs obey a negative linear relation. The parameter l, on the other hand, can be considered as the payoff that a player would get against himself (see Figure 1). We thus call the parameter l the *baseline payoff*, and we refer to s as the *slope* of an $\mathcal{LR}$–strategy (in fact, the slope s is just the inverse of the extortion factor χ).

We consider an iterated prisoner's dilemma and make the common assumption that the payoffs of the one-shot game fulfill the relation $T > R > P > S$, and $R > (T+S)/2 > P$, such that mutual cooperation is the best outcome and mutual defection is the worst outcome. As payoffs then need to be in the interval $[S,T]$, and because memory-one strategies need to consist of four probabilities, there are restrictions on the linear relations that a player can enforce. In the Methods section, we show that a pair (l,s) is enforceable if

$$P \leq l \leq R$$
$$-\min\left(\frac{T-l}{l-S}, \frac{l-S}{T-l}\right) \leq s \leq 1. \qquad (6)$$

For example, the set of extortioners corresponds to the set of pairs (l,s) with $l = P$ and $s > 0$. The set of compliers is given by those memory-one strategies for which $l = R$ and $s > 0$. In the following, we study the evolution of zero-determinant strategies by considering the dynamics on the (l,s)-plane. That is, we assume that each player determines an enforceable pair (l,s) and then picks a p from the corresponding class of $\mathcal{LR}$-strategies. Depending on the player's performance in the game, the enforceable pair (l,s) may then be adopted by others, a process that we will describe with adaptive dynamics and individual-based simulations.

Adaptive Dynamics in Infinite Populations

In order to derive the adaptive dynamics on the (l,s)–plane, we first have to calculate the payoffs for each player. While the payoff
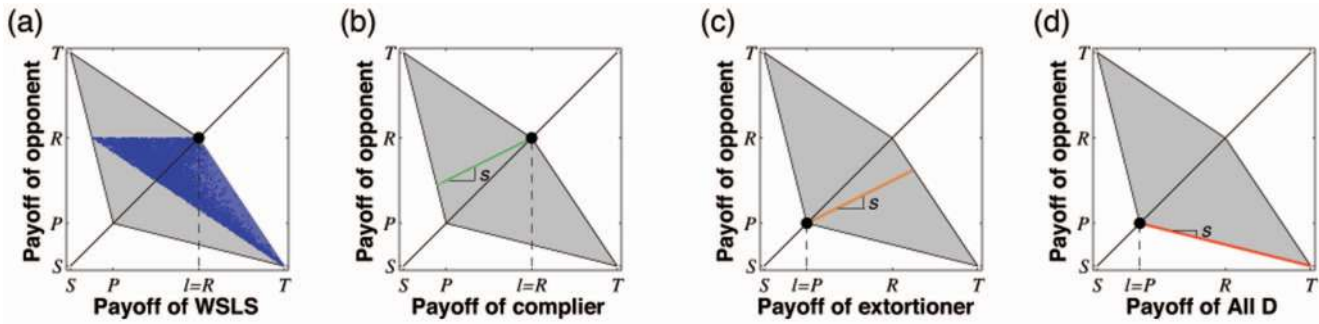
Figure 1. Illustration of zero-determinant strategies for an iterated prisoner's dilemma with $T=5$, $R=3$, $P=1$ **and** $S=0$. All graphs show the possible payoffs of the focal player (on the horizontal axis) and the resulting payoff for the opponent (on the vertical axis) as colored areas or lines. The colored points represent the payoff pairs for 10^3 randomly chosen opponents. (a) In general, as for example when the focal player applies the win-stay lose-shift strategy $p=(1,0,0,1)$, the possible payoff pairs form a convex polygon. (b) However, if the focal player applies a compliant strategy, the set of all possible payoff pairs degenerates to a line with positive slope s, which intersects the diagonal at $l=R$. (c) An extortioner enforces payoff relations that are on a line with positive slope s, intersecting the diagonal at $l=P$. (d) The strategy $AllD=(0,0,0,0)$ enforces a linear relation between the payoffs of the two players although $AllD$ is not a zero-determinant strategy for the given parameters, as described in the Methods section.
doi:10.1371/journal.pone.0077886.g001

function for general memory-one strategies is highly non-trivial, these calculations become straightforward for $\mathcal{LR}$-strategies. Suppose a player wants to enforce the linear relation (l_1,s_1) by choosing an appropriate $\mathcal{LR}$-strategy p, whereas the co-player enforces the pair (l_2,s_2) by choosing $q\in\mathcal{LR}$. Then the payoffs are implicitly given by

$$
\begin{aligned}
s_1(A(p,q)-l_1) &= A(q,p)-l_1 \\
s_2(A(q,p)-l_2) &= A(p,q)-l_2.
\end{aligned}
\tag{7}
$$

From this, we recover the result that a player can set the co-player's score to a fixed value [20,27]: by choosing $s_2=0$, player 2 can guarantee that the first player's payoff is l_2 (i.e., the set of so-called equalizers corresponds to all enforceable pairs (l,s) with $s=0$).

Excluding the two non-generic cases that both players enforce the most extreme payoff relations ($s_1=s_2=1$ or $s_1=s_2=-1$), this system of two linear equations has a unique solution for the payoffs

$$
\begin{aligned}
A(l_1,s_1;l_2,s_2) &:= A(p,q) = \frac{(1-s_1)s_2}{1-s_1\cdot s_2}\cdot l_1 + \frac{1-s_2}{1-s_1\cdot s_2}\cdot l_2 \\
A(l_2,s_2;l_1,s_1) &:= A(q,p) = \frac{1-s_1}{1-s_1\cdot s_2}\cdot l_1 + \frac{(1-s_2)s_1}{1-s_1\cdot s_2}\cdot l_2.
\end{aligned}
\tag{8}
$$

It follows that if both players have the same baseline payoff, $l:=l_1=l_2$, then their payoff will be l, irrespective of their choice of the slopes s_1 and s_2. In particular, the payoff of a homogeneous (l,s)-population is l. As a consequence, if we consider homogeneous populations, and if we assume that the populations move towards the direction where mutants have the highest invasion fitness, then the resulting adaptive dynamics [28–30] is given by

$$
\begin{aligned}
\frac{d}{dt}s &= \frac{\partial}{\partial s_2}A(l_2,s_2;l_1,s_1)|_{l_2=l_1=l,s_2=s_1=s} = 0 \\
\frac{d}{dt}l &= \frac{\partial}{\partial l_2}A(l_2,s_2;l_1,s_1)|_{l_2=l_1=l,s_2=s_1=s} = \frac{s}{1+s}
\end{aligned}
\tag{9}
$$

The first equation implies that the slope s remains constant under adaptive dynamics. Nevertheless, the initial value of s determines the eventual fate of the population: if individuals enforce a positive correlation between payoffs ($s>0$), then the baseline payoff l increases over time. Eventually, such a population will thus yield the maximum payoff R, i.e. the population converges to the edge of compliers, see Fig. 2. On the other hand, for $s<0$ the population payoffs l decrease over time, and the dynamics leads to strategies in the neighborhood of $AllD$. Interestingly, although extortioners always outcompete their direct opponent, the edge of extortioners is unstable, as illustrated in Fig. 2. Along this edge, mutants with higher baseline payoff l can invade. By giving in to the extortioners' claim, they are able to yield a payoff that exceeds the payoff P that extortioners get against themselves. However, this argument rests on the assumption of an infinite population, such that the probability for an extortioner to interact with a rare, but profitable mutant is zero. In the following section, we therefore extend our analysis to finite populations.

Adaptive Dynamics in Finite Populations

Extortioners play a more prominent role in finite populations [23], where pairwise payoff advantages have a stronger effect (see also [14,31]). This is most intuitive when the population only consists of two individuals; since extortioners outperform their direct co-player by definition, extortion is expected to spread. These observations suggest that a given extortionate strategy can be stable as long as the population size is below some critical threshold. To calculate this threshold analytically, let us consider a homogeneous population of size N that enforces the pair (l_1,s_1). From time to time, a player may mutate to a different enforceable pair (l_2,s_2). If mutation (or exploration) events are sufficiently rare, the strategy of the mutant goes extinct, or fixates, before the next mutation occurs [32,33]. In this case, the fixation probability ρ is the decisive quantity for the evolutionary dynamics. It can be shown that such a process can be described with a modified form of the adaptive dynamics equation; instead of asserting that homogenous populations move towards the direction where mutants have the highest invasion fitness, it is assumed that the population moves towards the direction where mutants have the highest fixation probability. In Imhof and Nowak [34] it is shown that this direction can be found by calculating the adaptive dynamics for a slightly perturbed payoff matrix (called the effective
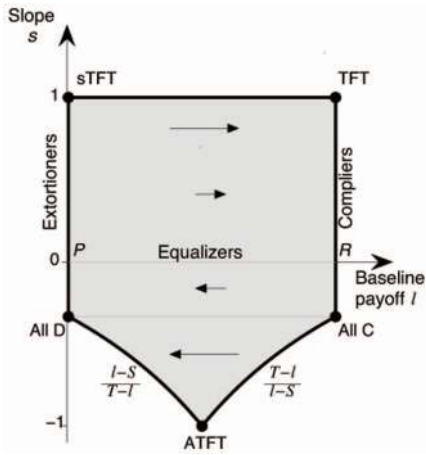
Figure 2. Adaptive dynamics in the (l,s)-plane. The grey-shaded state space represents the set of all enforceable linear relations that fulfill the inequalities (6). The corners of this state space consist of the payoff relations (l,s) that correspond to the five strategies Always Cooperate (AllC), Tit-for-Tat (TFT, which starts with cooperation, and then repeats the opponent's previous move), Suspicious Tit-for-Tat (sTFT, which starts with defection and then repeats the opponent's previous move), Always Defect (AllD), and an Anti-Tit-for-Tat strategy (ATFT, which always plays the opposite of the opponent's previous move). Three special subsets of this state space are of particular interest: (i) Extortioners are strategies for which l = P and s > 0. (ii) Equalizers are strategies with s = 0 (iii) Compliers correspond to the edge l = R and s > 0. The grey line between AllD and AllC corresponds to the set of linear relationships that can be enforced with unconditional strategies (in particular it follows that all unconditional strategies enforce linear relationships with a negative slope, see Methods section). The adaptive dynamics for this system is surprisingly simple: orbits are parallel to the l-axis; for s > 0, they converge towards the edge of compliers, whereas for s < 0, they converge towards the left boundary of the state space. Parameters: T = 3, R = 2, P = 0, S = −1.
doi:10.1371/journal.pone.0077886.g002

payoff matrix, or modified payoff matrix, see [35,36]),

$$\tilde{A}(p,q) := A(p,q) - \frac{A(p,q) + A(q,p)}{N}, \qquad (10)$$

The first correction term, $A(p,q)/N$ means that individuals cannot play against themselves, whereas the second correction term $A(q,p)/N$ corresponds to the competition effect in finite populations. In our case, the adaptive dynamics for finite populations becomes

$$
\begin{aligned}
\frac{d}{dt}s &= \frac{\partial \tilde{A}(l_2,s_2;l_1,s_1)}{\partial s_2}\Big|_{l_2=l_1=l,s_2=s_1=s} = 0 \\
\frac{d}{dt}l &= \frac{\partial \tilde{A}(l_2,s_2;l_1,s_1)}{\partial l_2}\Big|_{l_2=l_1=l,s_2=s_1=s} = \frac{(N-1)s-1}{(1+s)N}
\end{aligned}
\qquad (11)
$$

Remarkably, the slope s remains invariant for all population sizes. However, the dynamics for the baseline payoff l changes for small N: in the extreme case of $N = 2$, all trajectories in the interior of the state space lead to the lowest possible population payoff. For $N > 2$, a bistable situation emerges: if the value of s in the initial population exceeds $1/(N-1)$, then the population moves towards the edge of compliers (with $l = R$), whereas for smaller values of s populations move towards a non-cooperative

equilibrium (with $l = P$). Therefore, larger populations promote the evolution of cooperative behaviors, and in the limit of infinitely large populations, $N \to \infty$, we recover the original adaptive dynamics (9). The dynamical equations (11) also imply that a given extortionate strategy can only be stable if $s < 1/(N-1)$, or equivalently if the strategy's extortion factor $\chi = s^{-1}$ fulfills $\chi > N-1$. Thus, to be stable in a finite population, extortioners need to be sufficiently demanding ($\chi > N-1$), whereas compliers must not be too generous ($\chi < N-1$).

In order to confirm these predictions, we have simulated the dynamics in finite populations for a pairwise comparison process, where the probability to switch to the role model's strategy is given by a Fermi function [37,38]. We assume that mutations follow Gaussian distributions around l and s and focus on the distribution of strategies and on the distribution of payoffs. For $N = 2$ we find that the population clusters around the edge of low population payoffs (see Fig. 3a), and the density function for the payoffs has a single peak at $l = P$. Increasing the population size has a two-fold effect (Fig. 3b and 3c). First, compliant strategies with $s > 1/(N-1)$ become stable, such that the density function of the population payoffs has a second peak at $l = R$. Second, increasing the population size reduces the stochastic noise; as a consequence almost all the mass is concentrated around the two peaks $l = R$ and $l = P$. As predicted by adaptive dynamics, and in line with previous results [23], larger populations exhibit larger payoffs. For example, payoffs for a population size $N = 100$ exceed the payoffs for $N = 2$ by more than a factor of six.

Although extortioners seem to apply a fully selfish strategy, they are important as they can act as a catalyst for cooperation, by helping the population to escape from states with low payoffs [23]. Our adaptive dynamics formalism allows us to give an intuitive explanation for this effect: under a local mutation scheme, a population of AllD players can only be invaded by neutral drift, by moving along the vertical line of strategies with $l = P$. For cooperative strategies to have a selective advantage, the new resident population needs to have a positive slope s (i.e., only when the new resident applies an extortionate strategy, cooperation can evolve). In order to confirm this catalytic effect of extortionate strategies, we have removed a δ-neighborhood around the edge of extortioners from the set of enforceable pairs (see Fig. 4a; in [34] this method is called a knock-out experiment). That is, only those mutants are permitted that are sufficiently different from extortioners. The result is surprising: although extortioners are defined as strategies with the lowest payoff against themselves, their exclusion reduces the average payoff of the population for all population sizes $N > 2$ (Fig. 4b). This effect is especially pronounced in larger populations; for $N = 100$, Fig. 4b indicates that it is almost impossible to reach a cooperative regime without extortioners.

So far, we have assumed that a mutant's strategy is close to the parent's strategy (which allowed us to use derivatives to approximate the dynamics), and that mutations are rare (which allowed us to focus on games between a resident and one mutant strategy). Let us now weaken these assumptions and numerically explore the impact of non-local mutations, and of different mutation rates, respectively. In Fig. 5, we distinguish four simulations, according to whether the mutation rate is high or low ($\mu = 0.05$ vs. $\mu = 0.001$), and whether mutations occur on a local or on a global level (mutant strategies are drawn from a normal distribution around the parent's strategy, vs. mutant strategies are uniformly distributed over the set of enforceable pairs). These simulations indicate that all treatments follow the same pattern: average payoffs are close to the minimum P in small populations, and they increase with population size. However,
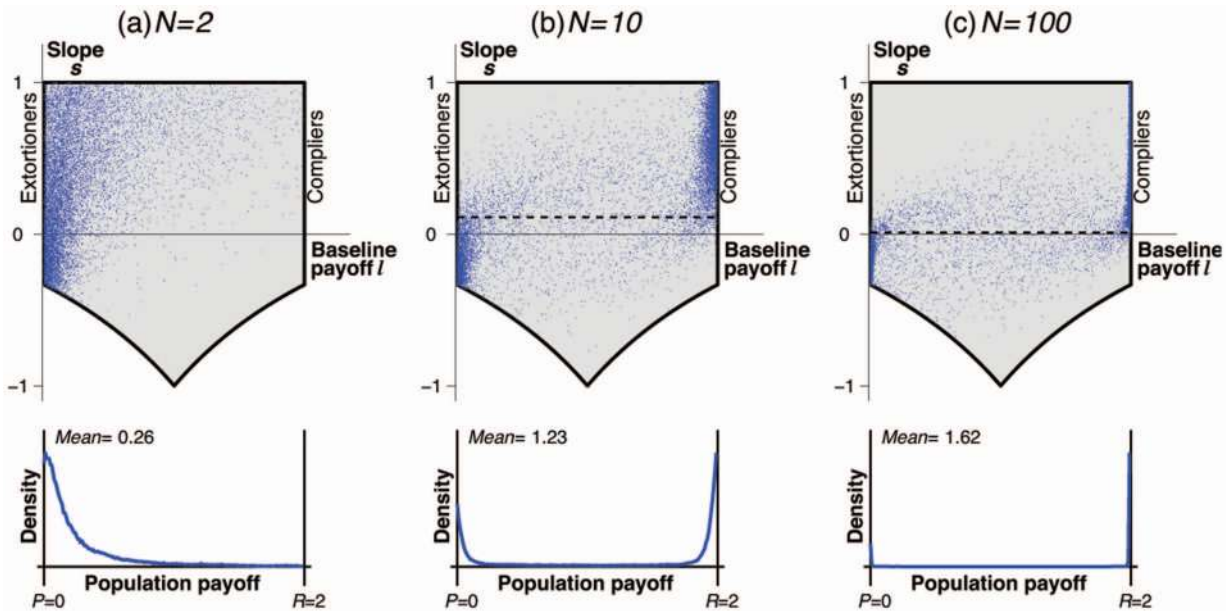
Figure 3. Stochastic dynamics for different population sizes. We consider a homogeneous population of size $N \in \{2,10,100\}$. Once a mutation occurs, the mutant strategy either takes over the whole population (with probability ρ), or goes extinct before the next mutation arises. This leads to a sequence of residents in the state space, which is shown in the upper three graphs (the dashed line corresponds to the threshold $1/(N-1)$). The lower three graphs give the distribution of the resulting payoffs in the population. (a) In the extreme case of $N=2$, most players enforce a strategy with baseline payoff $l=P$. In particular, extortion strategies can persist. (b) As population size increases, a bistable situation emerges: the population clusters along the edges with $(l=P, s<1/(N-1))$ and $(l=R, s>1/(N-1))$. (c) For large population sizes, this implies that the edge of compliers is (neutrally) stable, whereas the edge of extortioners is unstable. As a consequence, mean payoffs increase with population size. The figure shows simulation runs for 10^5 residents for a prisoner's dilemma with $T=3$, $R=2$, $P=0$, $S=-1$. New mutant strategies are randomly drawn from a Gaussian distribution around the parent strategy ($\sigma=0.05$). The invasion probability ρ of a mutant is calculated as $\rho=(1+\sum_{m=1}^{N-1}\Pi_{j=1}^{m}\exp(-\omega(A_M-A_R)))^{-1}$, where A_M and A_R are the respective payoffs of mutants and residents, and where $\omega=10$ is the strength of selection.
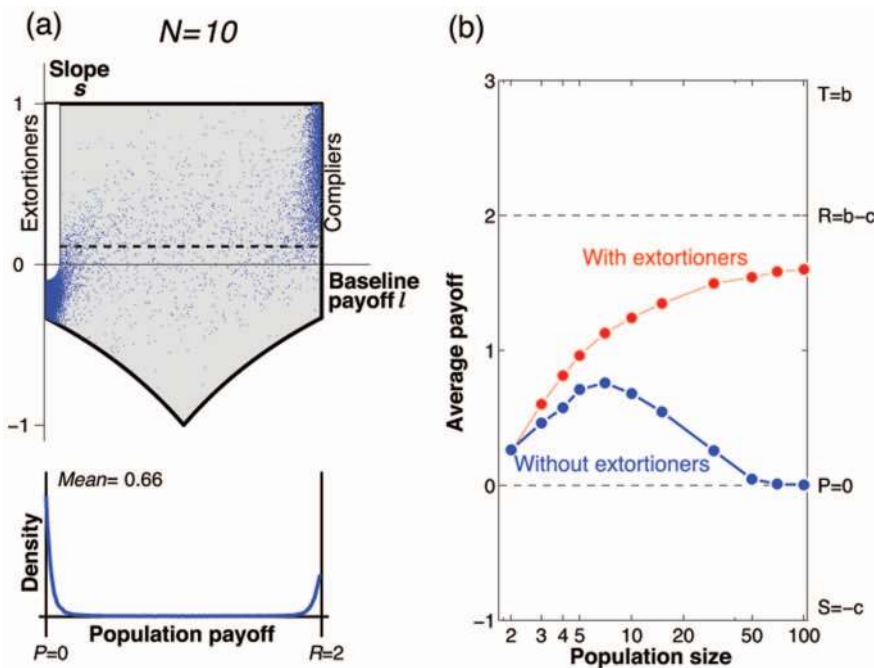doi:10.1371/journal.pone.0077886.g003

Figure 4. Extortioners facilitate cooperation. In order to study the impact of extortioners on the evolutionary dynamics, we have excluded all mutant strategies that are δ-close to the set of extortioners. (a) For the simulations we have used $\delta=0.1$, represented by the white area in the upper left corner of the panel. (b) As a result, we find for all population sizes $N>2$ that the removal of extortioners decreases the average payoff. This decrease is particularly dramatic in large populations, $N \geq 50$. Parameters are the same as in Fig. 3.
doi:10.1371/journal.pone.0077886.g004

there is a clear difference between treatments with local mutations and treatments with non-local mutations. If mutations are local, populations can be trapped in regions with a low payoff for a considerable time, although distant mutant strategies would offer an immediate escape. For example, we have seen that any strategy of the form ($l = P$, $s < 0$) forms a stable fixed point of the adaptive dynamics. However, once we allow mutants to adopt any strategy of the state space, mutants with s close to one and $l > P$ can easily invade (in fact, in Stewart and Plotkin [25] it is shown that in sufficiently large populations, compliant strategies with $s \approx 1$ can replace any noncooperative zero-determinant strategy). Overall, non-local mutations thus lead to a shift of the invariant distribution towards more cooperative strategies.

Discussion

The set of zero-determinant strategies exhibits a fascinating variety of possible behaviors, ranging from extortioners to compliant strategies, and from selfish strategies to altruists. To evaluate the evolutionary relevance of these different possible behaviors, previous studies focused on particular subsets. Adami and Hintze [22] demonstrated that neither extortioners nor equalizers are evolutionarily stable, and Hilbe et. al. [23] confirmed numerically that these two subsets are only favored by selection if the population is sufficiently small. In contrast, as shown by Stewart and Plotkin [25], large population sizes favor the emergence of compliant strategies, which are evolutionary robust (they can only be invaded by neutral drift), and which in turn are quite successful in invading other strategies. However, this focus on specific subsets of zero-determinant strategies comes at the risk of neglecting other important subsets. Thus, here we have systematically explored the space of all zero-determinant strategies.

To this end, we have derived the adaptive dynamics for all strategies that enforce a linear relation between the payoffs of the two players. This set of strategies includes all zero-determinant strategies [20] and all unconditional strategies such as *AllC* or *AllD* (see Methods section), but not all memory-one strategies (for example, it does not contain the win-stay lose-shift rule depicted in
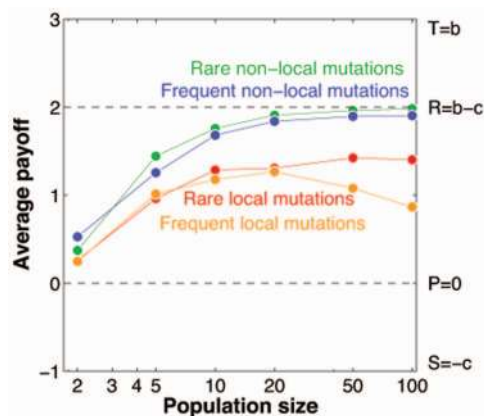

Figure 5. Average payoffs for the four different mutation treatments. In rare-mutation treatments, the mutation rate is set to $\mu = 0.001$, whereas in frequent-mutation treatments the mutation rate is $\mu = 0.05$. Local mutations are randomly drawn from a Gaussian distribution around the parent strategy, non-local mutations are randomly drawn from the entire state space. The rare local mutations correspond to the previous simulations in Figs. 3 and 4. All other parameters are the same as before.
doi:10.1371/journal.pone.0077886.g005

Figure 1a). The focus on this strategy space allows us to describe the evolutionary dynamics with an analytically tractable model. The resulting dynamics in large populations is bistable and the state space contains two neutrally stable sets. When the initial population enforces a positive relation between payoffs ($s > 0$), the population is most likely to end up at the edge of compliers. This subset of strategies shares the following three properties: (i) compliers enforce a linear relation between the payoffs of the two players, (ii) a population of compliers yields the maximum possible payoff $l = R$, and (iii) compliers play a best response to themselves (no strategy can yield a payoff higher than R when playing against a complier, see also [24] for a characterization of such strategies). However, compliers have one shortcoming: they can be neutrally invaded by altruistic strategies (strategies that accept a decrease of their own payoff to increase the opponent's payoff, such as *AllC* with $s = -c/b$). Such altruistic strategies give rise to selfish behaviors, leading the population to a neighborhood of *AllD*. To escape from that neighborhood, extortioners play an important role [23]: they can invade *AllD* by neutral drift and promote the emergence of compliant strategies. Thus, the route from cooperation to defection goes via altruism, whereas the route from defection to cooperation goes via extortion.

It is natural to ask which of these dynamical results on the space of all zero-determinant strategies are robust when we consider evolution in more general strategy spaces, such as memory-one strategies, or strategies encoded by a finite automaton (see, for example, [5]). Further simulations suggest that our results hold more generally: for Fig. 6 we consider the adaptive dynamics on the space of all memory-one strategies (similar simulations are also presented in [23,25]). The numerical results confirm our analytical predictions based on the adaptive dynamics framework: extortioners are strongest in small populations, whereas compliers succeed in large populations. Note, however, that zero-determinant strategists in general are disfavored by selection as the population size increases. In fact, as our analysis suggests, a large proportion of zero-determinant strategies only play a transient role in the evolutionary dynamics. For most of the time, the population applies a strategy that is close to one of the boundaries $l = P$ and $l = R$, whereas interior states are hardly visited. The dynamics is centered around the edge of selfish strategies and extortioners, and around the edge of compliers and altruists, whereas the evolutionary importance of other zero-determinant strategies seems negligible.

Our results on the adaptive dynamics of zero-determinant strategies resemble the results for the evolution of reactive strategies (i.e., memory-one strategies with $p_R = p_T$ and $p_S = p_P$, [5,28,34]). In both models, there are two regimes. There is a *cooperation rewarding zone* where populations evolve towards an edge of fully cooperative strategies (the edge of compliers, or the edge between tit-for-tat and generous tit-for-tat, respectively). Outside of this cooperation rewarding zone, populations move towards lower population payoffs (ending up at a neighborhood of *AllD*). These similarities are not a mere coincidence. Instead, for games with equal gains from switching (when $R + P = S + T$), every reactive strategy is a zero-determinant strategy [23] and thus reactive strategies form a subset of $\mathcal{LR}$. Conversely, we show in the Methods section that any enforceable payoff relation (l, s) can be enforced by a reactive strategy in this case. Thus, for games with equal gains from switching, the space $\mathcal{LR}$ is essentially equivalent to the space of reactive strategies.

Throughout this manuscript, we have focused on the dynamics of an iterated prisoner's dilemma. However, only a few of our results actually depend on the characteristic order of payoffs, $T > R > P > S$. In fact, the only result specific to the prisoner's
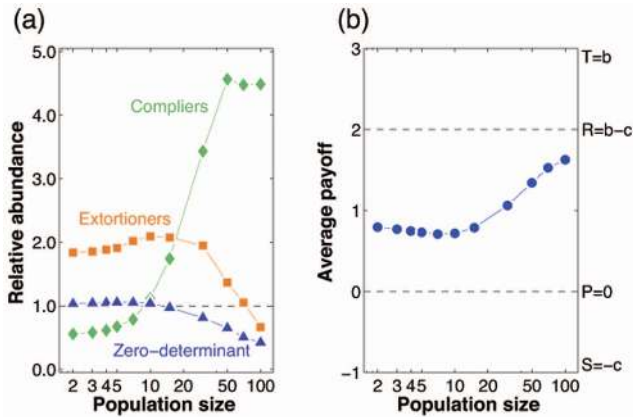
Figure 6. Statistics for the stochastic dynamics on the space of all memory-one strategies. Instead of taking the enforceable pairs (l,s) as the evolving traits, we consider the adaptive dynamics on the space of memory-one strategies $p = (p_R, p_S, p_T, p_P)$, see also [23,25]. (a) To assess the impact of zero-determinant strategies, extortioners, and compliers, we record how often the evolving population is in a δ-neighborhood of these strategy sets, and compare this to their expected abundance in a neutral process. A given strategy set is thus favored by selection if its relative abundance exceeds one. Our simulations indicate that in small populations extortioners are favored by selection, whereas in large populations compliers are favored. (b) As a consequence, average payoffs increase with population size. Simulations are run for a sequence of 10^6 mutants. We assume that mutant strategies are uniformly distributed over the space of memory-one strategies, and use the parameters $\omega = 1$ and $\delta = 0.1$. The other parameters of the evolutionary process are the same as in the previous figures.

dilemma concerns the characterization of enforceable (l,s) pairs in Eq. (6). For games that are different from the prisoner's dilemma, the geometry of the state space may thus be different, but the dynamics on the respective state space remains unchanged. In Figure 7, we illustrate this observation by considering the dynamics of an iterated snowdrift game (which is defined by the payoff relations $T = b$, $R = b - c/2$, $S = b - c$, $P = 0$ with $0 < c < b$ such that $T > R > S > P$, see [39,40]). For snowdrift games we observe that only a subset of extortionate strategies is feasible [41]: extortionate strategies with $l = P$ need to fulfill the requirement $s \geq (b-c)/b$ (i.e. the maximum extortion factor is $\chi = 1/s = b/(b-c)$). Moreover, only strategies that yield a baseline payoff higher than $l = S$ can enforce a payoff relation with negative slope, $s < 0$. As a consequence, any sufficiently large initial population that yields a payoff less than S against itself can be replaced by more cooperative mutant strategies with higher baseline payoffs. As in the prisoner's dilemma, this dynamics leads to the edge of compliers, which can only be left by neutral invasion of altruists.

Similar results may be feasible for social dilemmas with a continuous action space, as for example considered in [42–46]. However, transferring our findings to the continuous case is not straightforward. First, the existing literature on zero-determinant strategies exclusively deals with games where the players can only choose among two actions (either to cooperate or to defect), and it is not obvious how the corresponding proofs can be generalized to iterated games with continuous action spaces. Moreover, even if continuous games admit zero-determinant strategies, one may wonder which linear relations (l,s) these strategies can enforce. Is there an upper bound on the extortion factor? Which payoffs can

be enforced by an equalizer strategy? The answers to these questions are likely to depend on specific details of the benefit and cost function, representing an interesting topic for future research.

Our results confirm that extortionate behaviors can only prevail in small populations. In large populations, the evolutionary steady state is increasingly biased in favor of cooperative strategies. This may come as a surprise, as it has been shown that intermediate population sizes are optimal for the fixation of rare cooperative mutants in a population of defectors [14]. However, compliant strategies do not need to invade defectors directly. Instead, in sufficiently large populations extortioners always provide an escape path to leave non-cooperative populations. More importantly, once compliant strategies are common, they are evolutionary robust [25], with the neutral invasion of overly altruistic strategies as their only weak spot. Overall, compliance succeeds.

Methods

The Geometry of the State Space

Let us first show that the set of all strategies that fulfill condition (2) coincides with the set $\mathcal{LR}$, as defined by (5). If we multiply the condition

$$s(A(p,q) - l) = A(q,p) - l \tag{12}$$

with some $\Phi \neq 0$, then we can relate (2) and (5) by the following transform of coordinates

$$
\begin{aligned}
\alpha &= \Phi s \\
\beta &= -\Phi \\
\gamma &= \Phi(1-s)l
\end{aligned}
\tag{13}
$$

It then follows from (1) that a zero-determinant strategy p enforces the pair (l,s) if and only if there is a Φ such that p has the form

$$
\begin{aligned}
p_R &= 1 - \Phi(1-s)(R-l) \\
p_S &= 1 - \Phi(s(l-S) + (T-l)) \\
p_T &= \Phi((l-S) + s(T-l)) \\
p_P &= \Phi(1-s)(l-P).
\end{aligned}
\tag{14}
$$

Since all entries p_i need to be in the interval $[0,1]$, there are restrictions on the pairs (l,s) that can be enforced by zero-determinant strategies. For the parameters of the prisoner's dilemma, it follows by $p_P \geq 0$ and $p_R \leq 1$ that baseline payoffs l need to fulfill the condition $P \leq l \leq R$. Again because $p_P \geq 0$ and $p_R \leq 1$, we may then conclude that $\Phi(1-s) \geq 0$. As a consequence, the requirement $p_S \leq 1$ yields $s \leq 1$ and $\Phi > 0$. Then $p_T \geq 0$ leads to the restriction $s \geq -(l-S)/(T-l)$, whereas $p_S \leq 1$ implies $s \geq -(T-l)/(l-S)$. In summary, we conclude that for all pairs (l,s) that fulfill

$$
\begin{aligned}
P \leq l \leq R \\
-\min\left(\frac{T-l}{l-S}, \frac{l-S}{T-l}\right) \leq s \leq 1.
\end{aligned}
\tag{15}
$$

there is a corresponding zero-determinant strategy p of the form (14) such that $p \in [0,1]^4$ (we only have to choose a Φ that is
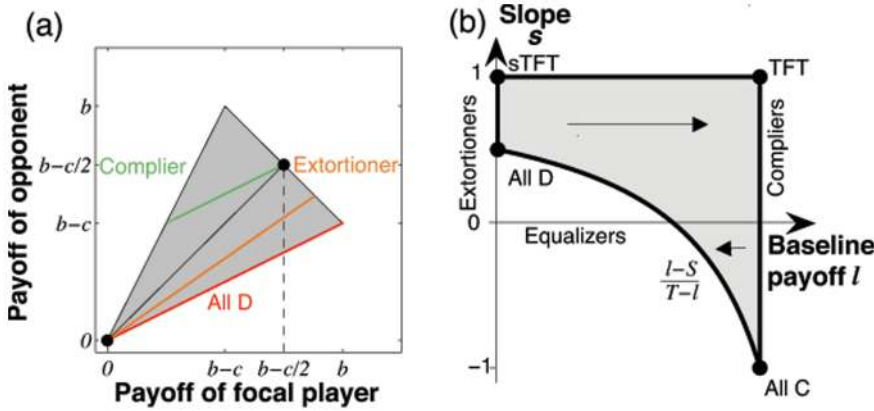
Figure 7. Zero-determinant strategies for the iterated snowdrift game (with $b=2$, $c=1$, and $T=b$, $R=b-c/2$, $S=b-c$, $P=0$). (a) The grey shaded area gives the space of feasible payoff pairs in the snowdrift game. The three colored lines give three examples of possible payoff combinations if the focal player uses a strategy that enforces a linear relation between payoffs. Unlike in the iterated prisoner's dilemma, the slope of *AllD* is positive, $s=(b-c)/b$. (b) The grey-shaded area depicts the space of possible combinations of baseline payoff l and slopes s that are enforceable in the snowdrift game. A comparison with Fig. 2 shows that the state space differs considerably from the state space of a prisoner's dilemma game. However, the qualitative dynamics within the state space remains unchanged.
doi:10.1371/journal.pone.0077886.g007

sufficiently small). Conversely, the linear relations (l,s) that can be enforced by zero-determinant strategies are in fact all possible linear relations that can be enforced in an iterated prisoner's dilemma with $R>(T+S)/2>P$. To see this, we note that for any memory-one strategy p we have:

1. The payoff pair $(A(p,AllD), A(AllD,p))$ is on the line between (S,T) and (P,P), whereas
2. the payoff pair $(A(p,AllC), A(AllC,p))$ is on the line between (T,S) and (R,R).

Thus, any linear payoff relation (l,s) enforced by some $p \in \mathcal{LR}$ connects the line segment between (S,T) and (P,P) with the line segment between (T,S) and (R,R) (see also Figs. 1b–1d). A straightforward computation verifies that any such linear payoff relation (l,s) needs to meet the conditions (15).

The set $\mathcal{LR}$ is a proper super set of the zero-determinant strategies. For example, the strategy $AllD = (0,0,0,0)$ is not a zero-determinant strategy in the general prisoner's dilemma (it is only a zero-determinant strategy in games with equal gains from switching, i.e. when $R+P=S+T$). However, $AllD \in \mathcal{LR}$ holds true in all prisoner's dilemma games. In fact, every unconditional strategy (r,r,r,r) is an element of $\mathcal{LR}$, with parameters

$$
\begin{aligned}
l &= (1-r)^2 P + r(1-r)(T+S) + r^2 R \\
s &= -\frac{(1-r)(P-S) + r(T-R)}{(1-r)(T-P) + r(R-S)} \\
\Phi &= (1-r)(T-P) + r(R-S)
\end{aligned}
\tag{16}
$$

In particular, it follows that unconditional strategies can only enforce linear payoff relations with negative slopes s. As previously suggested, these values of l and s satisfy the inequalities (15) for all r; any linear relation (l,s) that can be enforced by an unconditional strategy can also be enforced by a zero-determinant strategy.

Given a triplet (α,β,γ), the corresponding zero-determinant strategy p is uniquely determined by (1). However, for a given pair (l,s) there will generally be many zero-determinant strategies p that enforce the corresponding linear relationship in (5) - one for every Φ in (14). We call two strategies $p,p' \in \mathcal{LR}$ equivalent, and write

$p \sim p'$, if they give rise to the same pair (l,s). To study the evolutionary dynamics of $\mathcal{LR}$-strategies, we consider the dynamics on the space of equivalence classes $\mathcal{LR}/_\sim$. That is, we assume that each player determines a pair (l,s) and then picks a p from the corresponding class of $\mathcal{LR}$-strategies. The dynamics is well-defined in the sense that the adaptive dynamics does not depend on the choice of the class representative p.

$\mathcal{LR}$-strategies Versus Reactive Strategies

When payoffs fulfill equal gains from switching, $R+P=T+S$, we can choose $\Phi=1/(R-S+s\cdot(T-R))$ such that the zero-determinant strategies according to Eqs. (14) are given by

$$
\begin{aligned}
p_R &= \frac{s(T-l)+l-S}{s(T-R)+R-S} \\
p_S &= \frac{(1-s)(l-P)}{s(T-R)+R-S} \\
p_T &= \frac{s(T-l)+l-S}{s(T-R)+R-S} \\
p_P &= \frac{(1-s)(l-P)}{s(T-R)+R-S}
\end{aligned}
\tag{17}
$$

In particular, $p_R = p_T$ and $p_S = p_P$, i.e. all resulting zero-determinant strategies are reactive strategies. For such reactive strategies it follows that for $P \le l \le R$ the conditions $0 \le p_R \le 1$ and $0 \le p_S \le 1$ are equivalent to the conditions

$$
\begin{aligned}
-\frac{l-S}{T-l} &\le s \le 1 \\
-\frac{T-l}{l-S} &\le s \le 1,
\end{aligned}
\tag{18}
$$

respectively. From this, we conclude that for games with equal gains from switching, all payoff relations (l,s) that can be enforced by zero-determinant strategies (given by the conditions (15)) can already be enforced by reactive strategies.

Acknowledgments

Author Contributions

Conceived and designed the experiments: CH MAN AT. Performed the experiments: CH MAN AT. Analyzed the data: CH MAN AT. Wrote the paper: CH MAN AT.

References

1. Rapoport A, Chammah AM (1965) Prisoner's Dilemma. University of Michigan Press, Ann Arbor.
2. Trivers RL (1971) The evolution of reciprocal altruism. The Quarterly Review of Biology 46: 35–57.
3. Axelrod R (1984) The Evolution of Cooperation. New York, NY: Basic Books.
4. Nowak MA (2006) Evolutionary Dynamics. Harvard University Press, Cambridge.
5. Sigmund K (2010) The calculus of selfishness. Princeton Univ. Press.
6. Friedman J (1971) A non-cooperative equilibrium for supergames. Review of Economic Studies 38: 1–12.
7. Molander P (1985) The optimal level of generosity in a selfish, uncertain environment. Journal of Conflict Resolution 29: 611–618.
8. Fudenberg D, Maskin E (1986) The folk theorem in repeated games with discounting or with incomplete information. Econometrica 54: 533–554.
9. Milinski M (1987) Tit For Tat in sticklebacks and the evolution of cooperation. Nature 325: 433–435.
10. Boyd R, Lorberbaum J (1987) No pure strategy is evolutionary stable in the iterated prisoner's dilemma game. Nature 327: 58–59.
11. Nowak MA, Sigmund K (1993) A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. Nature 364: 56–58.
12. Frean MR (1994) The prisoner's dilemma without synchrony. Proceedings of the Royal Society B 257: 75–79.
13. Hauert C, Schuster HG (1997) Effects of increasing the number of players and memory size in the iterated prisoner's dilemma: a numerical approach. Proceedings of the Royal Society B 264: 513–519.
14. Nowak MA, Sasaki A, Taylor C, Fudenberg D (2004) Emergence of cooperation and evolutionary stability in finite populations. Nature 428: 646–650.
15. Imhof LA, Fudenberg D, Nowak MA (2005) Evolutionary cycles of cooperation and defection. Proceedings of the National Academy of Sciences USA 102: 10797–10800.
16. Perc M, Szolnoki A, Szabó G (2008) Restricted connections among distinguished players support cooperation. Physical Review E 78: 066101.
17. Grujic J, Fosco C, Araujo L, Cuesta J, Sanchez A (2010) Social experiments in the mesoscale: Humans playing a spatial prisoner's dilemma. PLoS One 5: e13749.
18. Perc M, Wang Z (2010) Heterogeneous aspirations promote cooperation in the prisoner's dilemma game. PLoS One 5: e15117.
19. van Veelen M, García J, Rand DG, Nowak MA (2012) Direct reciprocity in structured populations. Proceedings of the National Academy of Sciences USA 109: 9929–9934.
20. Press WH, Dyson FD (2012) Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. Proceedings of the National Academy of Sciences USA 109: 10409–10413.
21. Stewart AJ, Plotkin JB (2012) Extortion and cooperation in the prisoner's dilemma. Proceedings of the National Academy of Sciences USA 109: 10134–10135.
22. Adami C, Hintze A (2013) Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything. Nature Communications 4: 2193.
23. Hilbe C, Nowak MA, Sigmund K (2013) The evolution of extortion in iterated prisoner's dilemma games. Proceedings of the National Academy of Sciences USA 110: 6913–6918.
24. Akin E (2013) Stable cooperative solutions for the iterated prisoner's dilemma. arXiv : 1211.0969v2.
25. Stewart AJ, Plotkin JB (2013) From extortion to generosity, evolution in the iterated prisoner's dilemma. Proceedings of the National Academy of Sciences USA (in press).
26. Ball P (2012) Physicists suggest selfishness can pay. Nature, doi:10.1038/nature.2012.11254.
27. Boerlijst MC, Nowak MA, Sigmund K (1997) Equal pay for all prisoners. American Mathematical Monthly 104: 303–307.
28. Nowak MA, Sigmund K (1990) The evolution of stochastic strategies in the prisoner's dilemma. Acta Applicandae Mathematicae 20: 247–265.
29. Metz JAJ, Geritz SAH, Meszena G, Jacobs FJA, van Heerwaarden JS (1996) Adaptive dynamics: a geometrical study of the consequences of nearly faithful replication. In: van Strien SJ, Ver18 duyn Lunel SM, editors, Stochastic and Spatial Structures of Dynamical Systems, Amsterdam: North Holland. 183–231.
30. Geritz SAH, Kisdi E, Meszéna G, Metz JAJ (1998) Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evolutionary Ecology Research 12: 35–57.
31. Hilbe C, Traulsen A (2012) Emergence of responsible sanctions without second order free riders, antisocial punishment or spite. Nature Scientific Reports 2: 458.
32. Antal T, Scheuring I (2006) Fixation of strategies for an evolutionary game in finite populations. Bulletin of Mathematical Biology 68: 1923–1944.
33. Wu B, Gokhale CS, Wang L, Traulsen A (2012) How small are small mutation rates? Journal of Mathematical Biology 64: 803–827.
34. Imhof LA, Nowak MA (2010) Stochastic evolutionary dynamics of direct reciprocity. Proceedings of the Royal Society B 277: 463–468.
35. Lessard S (2011) Effective game matrix and inclusive payoff in group-structured populations. Dynamic Games and Applications 1: 301–318.
36. Hilbe C (2011) Local replicator dynamics: A simple link between deterministic and stochastic models of evolutionary game theory. Bulletin of Mathematical Biology 73: 2068–2087.
37. Blume LE (1993) The statistical mechanics of strategic interaction. Games and Economic Behavior 5: 387–424.
38. Traulsen A, Nowak MA, Pacheco JM (2006) Stochastic dynamics of invasion and fixation. Physical Review E 74: 011909.
39. Sugden R (1986) The Economics of Rights, Co-operation and Welfare. Oxford and New York: Blackwell.
40. Doebeli M, Hauert C (2005) Models of cooperation based on the prisoner's dilemma and the snowdrift game. Ecology Letters 8: 748–766.
41. Roemheld L (2013) Evolutionary extortion and mischief - zero determinant strategies in 2×2 games. arXiv 1308.2576.
42. Roberts G, Sherratt TN (1998) Development of cooperative relationships through increasing in vestment. Nature 394: 175–179.
43. Killingback T, Doebeli M (1999) 'Raise the stakes' evolves into a defector. Nature 400: 518.
44. Wahl LM, Nowak MA (1999) The continuous prisoner's dilemma: I. Linear reactive strategies. Journal of Theoretical Biology 200: 307–321.
45. Wahl LM, Nowak MA (1999) The continuous prisoner's dilemma: II. Linear reactive strategies with noise. Journal of Theoretical Biology 200: 323–338.
46. Killingback T, Doebeli M (2002) The continuous Prisoner's Dilemma and the evolution of cooperation through reciprocal altruism with variable investment. The American Naturalist 160: 421–438.