International Institute for
Applied Systems Analysis
Schlossplatz 1
A-2361 Laxenburg, Austria

Tel: +43 2236 807 342
Fax: +43 2236 71313
E-mail: publications@iiasa.ac.at
Web: www.iiasa.ac.at

**Interim Report** **IR-10-040**

# Incentives and opportunism: From the carrot to the stick

Christian Hilbe (christian.hilbe@univie.ac.at)
Karl Sigmund (ksigmund@iiasa.ac.at)

**Approved by**

Ulf Dieckmann
Program Leader, EEP

June 2011

# Incentives and Opportunism: from the Carrot to the Stick

Christian Hilbe[1] & Karl Sigmund[1,2]
[1]Faculty of Mathematics, University of Vienna,
Nordbergstrasse 15, A-1090 Vienna, Austria
[2]IIASA, A-2361 Laxenburg, Austria

March 15, 2010

### Abstract

Cooperation in public good games is greatly promoted by positive and negative incentives. In this paper, we use evolutionary game dynamics to study the evolution of opportunism (the readiness to be swayed by incentives) and the evolution of trust (the propensity to cooperate in the absence of information on the co-players). If both positive and negative incentives are available, evolution leads to a population where defectors are punished and players cooperate, except when they can get away with defection. Rewarding behavior does not become fixed, but can play an essential role in catalyzing the emergence of cooperation, especially if the information level is low.

**Key index words:** evolutionary game theory, cooperation, reward, punishment, reputation

**Short title for page headings: Incentives and Opportunism**

**Length of abstract: 134 words**

**Length of text (excluding formulas, including captions): 5,670 words**

**Number of figures: 4**

# 1 Introduction

Social dilemmas are obstacles to the evolution of cooperation. Examples such as the Prisoner's Dilemma show that self-interested motives can dictate self-defeating moves, and thus suppress cooperation. Positive and negative incentives (the carrot and the stick) can both induce cooperation in a population of self-regarding agents (see e.g. Olson 1965, Ostrom & Walker 2003, Sigmund 2007). The provision of such incentives is costly, however, and therefore raises a second-order social dilemma. This issue has been addressed in many papers, particularly for the case of negative incentives. We mention for example Yamagishi 1986, Boyd & Richerson 1992, Fehr & Gächter 2002, Walker & Halloran 2004, Bowles & Gintis 2004, Gardner & West 2004, Nakamaru & Iwasa 2006, Sefton *et al* 2007, Carpenter 2007, Lehmann *et al.* 2007 and Kiyonari & Barclay 2008.

It is easily seen that the efficiency of the two types of incentives relies on contrasting and even complementary circumstances. Indeed, if most players cooperate, then it will be costly to reward them all, while punishing the few defections will be cheap: often, the mere threat of a sanction suffices (Boyd *et al* 2003, Gächter *et al* 2008). On the other hand, if most players defect, then punishing them all will be a costly enterprise, while rewarding the few cooperators will be cheap. Obviously, therefore, the best policy for turning a population of defectors into a population of cooperators would be, first, to use the carrot, and at some later point, the stick.

In the absence of a proper institution to implement such a policy, members of the population can take the job onto themselves. But what is *their* incentive to do so? It pays only if the threat of a punishment, or the promise of a reward, should turn a co-player from a defector into a cooperator. Hence, the co-players must be opportunistic, i.e., prone to be swayed by incentives.

In order to impress a co-player, the threat (or promise) of an incentive must be sufficiently credible. In the following model, we shall assume that the credibility is provided by the players' reputation, i.e. by their history, and thus assume several rounds of the game, not necessarily with the same partner (see e.g. Sigmund *et al* 2001, Fehr & Fischbacher 2003, Barclay 2006). Credibility could alternatively be provided by a verbal commitment, for example. Since mere talk is cheap, however, such commitments need to be convincing; ultimately, they must be backed up by actions, and hence again rely on reputation. Whether a player obtains information about the co-players' previous actions from direct experience, or from witnessing them at a distance, or hearing about them through gossip, can be left open at this stage. In particular, we do not assume repeated rounds between the same two players, but do not exclude them either. Basically, the carrot or the stick will be applied after the cooperation, or defection, and hence are forms of targeted reciprocation (while conversely, of course, the promise to return good with good and bad with bad, can act as an incentive).

In the following, we present a simple game theoretic model to analyze the evolution of opportunism, and to stress the smooth interplay of positive and negative incentives. The model is based on a previous paper (Sigmund *et al* 2001, see also Hauert *et al* 2004), which analyses punishment and reward separately and which *presumes* opportunistic agents. Here, we show how such opportunistic agents evolve via social learning, and how first rewards, then punishment lead to a society dominated by players who cooperate, except when they expect that they can get away with defection. Rewards will not become stably established; but they can play an essential role in the transition to cooperation, especially if the information level is below a specific threshold. Whenever the benefit-to-cost ratio for the reward is larger than one, the eventual demise of rewarders is surprising, since a homogeneous population of rewarding cooperators would obtain a higher payoff than a homogeneous population of punishing cooperators. We first analyze the model by means of the replicator dynamics, then by means of a stochastic learning model based on the Moran process. Thus both finite populations and the limiting case of infinite populations will be covered. In the discussion, we study the role of errors, compare our results with experiments and point out the need to consider a wider role for incentives.

## 2 The model

Each round of the game consists of two stages, a helping stage and an incentive stage. Individuals in the population are randomly paired. A dice decides who is in the role of the (potential) Donor, and who is Recipient. In the first stage, Donors may transfer a benefit $b$ to their Recipients, at their own cost $c$, or they may refuse to do so. These two alternatives are denoted by $\mathbf{C}$ (for cooperation) resp. $\mathbf{D}$ (for defection). In the second stage, Recipients can reward their Donors, or punish them, or refuse to react. If rewarded, Donors receive an amount $\beta$; if punished, they must part with that amount $\beta$; in both cases, Recipients must pay an amount $\gamma$, since both rewarding and punishing is costly. As usual, we assume that $c < b$, as well as $c < \beta$ and $\gamma < b$. Using the same parameter values $\beta$ and $\gamma$ for both types of incentives is done for convenience only: basically, all that matters are the inequalities. They ensure that Donors are better off by choosing **C**, if their Recipients use an incentive; and that in the case of rewards, both players have a positive payoff. But material interests speak against using incentives, since they are costly; and in the absence of incentives, helping behavior will not evolve.

The four possible moves for the second stage will be denoted by **N**, to do nothing; **P**, to punish defection; **R**, to reward cooperation; and **I**, to provide for both types of incentives, i.e. to punish defection *and* to reward cooperation. For the first stage, next to the two unconditional moves **AllC**, to always cooperate, and **AllD**, to always defect, we also consider the opportunistic move: namely to defect except if prodded by an incentive. We shall, however, assume that information about the co-player may be incomplete. Let $\mu$ denote the probability to know whether the co-player provides an incentive or not, and set $\bar{\mu} = 1 - \mu$. We consider two types of opportunists, who act differently under uncertainty: players of type $\mathbf{O_C}$ defect only if they know that their co-player provides no incentive, and players of type $\mathbf{O_D}$ defect except if they know that an incentive will be delivered. Hence in the absence of information, $\mathbf{O_C}$ players play **C** and $\mathbf{O_D}$-players **D**. This yields sixteen strategies, each given by a pair $[i, j]$, with $i \in M_D := \{\mathbf{AllC}, \mathbf{O_C}, \mathbf{O_D}, \mathbf{AllD}\}$ specifying how the player acts as a Donor and $j \in M_R := \{\mathbf{N}, \mathbf{P}, \mathbf{R}, \mathbf{I}\}$ how the player acts as Recipient. If player I is Donor and player II Recipient, the pair

$(p_I, p_{II})$ of their payoff values is determined by their moves in the corresponding roles. Hence we can describe these pairs by a $4 \times 4$ matrix $(a_{[ij]}, b_{[ij]})$ given by

| $*$ | **N** | **P** | **R** | **I** |
|---|---|---|---|---|
| **AllC** | $(-c, b)$ | $(-c, b)$ | $(\beta - c, b - \gamma)$ | $(\beta - c, b - \gamma)$ |
| $\mathbf{O_C}$ | $(-\bar{\mu}c, \bar{\mu}b)$ | $(-c, b)$ | $(\beta - c, b - \gamma)$ | $(\beta - c, b - \gamma)$ |
| $\mathbf{O_D}$ | $(0, 0)$ | $(-\mu c - \bar{\mu}\beta, \mu b - \bar{\mu}\gamma)$ | $(\mu(\beta - c), \mu(b - \gamma))$ | $(-(1 - 2\mu)\beta - \mu c, \mu b - \gamma)$ |
| **AllD** | $(0, 0)$ | $(-\beta, -\gamma)$ | $(0, 0)$ | $(-\beta, -\gamma)$ |

This specifies the payoff values for the corresponding symmetrized game, which is given by a $16 \times 16$-matrix. A player using $[i, j]$ against a player using $[k, l]$ is with equal probability in the role of the Donor or the Recipient and hence obtains as payoff $(a_{[i,l]} + b_{[k,j]})/2$. The state of the population $\mathbf{x} = (x_{[i,j]})$ is given by the frequencies of the 16 strategies.

There exist a wealth of possible evolutionary dynamics, describing how the frequencies of the strategies change with time under the influence of social learning (Hofbauer and Sigmund 1998). We shall consider only one updating mechanism, but stress that the results hold in many other cases too. For the learning rule, we shall use the familiar Moran-like 'death-birth' process (Nowak 2006): we thus assume that occasionally, players can update their strategy by copying the strategy of a 'model', i.e., a player chosen at random with a probability which is proportional to that player's fitness. This fitness in turn is assumed to be a convex combination $(1-s)B + sP$, where $B$ is a 'baseline fitness' (the same for all players), $P$ is the payoff (which depends on the model's strategy, and the state of the population), and $0 \leq s \leq 1$ measures the 'strenght of selection', i.e. the importance of the game for overall fitness. (We shall always assume $s$ small enough to avoid negative fitness values). This learning rule corresponds to a Markov process. The rate for switching from strategy $[k, l]$ to strategy $[i, j]$ is $(1 - s)B + sP_{[i,j]}$, independently of $[k, l]$.

## 2.1 Large populations

The learning rule leads, in the limiting case of an infinitely large population, to the replicator equation for the relative frequencies $x_{[ij]}$: the growth rate of any strategy is given by the difference between its payoff and the average payoff in the population (Hofbauer & Sigmund, 1998). This yields an ordinary differential equation which can be analyzed in a relatively straightforward way, despite being 15-dimensional.

Let us first note that **I** is weakly dominated by **P**, in the sense that **I**-players never do better, and sometimes less well, than **P**-players. Hence, no state where all the strategies are played can be stationary. The population always evolves towards a region where at least one strategy is missing. Furthermore, **AllC** is weakly dominated by $\mathbf{O_C}$, and **All D** by $\mathbf{O_D}$. This allows to reduce the dynamics to lower dimensional cases. Of particular relevance are the states where only two strategies are present, and where these two strategies prescribe the same move in one of the two stages of the game. The outcome of such pairwise contests is mostly independent of the parameter values, with three exceptions:

(a) In a homogeneous $\mathbf{O_C}$-population, **R** dominates **N** if and only if $\mu > \frac{\gamma}{b}$;

(b) In a homogeneous $\mathbf{O_D}$-population, **P** dominates **N** if and only if $\mu > \frac{\gamma}{b+\gamma}$;

(c) In a homogeneous $\mathbf{O_D}$-population, **P** dominates **R** if and only if $\mu > 1/2$.

In each case, it is easy to understand why higher reputation will have the corresponding effect. Owing to our assumption $\gamma < b$, all these thresholds for $\mu$ lie in the open interval (0,1).

One can obtain a good representation of the dynamics by looking at the situations where there are two alternatives for the first stage (namely **AllD** and $\mathbf{O_C}$, or **AllD** and $\mathbf{O_D}$, or $\mathbf{O_C}$ and $\mathbf{O_D}$), and the three alternatives **N**, **P** and **R** for the second stage. In each such case, the state space of the population can be visualized by a prism (Fig. 1). Here, each of its 'square faces' stands for the set of all mixed populations with only four strategies present. For instance, if the population consists only of the four strategies $[\mathbf{O_C}, \mathbf{N}]$, $[\mathbf{O_C}, \mathbf{R}]$, $[\mathbf{O_D}, \mathbf{N},]$ and $[\mathbf{O_D}, \mathbf{R}]$, then the state corresponds to a point in the three dimensional simplex spanned by the corresponding four monomorphic populations. But since the double ratios $x_{[ij]}x_{[kl]}/x_{[il]}x_{[kj]}$ are invariant under the replicator dynamics (see Hofbauer & Sigmund 1998, pp. 122–125), the state cannot leave the corresponding two-dimensional surface, which may be represented by a square (see Fig. 2).

For several pairs of strategies (such as $[\mathbf{O_C}, \mathbf{P}]$ and $[\mathbf{AllC}, \mathbf{P}]$, or $[\mathbf{AllD}, \mathbf{N}$ and $[\mathbf{O_D}, \mathbf{N}]$), all populations which are mixtures of the corresponding two strategies are stationary. There is no selective force favouring one strategy over the other. We shall assume that in this case, small random shocks will cause the state to evolve through neutral drift. This implies that evolution then leads ultimately to $[\mathbf{O_C}, \mathbf{P}]$, and hence to a homogeneous population which stably cooperates in the most efficient way. Indeed, it is easy to see that no other strategy can invade a monomorphic $[\mathbf{O_C}, \mathbf{P}]$-population through selection. The only flaw is that $[\mathbf{AllC}, \mathbf{P}]$ can enter through neutral drift. Nevertheless, $[\mathbf{O_C}, \mathbf{P}]$ is a Nash equilibrium.

But how can $[\mathbf{O_C}, \mathbf{P}]$ get off the ground? Let us first consider what happens if the possibility to play **R**, i.e. to reward a cooperative move, is excluded. The asocial strategy $[\mathbf{AllD}, \mathbf{N}]$ is stable. It can at best be invaded through neutral drift by $[\mathbf{O_D}, \mathbf{N}]$. If $\mu > \gamma/(b + \gamma)$, this can in turn be invaded by $[\mathbf{O_D}, \mathbf{P}]$, which then leads to $[\mathbf{O_C}, \mathbf{P}]$. If $\mu$ is smaller, however, that path is precluded and the population would remain in an un-cooperative state. It is in this case that the **R**-alternative plays an essential role. By neutral drift, $[\mathbf{AllD}, \mathbf{R}]$ can invade $[\mathbf{AllD}, \mathbf{N}]$. More importantly, $[\mathbf{O_D}, \mathbf{R}]$ dominates $[\mathbf{O_D}, \mathbf{N}]$, $[\mathbf{AllD}, \mathbf{R}]$ and $[\mathbf{AllD}, \mathbf{N}]$. From $[\mathbf{O_D}, \mathbf{R}]$, the way to $[\mathbf{O_C}, \mathbf{R}]$ and then to $[\mathbf{O_C}, \mathbf{P}]$ is easy.

The essential step of that evolution occurs in the transition from $\mathbf{O_D}$ to $\mathbf{O_C}$, when players start cooperating by default, i.e., in the absence of information (see the third column in Fig. 1). If the **R**-alternative is not available, then for small values of $\mu$, the population can be trapped in $[\mathbf{O_D}, \mathbf{N}]$. But if the **R**-alternative can be used, it can switch from $[\mathbf{O_D}, \mathbf{N}]$ to $[\mathbf{O_D}, \mathbf{R}]$ . In a population where the first move is either $\mathbf{O_D}$ or $\mathbf{O_R}$, and the second move either **N** or **R**, there is a (four-membered) Rock-Paper-Scissors cycle, see Fig. 2: one strategy is superseded by the next. There exists a unique stationary state where these four alternatives are used. We show in the electronic supplementary material that for $\mu < \gamma/2b$, this stationary state cannot be invaded by any strategy using $\mathbf{P}$. But due to the Rock-Paper-Scissors dynamics, it is inherently unstable. The population will eventually use mostly strategy $[\mathbf{O_C}, \mathbf{R}]$. There, the strategy $[\mathbf{O_C}, \mathbf{P}]$ can invade and become fixed.

In the competition between $[\mathbf{O_D}, \mathbf{N}]$ and $[\mathbf{O_C}, \mathbf{P}]$, the latter is dominant if and only if $\mu > (c + \gamma)/(c + \gamma + b)$ (a condition which is independent of $\beta$). If not, then the competition is bistable, meaning that neither strategy can invade a homogeneous population adopting the other strategy. An equal mixture of both strategies converges to the pro-social strategy $[\mathbf{O_C}, \mathbf{P}]$ if and only if $\mu(\beta - 2c - 2b - \gamma) < \beta - 2c - \gamma$. In the case $\gamma = \beta$, this simply reduces to $\mu > c/(c + b)$.

We thus obtain a full classification of the replicator dynamics in terms of the parameter $\mu$. The main bifurcation values are $\frac{\gamma}{2b} < \frac{\gamma}{b+\gamma} < \frac{\gamma}{b}$ and $\frac{1}{2}$. These can be arranged in two ways, depending on whether $b < 2\gamma$ or not. But the basic outcome is the same in both case (see Fig. 1 and the electronic supplement).

It is possible to modify this model by additionally taking into account the recombination of the traits affecting the first and the second stage of the game. Indeed, recombination does not only occur for genetic transmission of strategies, but also for social learning. A modification of an argument from Gaunersdorfer *et al.* (1991) allows to show that in this case, the double ratios $x_{[ij]}x_{[kl]}/x_{[il]}x_{[kj]}$ converge to 1, so that the traits for the first and the second stage of the game become statistically independent of each other. Hence the previous analysis still holds. In Lehmann *et al.* (2007) and Lehmann & Rousset (2009) it is shown, in contrast, that recombination greatly affects the outcome in a lattice and in a finite population model without reputational effects.

## 2.2 Small mutation rates

In the case of a finite population of size $M$, the learning process corresponds to a Markov chain on a state space which consists of the frequencies of all the strategies (which sum up to $M$). The absorbing states correspond to the homogeneous populations: in such a homogeneous population, imitation cannot introduce any change. If we add to the learning process a 'mutation rate' (or more precisely, an exploration rate), by assuming that players can also adopt a strategy by chance, rather than imitation, then the corresponding process is recurrent (a chain of transitions can lead from every state to every other) and it admits a unique stationary distribution. This stationary distribution describes the frequencies of the states in the long run. It is in general laborious to compute, since the number of possible states grows polynomially in $M$. However, in the limiting case of a very small exploration rate (the so-called adiabatic case), we can assume that the population is mostly in a homogeneous state, and we can compute the transition probabilities between these states (Nowak 2006). This limiting case is based on the assumption that the fate of mutant (i.e., whether it will be eliminated or fixed in the population) is decided before the next mutation occurs. We can confirm the results from the replicator dynamics. For simplicity, we confine ourselves to the non-dominated strategies $\mathbf{O_C}$, $\mathbf{O_D}$, resp. $\mathbf{N}$, $\mathbf{P}$ and $\mathbf{R}$; similar results can be obtained by considering the full strategy space.

In the stationary distribution, the population is dominated by the strategy $[\mathbf{O_C}, \mathbf{P}]$, but for smaller values of $\mu$, it needs the presence of the $\mathbf{R}$-alternative to emerge. This becomes particularly clear if one looks at the transition probabilities (see electronic supplementary material). Except for large values of $\mu$, only the strategy $[\mathbf{O_D}, \mathbf{R}]$ can invade the asocial $[\mathbf{O_D}, \mathbf{N}]$ with a fixation probability which is larger than the neutral fixation probability $1/M$.

If $[\mathbf{O_C}, \mathbf{P}]$ dominates $[\mathbf{O_D}, \mathbf{N}]$, or when it fares best in an equal mixture of both strategies, then it needs not the help of $\mathbf{R}$-players to become the most frequent strategy in the long run (i.e. in the stationary distribution). But for smaller values of $\mu$, rewards are essential. In Fig. 3, it is shown that that the existence of rewarding strategies allows the social strategy $[\mathbf{O_C}, \mathbf{P}]$ to supersede the asocial $[\mathbf{O_D}, \mathbf{N}]$ even in cases in which the players have hardly any information about their co-players. The time until the system leaves $[\mathbf{O_D}, \mathbf{N}]$ is greatly reduced if rewarding is available (see Fig. 4). In the electronic supplementary material it is shown that the state $[\mathbf{O_C}, \mathbf{P}]$ is usually reached from $[\mathbf{O_C}, \mathbf{R}]$, while the strategy most likely to invade the asocial $[\mathbf{O_D}, \mathbf{N}]$ is $[\mathbf{O_D}, \mathbf{R}]$. These outcomes are robust, and depend little on the parameter choices. Moreover, they are barely affected by the mutation structure. If, instead of assuming that all mutations are equally likely, we only allow for mutations in the behavior in one of the two stages (i.e., no recombination between the corresponding traits), the result is very similar. Apparently, if it is impossible to mutate directly from $[\mathbf{O_D}, \mathbf{N}]$ to $[\mathbf{O_C}, \mathbf{P}]$, then the detour via $[\mathbf{O_D}, \mathbf{P}]$ works almost as well.

Even for the limiting case $\mu = 0$ (no reputation effects), the role of rewards is strongly noticeable. Without rewards, the stationary probability of the asocial strategy $[\mathbf{O_D}, \mathbf{N}]$ is close to 100 percent; with the possibility of rewards, it is considerably reduced.

# 3 Discussion

We have analyzed a two-person, two-stages game. It is well-known that it corresponds to a simplified version of the Ultimatum game (Güth *et al.* 1982), in the punishment case, or of the Trust game (Berg *et al.* 1995), in the reward case (De Silva & Sigmund 2009, Sigmund 2010). Similar results also hold for the N-person Public Good game with reward and punishment (e.g. Hauert *et al.* 2004). However, the many-person game offers a wealth of variants having an interest of their own (as, for instance, when players decide to mete out punishment only if they have a majority on their side; see Boyd *et al.*, to appear). In this paper, we have opted for the simplest set-up and considered pairwise interactions only.

In classical economic thought, positive and negative incentives have often been treated on equal footing, so to speak (Olson 1965). In evolutionary game theory, punishing is studied much more frequently than rewarding. The relevance of positive incentives is sometimes queried, on the ground that helping behavior makes only sense if there is an asymmetry in resource level between donor and recipient. If A has a high pile of wood, and B has no fuel, A can give some wood away at little cost, and provide a large benefit to B. This is the cooperative act. Where is the positive incentive? It would be absurd to imagine that B rewards A by returning the fuel. But B can reward A by donating some other resource, such as food, or fire, which A is lacking.

In experimental economics, punishing behavior has been studied considerably more often than rewarding behavior (Yamagishi 1986; Fehr & Gächter 2002; Barclay 2006; Dreber *et al.* 2008). In the last few years, there has been a substantial amount of empirical work on the interplay of the two forms of incentives (Andreoni *et al.* 2003; Rockenbach & Milinski 2006; Sefton *et al.* 2007). The results, with two exceptions to be discussed presently, confirm our theoretical conclusion: punishment is the more lasting factor, but the combination of reward and punishment works best. This outcome is somewhat surprising, because in most experiments, players are anonymous and know that they cannot build up a reputation. One significant exception is the investigation, in Fehr & Fischbacher (2003), of the Ultimatum game, which has essentially the same structure as our two-stage game with punishment. In that case, the treatment without information on the co-player's past behavior yields a noticeably lower level of cooperation than the treatment with information. Nevertheless, even in the no-information treatment, both the level of cooperation (in the form of fair sharing) and of punishment (in the form of rejection of small offers) are remarkably high.

A serious criticism of the model presented in this paper is thus that it does not seem to account for the pro-social behavior shown by players who know that reputation-building is impossible. We believe that this effect is due to a maladaptation. Our evolutionary past has not prepared us to expect anonymity. In hunter-gatherer societies and in rural life, it is not often that one can really be sure to be unobserved. Even in modern life, the long phase of childhood is usually spent under the watchful eyes of parents, educators or age-peers. Ingenious experiments uncover our tendency to over-react to the slightest cues indicating that somebody may be watching (for instance, the mere picture of an eye, see Haley & Fessler 2005 and Bateson *et al.* 2006, or three dots representing eyes and mouth, see Rigdon *et al.* 2009). The idea of personal deities scrutinizing our behavior, which seems to be almost universal, is probably a projection of this deep-seated conviction (Johnson & Bering 2006). The concept of conscience was famously described, by Mencken, as 'the inner voice that warns us somebody may be looking' (cf. Levin 2009).

In several experimental papers, however, the role of reputation is very explicit. In Rand *et al.* (2009), players are engaged in fifty rounds of the Public Goods game with incentives, always with the same three partners. Hence they know the past actions of their co-players. In this case, we can be sure that $\mu > \gamma/b$. Thus in a homogeneous $\mathbf{O_C}$-population, $\mathbf{R}$ should dominate $\mathbf{N}$. Moreover, as the leverage for both punishment and reward is 1:3 in this experiment (as in many others), an $[\mathbf{O_C}, \mathbf{R}]$-population obtains a payoff $b - c + \beta - \gamma$ which is substantially larger than that of an $[\mathbf{O_C}, \mathbf{P}]$-population. In the experiment, rewarding performs indeed much better than punishing, and Rand *et al.* conclude that 'Positive reciprocity should play a larger role than negative reciprocity in maintaining public cooperation in repeated situations.'

Nevertheless, according to our model, $\mathbf{P}$-players ought to invade. This seems counter-intuitive. Punishers do not have to pay for an incentive (since everyone cooperates), but they will nevertheless be rewarded, since they cooperate in the Public Goods stage. Thus $[\mathbf{O_C}, \mathbf{P}]$ should take over, thereby lowering the average payoff. By contrast, in the repeated game considered by Rand *et al.*, it is clear that cooperative players who have not been rewarded by their co-player in the previous round will feel cheated, and stop rewarding that co-player. They will not be impressed by the fact that the co-player is still providing an incentive by punishing defectors instead. In other words, in this experiment rewards are not only seen as incentives, but as contributions in their own right, in a Repeated Prisoner's Dilemma game. Players will reciprocate not only for the Public Goods behavior, but for the 'mutual reward game' too. In fact, if there had been two players only in the experiment by Rand *et al*, it would reduce to a Repeated Prisoner's Dilemma game with 100 rounds.

This aspect is not covered in our model, where the incentives are only triggered by the behavior in the Public Goods stage, but not by previous incentives. In particular, rewarding behavior cannot be rewarded, and fines do not elicit counter-punishment. This facilitates the analysis of incentives as instruments for promoting cooperation, but it obscures the fact that in real life, incentives have to be viewed as economic exchanges in their own right.

A similar experiment as in Rand *et al.* was studied by Milinski *et al.* (2002), where essentially the Public Goods rounds alternate with an indirect reciprocity game (see also Panchanathan & Boyd 2006). Helping, in such an indirect reciprocity game, is a form of reward. In Milinski's experiment, punishment was not allowed, but in Rockenbach & Milinski (2006), both types of incentives could be used. Groups were rearranged between rounds, as players could decide whether to leave or to stay. Players knew each other's past behavior in the previous public goods rounds and the indirect reciprocity rounds (but not their punishing behavior). It was thus possible to acquire a reputation as a rewarder, but not as a punisher. This treatment usually led to a very cooperative outcome, with punishment focused on the worst cheaters, and a significant interaction between reward and punishment.

In our numerical examples, we have usually assumed $\gamma = \beta$, but stress that this does not affect the basic outcome (see electronic supplementary material for the case $\gamma < \beta$). In most experiments, the leverage of the incentive is assumed to be stronger. Clearly, this encourages the Recipients to use incentives (Carpenter 2006; Egas & Riedl 2007; Vyrastekova & van Soest 2008). But it has been shown (Carpenter 2006; Sefton *et al.* 2007) that many are willing to punish exploiters even if it reduces their own account by as much as that of the punished player. In the Trust game, it is also usually assumed that the second stage is a zero-sum game. In most of the (relatively few) experiments on rewarding, the leverage is 1:1 (Walker & Halloran 2004; Sefton *et al.* 2007), in Rockenbach & Milinski and Rand *et al* it is 1:3. In Vyrastekova & van Soest (2008), it is shown that increasing this leverage makes rewarding more efficient. In our view, it is natural to assume a high benefit-to-cost ratio in the first stage (the occasion for a Public Goods game is precisely the situation when mutual help is needed), but it is less essential that a high leverage also applies in the second stage. Punishment, for instance, can be very costly it the other player retaliates, as seems quite natural to expect (at least in pairwise interactions; in N-person games, sanctions can be inexpensive if the majority punishes a single cheater).

For the sake of simplicity, we have not considered the probability of errors in implementation. But it can be checked in a straightforward way that the results are essentially unchanged if we assume that with a small probability $\epsilon > 0$, an intended donation fails (either due to a mistake of the player, or to unfavorable conditions). The other type of errors in implementation (namely helping without wanting it) seems considerably less plausible. We note that in a homogeneous $[\mathbf{O_C}, \mathbf{P}]$-population, usually there is no need to punish co-players, and hence no way of building up a reputation as a punisher. But if errors in implementation occur, there will be opportunities for punishers to reveal their true colours. In Sigmund (2010), it is shown that if there are sufficiently many rounds of the game, occasional errors will provide enough opportunities for building up a reputation.

# References

Andreoni, J., Harbaugh, W. & Vesterlund, L. 2003 The carrot or the stick: rewards, punishments, and cooperation. *Am. Econ. Rev.* **93**, 893-902.

Barclay, P. 2006 Reputational benefits for altruistic punishment. *Evolution and Human behavior* **27**, 325-344.

Bateson, M., Nettle, D. & Roberts, G. 2006 Cues of being watched enhance cooperation in a real-world setting. *Biol. Lett.* **2**, 412-414.

Berg, J., Dickhaut, J. & McCabe, K. 1995 Trust, reciprocity, and social history. *Games and Econ. Behav.* **10**, 122-42.

Bowles, S, & Gintis, H. 2004 The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theor. Pop. Biol.* **65**, 17-28.

Boyd, R. & Richerson, P. J. 1992 Punishment allows the evolution of cooperation (or anything else) in sizeable groups. *Ethol. Sociobiol.* **13**, 171-195.

Boyd, R., Gintis, H., Bowles, S. & Richerson, P. J. 2003 The evolution of altruistic punishment *Proc. Natl. Acad. Sci.* **100**, 3531-3535.

Boyd, R., Gintis, H. & Bowles, S. Coordinated contingent punishment is group-beneficial and can proliferate when rare (to appear).

Carpenter, J. P. 2007 The demand for punishment. *J. Econ. Behav. Organ.* **62**, 522-542.

De Silva, H. & Sigmund, K. 2009 Public Good Games with Incentives: the role of reputation. In *Games, Groups and the Global Good* (ed. S. A. Levin), pp. 85-114. New York: Springer.

Dreber A., Rand, D. G., Fudenberg, D. & Nowak, M. A. 2008 Winners don't punish. *Nature*, **452** 348-351.

Egas, M. & Riedl A. 2007 The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B* **275**, 871-878.

Fehr, E. & Gächter, S. 2002 Altruistic punishment in humans. *Nature* **415**, 137-140.

Fehr, E. & Fischbacher, U. 2003 The nature of human altruism. *Nature* **425**, 785-791.

Gächter, S., Renner, E. & Sefton, M. 2008 The long-run benefits of punishment. *Science* **322**, 1510.

Gaunersdorfer, A., Hofbauer, J. & Sigmund, K. 1991 On the dynamics of asymmetric games. *Theor. Pop. Biol.* **39**, 345-357.

Gardner, A. & West, S. 2004 Cooperation and punishment, especially in humans. *Amer. Naturalist* **164**, 753-764.

Güth, W., Schmittberger, R. & Schwarze, B. 1982 An experimental analysis of ultimatum bargaining. *J. Econ. Behav. Organ.* **3**, 367-388.

Haley, K. & Fessler D. 2005 Nobodys watching? Subtle cues affect generosity in an anonymous economic game. *Evol. Hum. Behav.* **26**, 245-256.

Hauert, C., Haiden, N. & Sigmund, K. 2004 The dynamics of public goods. *Discrete Contin. Dynam. Systems B* **4**, 575-585.

Hofbauer, J. & Sigmund, K. 1998 *Evolutionary Games and Population Dynamics*. Cambridge, UK: Cambridge University Press.

Johnson, D. & Bering, J. 2006 Hand of God, mind of man: punishment and cognition in the evolution of cooperation. *Evol. Psych.* **4**, 21923.

Kiyonari, T. & Barclay, P. 2008 Cooperation in social dilemmas: free riding may be thwarted by second-order reward rather than by punishment. *J. Pers. Soc. Psychol.* **95**, 826-842.

Lehmann, L., Rousset, F., Roze, D. & Keller, L. 2007 Strong reciprocity or strong ferocity? A population genetic view of the evolution of altruistic punishment. *Amer. Nat.* **170**, 21-36.

Lehmann, L. & Rousset, F. 2009 Perturbation expansions of multilocus fixation probabilities for frequency-dependent selection with applications to the Hill-Robertson effect and to the joint evolution of helping and punishment. *Theor. Pop. Biol.* **76**, 35-51.

Levin, S. A. 2009 Preface to *Games, Groups, and the Global Good* (ed. S. A. Levin). New York: Springer.

McCabe, K. A., Rigdon, M. L. & Smith, V. L. 2003 Positive reciprocity and intentions in trust games. *J. Econ. Behav. Organ.* **52**, 267-275.

Milinski, M., Semmann, D. & Krambeck, H. J. 2002 Reputation helps solve the Tragedy of the Commons. *Nature* **415**, 424-426.

Nakamaru, M. and Iwasa, Y. 2006 The coevolution of altruism and punishment: role of the selfish punisher. *Journ. Theor. Biol.* **240**, 475-488.

Nowak, M. A., Page, K., & Sigmund, K. 2000 Fairness versus reason in the Ultimatum Game. *Science* **289**, 1773-1775.

Nowak, M. A. 2006 *Evolutionary dynamics.* Cambridge, MA: Harvard University Press.

Olson, M. 1965 *The Logic of Collective Action.* Cambridge, MA: Harvard University Press.

Ostrom, E. & Walker, J. 2003 *Trust and Reciprocity: Interdisciplinary Lessons from Experimental Research.* New York: Russel Sage Foundation.

Panchanathan, K. & Boyd, R. 2006 Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature* **432**, 499-502.

Rand, D. G., Dreber, A., Fudenberg, D., Ellingsen, T. & Nowak, M. A. 2009 Positive interactions promote public cooperation. *Science* **325**, 1272-1275.

Rigdon, M., Ishii, K., Watabe, M. and Kitayama, S. 2009 Minimal social cues in the dictator game, *J. Econ. Psychol.* **30**, 358-367.

Rockenbach, B. & Milinski, M. 2006 The efficient interaction of indirect reciprocity and costly punishment. *Nature* **444**, 718-723.

Sefton, M., Shupp, R. & Walker, J. M. 2007 The effects of rewards and sanctions in provision of public goods. *Econ. Inquiry* **45**, 671-690.

Sigmund, K., Hauert, C. & Nowak, M. A. 2001 Reward and punishment. *Proc. Natl. Acad. Sci.* **98**, 10757-10762.

Sigmund, K. 2007 Punish or Perish? Retaliation and Collaboration among humans. *Trends Ecol. Evol.* **22**, 593-600.

Sigmund, K. 2010 *The calculus of selfishness.* Princeton NJ: Princeton University Press.

Vyrastekova, J. & van Soest, D. P. 2008 On the (in)effectiveness of rewards in sustaining cooperation. *Exp. Econ.* **11**, 53-65.

Walker, J. M. & Halloran, W. A. 2004 Rewards and sanctions and the provision of public goods in one-shot settings. *Exp. Econ.* **7**, 235-247.

Yamagishi, T. 1986 The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110-116.

# Figure Captions

**Figure 1**: Dynamics of a population consisting of $\mathbf{O_C}$ and $\mathbf{AllD}$ (left column), $\mathbf{AllD}$ and $\mathbf{O_D}$ (middle column) resp. $\mathbf{O_C}$ and $\mathbf{O_D}$ (right column). Black circles represent Nash-equilibria, white circles indicate unstable fixed points. The arrows on the edges indicate the direction of the dynamics if only the two strategies corresponding to the end-points are present. The orange grid is the manifold that separates initial values with different asymptotic behavior. The blue curves represent the typical dynamics for a given initial population. Parameter values: $b = 4,\ c = 1,\ \beta = \gamma = 2$ and $\mu = 30\%$ (hence $\frac{\gamma}{2b} < \mu < \frac{\gamma}{\gamma+b}$).

**Figure 2**: The state space of a game involving the four strategies $[\mathbf{O_C}, \mathbf{N}]$, $[\mathbf{O_C}, \mathbf{R}]$, $[\mathbf{O_D}, \mathbf{N}]$ and $[\mathbf{O_D}, \mathbf{R}]$. The corners of the three-dimensional simplex correspond to the homogeneous populations using that strategy, the interior points denote mixed populations. For each initial state, the evolution of the system is restricted to a two-dimensional saddle-like manifold that can be represented by a square (right). If $\mu < \frac{\gamma}{b}$, the competition between these four strategies is characterized by a rock-paper-scissors like dynamics, as indicated by the orientation of the edges.

**Figure 3**: Strategy selection in finite and infinite populations, depending on the information parameter $\mu$. The left column shows the outcome of a simulation of the replicator equation for 1,000 randomly chosen initial populations. If only punishment is available to sway opportunistic behavior, then cooperative outcomes become more likely if $\mu$ exceeds roughly $1/3$ (in which case $[\mathbf{O_C}, \mathbf{P}]$ becomes fixed). As soon as rewards are also allowed, punishment-enforced cooperation becomes predominant as soon as $\mu > \gamma/2b = 1/4$. Additionally, for smaller values of $\mu$ the population may tend to cycle between the strategies $[\mathbf{O_C}, \mathbf{R}]$, $[\mathbf{O_C}, \mathbf{N}]$, $[\mathbf{O_D}, \mathbf{N}]$ and $[\mathbf{O_D}, \mathbf{R}]$, represented by the orange line in the lower left graph.

The right column shows the stationary distribution of strategies in a finite population. Again, without rewards a considerably higher information level $\mu$ is nec-

essary to promote punishment-enforced cooperation (either $[\mathbf{O_C}, \mathbf{P}]$ or $[\mathbf{O_D}, \mathbf{P}]$); note that both opportunist strategies become indistinguishable in the limit case of complete information. In finite populations, rewarding strategies act merely as a catalyst for the emergence of punishment; even for small $\mu$, the outcomes $[\mathbf{O_C}, \mathbf{R}]$ resp. $[\mathbf{O_D}, \mathbf{R}]$ never prevail.
Parameter values: $b = 4, c = 1, \beta = \gamma = 2$. For finite populations, the population size is $M = 100$ and the selection strength $s = 1/10$.

**Figure 4**: Average number of mutations needed until a population of $[\mathbf{O_D}, \mathbf{N}]$ players is successfully invaded. Adding the possibility of rewards reduces the waiting time considerably (for $\mu = 0\%$ it takes 500 mutations with rewards and almost 500,000 mutations without). As the information level increases, this catalytic effect of rewarding disappears. Parameter values: Population size $M = 100$, selection strength $s = 1/10$; $b = 4, c = 1, \beta = \gamma = 2$.

**Figure 1**

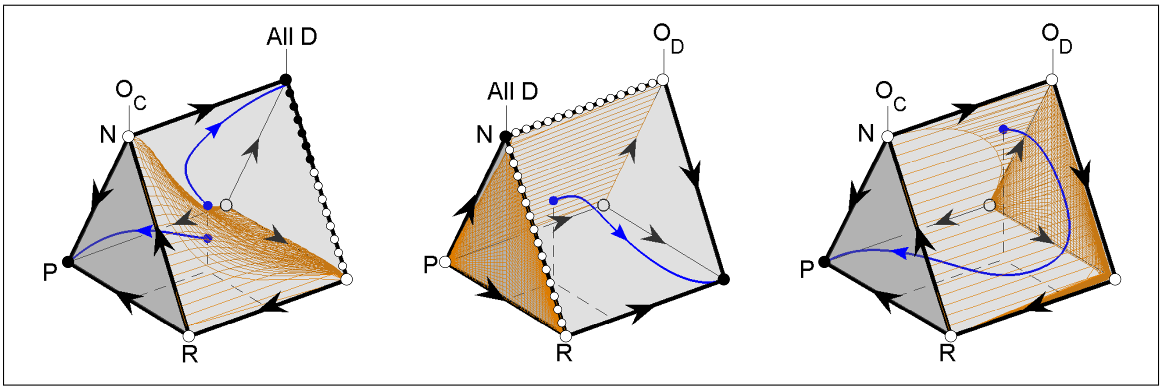
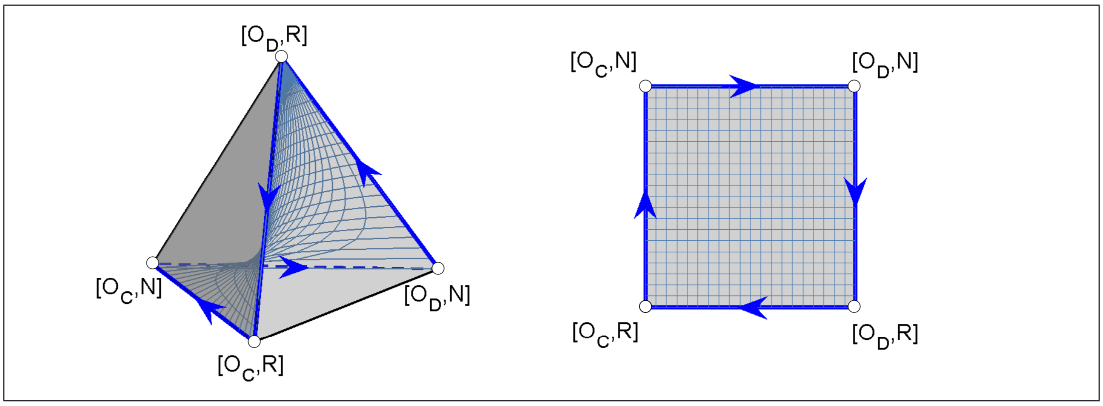


**Figure 2**



# Figure 3

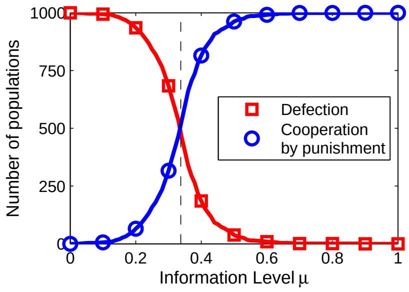
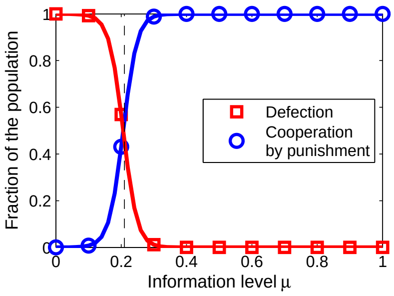
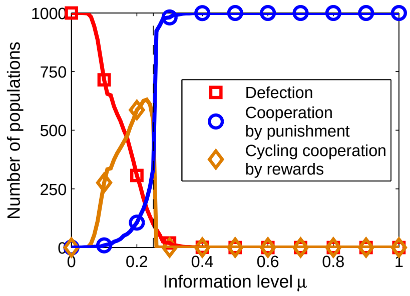
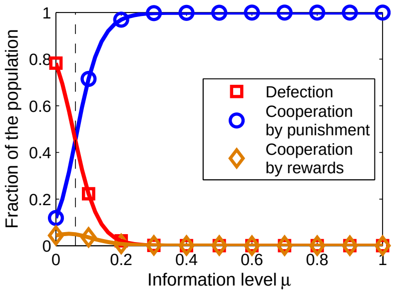
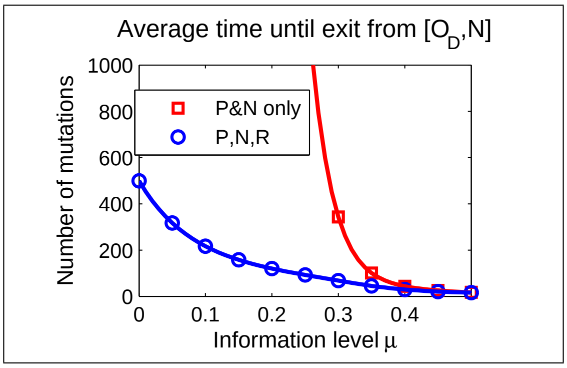
| | Infinite populations | Finite populations |
|---|---|---|
| **Punishment only** |  |  |
| **Punishment and rewards** |  |  |

**Figure 4**


Average time until exit from $[O_D,N]$
Number of mutations
1000
800
600
400
200
0
0
0.1
0.2
0.3
0.4
Information level $\mu$
P&N only
P,N,R

# Electronic supplementary material
# Incentives and Opportunism: from the Carrot to the Stick

Christian Hilbe and Karl Sigmund

## 1 Large populations

### 1.1 The dynamics on a prism

If we consider the two strategies $\mathbf{O_C}$ and $\mathbf{O_D}$ in the first stage and the alternatives $\mathbf{P}$, $\mathbf{N}$ and $\mathbf{R}$ for the second, we obtain (up to a factor 1/2) the following payoff matrix:

$$\begin{array}{c|cccccc}
 & [\mathbf{O_C},\mathbf{N}] & [\mathbf{O_D},\mathbf{N}] & [\mathbf{O_C},\mathbf{R}] & [\mathbf{O_D},\mathbf{R}] & [\mathbf{O_C},\mathbf{P}] & [\mathbf{O_D},\mathbf{P}] \\
\hline
[\mathbf{O_C},\mathbf{N}] & \bar{\mu}(b-c) & -\bar{\mu}c & \beta-c+\bar{\mu}b & \beta-c & \bar{\mu}b-c & -c \\
[\mathbf{O_D},\mathbf{N}] & \bar{\mu}b & 0 & \mu(\beta-c)+\bar{\mu}b & \mu(\beta-c) & \bar{\mu}(b-\beta)-\mu c & -\mu c-\bar{\mu}\beta \\
[\mathbf{O_C},\mathbf{R}] & b-\gamma-\bar{\mu}c & \mu(b-\gamma)-\bar{\mu}c & b-c+\beta-\gamma & \beta-c+\mu(b-\gamma) & b-c-\gamma & -c+\mu(b-\gamma) \\
[\mathbf{O_D},\mathbf{R}] & b-\gamma & \mu(b-\gamma) & b-\gamma+\mu(\beta-c) & \mu(b-c+\beta-\gamma) & b-\gamma-\mu c-\bar{\mu}\beta & \mu(b-\gamma-c)-\bar{\mu}\beta \\
[\mathbf{O_C},\mathbf{P}] & b-\bar{\mu}c & \mu b-\bar{\mu}(\gamma+c) & b-c+\beta & \beta-c+\mu b-\bar{\mu}\gamma & b-c & -c+\mu b-\bar{\mu}\gamma \\
[\mathbf{O_D},\mathbf{P}] & b & \mu b-\bar{\mu}\gamma & b+\mu(\beta-c) & \mu(b-c+\beta)-\bar{\mu}\gamma & b-\mu c-\bar{\mu}\beta & \mu(b-c)+\bar{\mu}(\beta-\gamma)
\end{array} \qquad (1)$$

Let $x_{[\mathbf{i},\mathbf{j}]}(t)$ denote the fraction of players with strategy $[\mathbf{i},\mathbf{j}]$ at time $t$. As the sum of the first and the fourth row equals the sum of the second and third row, one can easily verify that

$$V_{\mathbf{NR}}(t) = \frac{x_{[\mathbf{O_C},\mathbf{N}]}(t)\cdot x_{[\mathbf{O_D},\mathbf{R}]}(t)}{x_{[\mathbf{O_C},\mathbf{R}]}(t)\cdot x_{[\mathbf{O_D},\mathbf{N}]}(t)} \qquad (2)$$

is invariant under replicator dynamics (see Hofbauer & Sigmund 1998). Analogously, $V_{\mathbf{NP}} = x_{[\mathbf{O_C},\mathbf{N}]}x_{[\mathbf{O_D},\mathbf{P}]}/\left(x_{[\mathbf{O_C},\mathbf{P}]}x_{[\mathbf{O_D},\mathbf{N}]}\right)$ is an invariant of motion. This, together with the condition that the sum of all $x_{[\mathbf{i},\mathbf{j}]}$ equals one, reduces the 6-dimensional problem to a 3-dimensional system which can be represented by a prism.

### 1.2 A classification of the dynamics

If we return to the more convenient bimatrix notation, the game between the two strategies $\mathbf{O_C}$ and $\mathbf{O_D}$ is represented by

$$\begin{array}{c|ccc}
* & \mathbf{N} & \mathbf{P} & \mathbf{R} \\
\hline
\mathbf{O_C} & (-\bar{\mu}c,\bar{\mu}b) & (-c,b) & (\beta-c,b-\gamma) \\
\mathbf{O_D} & (0,0) & (-\mu c-\bar{\mu}\beta,\mu b-\bar{\mu}\gamma) & (\mu(\beta-c),\mu(b-\gamma))
\end{array}$$

in which the first entry denotes the payoff of the row player and the second entry the payoff of the column player. From this representation we may conclude that

1. In a homogeneous $\mathbf{O_C}$ population, $\mathbf{P}$ always dominates $\mathbf{N}$ and $\mathbf{R}$. Additionally, an $\mathbf{R}$ player obtains a higher payoff than an $\mathbf{N}$ player if and only if $b-\gamma > \bar{\mu}b$, i.e. iff

$$\mu > \gamma/b \qquad (3)$$

2. Similarly, in a homogeneous $\mathbf{O_D}$ population, $\mathbf{N}$ is always dominated by $\mathbf{R}$. $\mathbf{N}$ is also dominated by $\mathbf{P}$ iff $\mu b - \bar{\mu}\gamma > 0$, i.e. iff
$$\mu > \gamma/(b+\gamma) \tag{4}$$
$\mathbf{P}$ also dominates $\mathbf{R}$ iff $\mu b - \bar{\mu}\gamma > \mu(b-\gamma)$, i.e. iff
$$\mu > 1/2 \tag{5}$$
3. If incentives are used ($\mathbf{P}$ or $\mathbf{R}$), $\mathbf{O_C}$ dominates $\mathbf{O_D}$; in the absence of incentives ($\mathbf{N}$), this relation is reversed.

In particular, it follows that $\mathbf{N}$ is strictly dominated by $\mathbf{P}$ if $\mu > \gamma/(b+\gamma)$. Iterated elimination of strictly dominated strategies then leads to the conclusion that the only possible outcome of the dynamics is $[\mathbf{O_C}, \mathbf{P}]$.

However, if $\mu < \gamma/(b+\gamma)$ we find a more interesting behaviour. There is a uniqe fixed point $M$ in the interior of the square spanned by the strategies $[\mathbf{O_C}, \mathbf{N}]$, $[\mathbf{O_C}, \mathbf{R}]$, $[\mathbf{O_D}, \mathbf{N}]$ and $[\mathbf{O_D}, \mathbf{R}]$. This fixed point is surrounded by spiralling orbits (see Fig. 1). The asymptotic behaviour of these orbits depends on the initial condition: If $V_{\mathbf{NR}}(0) < 1$, all orbits converge to the boundary of the square; if $V_{\mathbf{NR}}(0) > 1$, M is a global attractor; and finally, if $V_{\mathbf{NR}}(0) = 1$, $M$ is surrounded by periodic orbits (see Hofbauer & Sigmund 1998). Independently of the initial condition, a population in this fixed point $M$ obtains a payoff of
$$\mu\left(-b + c - c^2/\beta + b^2/\gamma\right),$$
whereas both absent strategies, $[\mathbf{O_C}, \mathbf{P}]$ and $[\mathbf{O_D}, \mathbf{P}]$ would obtain the payoff $-\gamma + \mu\left(b + c - c^2/\beta + b^2/\gamma\right)$. Hence, a population in $M$ can be invaded by punishers if and only if
$$\mu > \gamma/(2b) \tag{6}$$

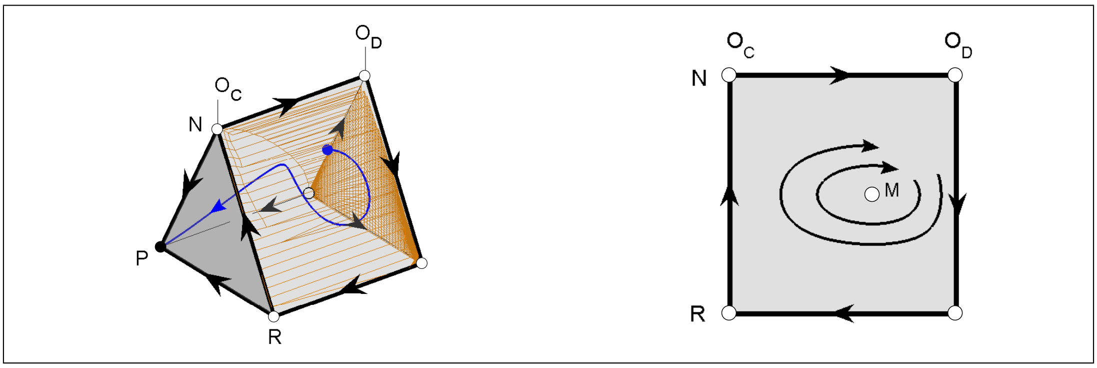


Figure 1: In the square spanned by $\mathbf{O_C}$, $\mathbf{O_D}$ resp. $\mathbf{N}$ and $\mathbf{R}$, the orbits cycle around the fixed point $M$.

If we also allow for the strategies **AllC** and **AllD** we do not see any additional bifurcations as the payoffs of these strategies do not depend on $\mu$. Hence, the dynamics of the game is fully described by the thresholds (3) - (6), which can be arranged in two possible ways:

1. If $\gamma < b/2$, we have $\frac{\gamma}{2b} < \frac{\gamma}{\gamma+b} < \frac{\gamma}{b} < 1/2$;
2. Otherwise we obtain $\frac{\gamma}{2b} < \frac{\gamma}{\gamma+b} < 1/2 < \frac{\gamma}{b}$

However, both cases induce the same long-term dynamics (see also Fig. 2, which depicts the borderline case $\gamma = b/2$): For any value of $\mu$, a population consisting of $\mathbf{O_C}$ and $\mathbf{AllD}$ evolves either towards punishment enforced cooperation, $[\mathbf{O_C}, \mathbf{P}]$, or towards a purely selfish regime, $[\mathbf{AllD}, \mathbf{N}]$ resp. $[\mathbf{AllD}, \mathbf{R}]$ (Fig. 2, first column). As $\mathbf{AllD}$ is always weakly dominated by its opportunistic counterpart, $\mathbf{O_D}$ might invade (Fig. 2, second column).
In this case the eventual outcome depends on the information level: If $\mu > \gamma/(b+\gamma)$, $[\mathbf{O_C}, \mathbf{P}]$ mutants succeed in an $\mathbf{O_D}$ population and lead to stable cooperation, independently of the additional inequalities $\mu > 1/2$ or $\mu > \gamma/b$ (Fig. 2, last two graphs in the last column). For $\mu < \gamma/(b+\gamma)$, however, only the possibility of rewards allows an $\mathbf{O_C}$ minority to invade a homogeneous $\mathbf{O_D}$ population, which may lead to oscillations between reward-driven cooperation and defection. If $\mu > \gamma/(2b)$ these cycles are unstable and open the way for $[\mathbf{O_C}, \mathbf{P}]$ (Fig. 2, upper two graphs in the last column).

### 1.3 Recombination

Classical replicator dynamics does not introduce new strategies (see Hofbauer & Sigmund 1998). For example, $[\mathbf{O_C}, \mathbf{N}]$ remains absent if it is not played in the initial population even if both pure strategies $\mathbf{O_C}$ and $\mathbf{N}$ exist (e.g. if the initial population consists of $[\mathbf{O_C}, \mathbf{P}]$ and $[\mathbf{O_D}, \mathbf{N}]$ players). In this section we introduce recombination, thereby extending the work of Gaunersdorfer *et al.* (1991).

We consider a general 2x$n$ role game with strategies $\mathbf{e_1}$ and $\mathbf{e_2}$ in the first stage and strategies $\mathbf{f_1}, ..., \mathbf{f_n}$ in the second. The fraction of players using strategy $[\mathbf{e_i}, \mathbf{f_j}]$ is denoted by $x_{ij}$; furthermore we define the fraction of all players using strategy $\mathbf{e_i}$ by $p_i := \sum_{k=1}^n x_{ik}$ and the fraction of all players using $\mathbf{f_j}$ by $q_j := \sum_{k=1}^2 x_{kj}$. Hence, if the payoffs of the game are given by the bimatrix

| | $\mathbf{f_1}$ | $\dots$ | $\mathbf{f_n}$ |
|---|---|---|---|
| $\mathbf{e_1}$ | $(a_{11}, b_{11})$ | $\dots$ | $(a_{1n}, b_{1n})$ |
| $\mathbf{e_2}$ | $(a_{21}, b_{21})$ | $\dots$ | $(a_{2n}, b_{2n})$ |

then an $[\mathbf{e_i}, \mathbf{f_j}]$ player obtains the payoff $m_{ij} = \frac{1}{2}\big(\sum_{k=1}^n a_{ik} \cdot q_k\big) + \frac{1}{2}\big(\sum_{k=1}^2 b_{kj} \cdot p_k\big)$. Then the average payoff of the population is given by $\bar{M} := \sum_{i=1}^2 \sum_{j=1}^n m_{ij} \cdot x_{ij}$. The replicator dynamics of this system is given by

$$\dot{x}_{ij} = x_{ij} \cdot \big[m_{ij} - \bar{M}\big]. \tag{7}$$

By an analogous argument as in section 1.1 the double ratios

$$Z_k := \frac{x_{11} \cdot x_{2k}}{x_{21} \cdot x_{1k}} \tag{8}$$

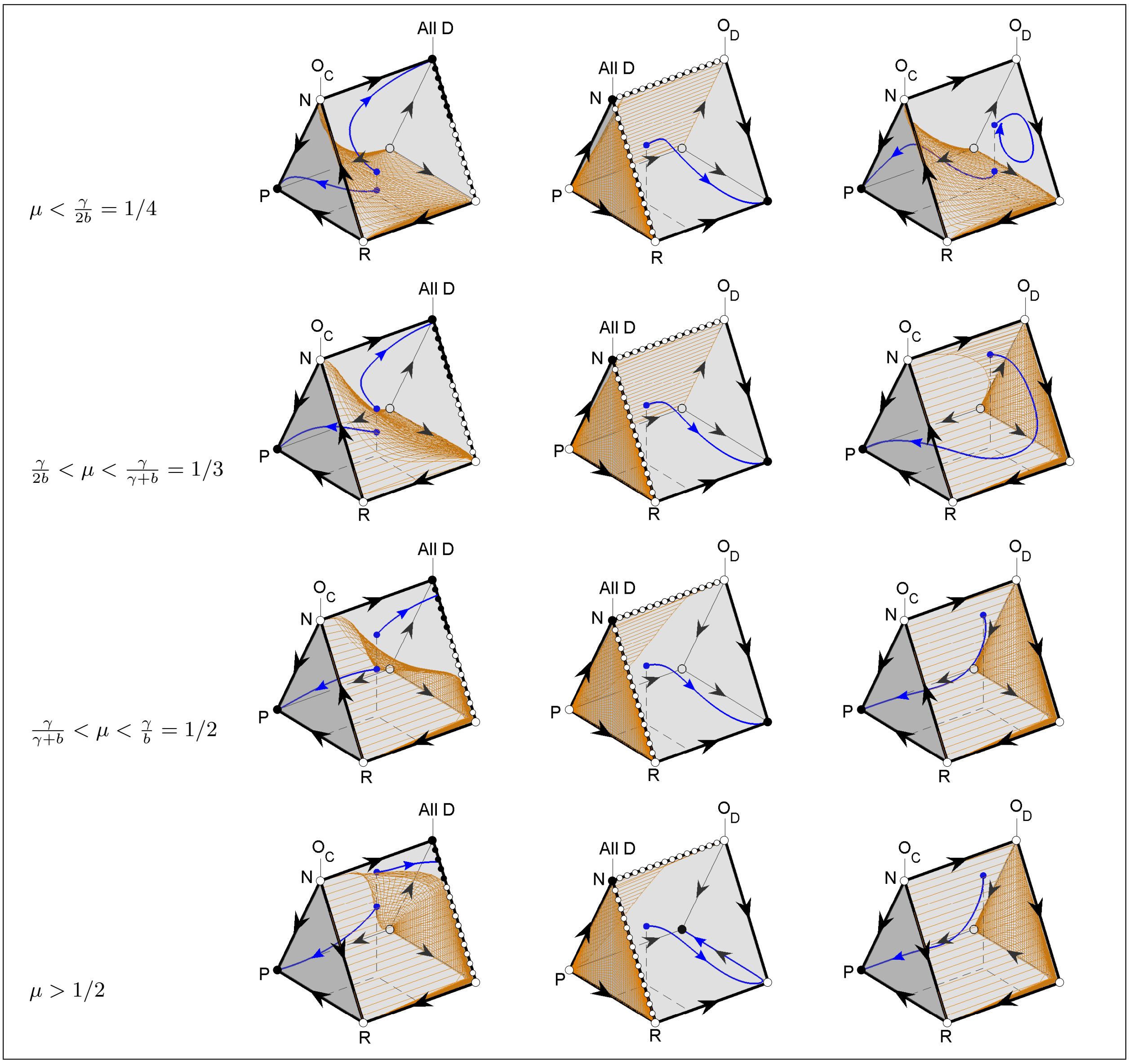


Figure 2: Replicator dynamics for $b = 4,\ c = 1,\ \beta = \gamma = 2$ and $\mu = 15\%, 30\%, 45\%$ resp. $\mu = 60\%$. The arrows on the edges indicate the direction of the dynamics if all other strategies are absent. Black points represent Nash-equilibria, white points indicate unstable fixed points. The orange grid is the manifold that separates initial values with different asymptotic behaviour (i.e., a separatrix). The blue curves represent the typical dynamics for a given initial population.

Note that this choice of parameters implies $\gamma/b = 1/2$; therefore there occur two bifurcations between the third and the fourth row (in the last row $\mathbf{R}$ dominates $\mathbf{N}$ in a homogeneous $\mathbf{O_C}$ population and $\mathbf{P}$ dominates $\mathbf{R}$ in a homogeneous $\mathbf{O_D}$ population). The dynamics in the interior of the prism, however, is the same no matter whether $\gamma/b < 1/2$ or $\gamma/b > 1/2$.

are invariants of motion for $1 \leq k \leq n$ under replicator dynamics, i.e., $\dot{Z}_k = 0$ in the interior of the state space (clearly, $Z_1$ is constant to 1). In particular this holds true for the *Wright manifold*

$$W := \big\{(x_{ij}) : Z_k = 1 \ \text{ for all } \ 2 \leq k \leq n\big\}. \tag{9}$$

Now we modify the replicator dynamics (7) by adding recombination:

$$\dot{x}_{ij} = x_{ij} \cdot \big[m_{ij} - \bar{M}\big] - \rho D_{ij}. \tag{10}$$

Here $\rho > 0$ is the recombination fraction and $D_{ij} := x_{ij} - p_i \cdot q_j$ is the linkage disequilibrium. (In an infinitesimal time interval of length $\Delta t$, the combination $[ij]$ is broken up with probability $(1-\rho)\Delta t$, and formed anew with probability $\rho p_i \cdot q_j \Delta t$.)

**Proposition** (Convergence to the Wright manifold)
Under the modified replicator equation (10), all orbits starting in the interior of the state space converge to the Wright manifold. On this manifold, the modified and the classical replicator dynamics coincide.

*Proof.* We only show $Z_2 \to 1$, all other $Z_k$ follow by a symmetry argument. We have

$$\frac{\dot{Z}_2}{Z_2} = -\rho \cdot \big[D_{11}/x_{11} + D_{22}/x_{22} - D_{12}/x_{12} - D_{21}/x_{21}\big] \tag{11}$$

It is easy to see that

$$\begin{aligned} D_{11}/x_{11} &= 1 - \tfrac{p_1 q_1}{x_{11}} = 1 - \tfrac{(x_{11}+\cdots+x_{1n})(x_{11}+x_{21})}{x_{11}} = \\ &= 1 - x_{11} - \cdots - x_{1n} - x_{21} - \tfrac{x_{12}x_{21}}{x_{11}} - \cdots - \tfrac{x_{1n}x_{21}}{x_{11}} = \\ &= x_{22} + \cdots + x_{2n} - \tfrac{1}{Z_2}x_{22} - \cdots - \tfrac{1}{Z_n}x_{2n} = \\ &= (1 - 1/Z_2)x_{22} + (1 - 1/Z_3)x_{23} + \cdots + (1 - 1/Z_n)x_{2n}, \end{aligned} \tag{12}$$

hence

$$D_{11}/x_{11} = \textstyle\sum_{k=1}^{n}(1 - \frac{Z_1}{Z_k})x_{2k} \tag{13}$$

and analogously

$$\begin{aligned} D_{12}/x_{12} &= \textstyle\sum_{k=1}^{n}(1 - \frac{Z_2}{Z_k})x_{2k} \\ D_{22}/x_{22} &= \textstyle\sum_{k=1}^{n}(1 - \frac{Z_k}{Z_2})x_{1k} \\ D_{21}/x_{21} &= \textstyle\sum_{k=1}^{n}(1 - \frac{Z_k}{Z_1})x_{1k} \end{aligned} \tag{14}$$

Plugging these identities into eq. (11) yields

$$\frac{\dot{Z}_2}{Z_2} = -\rho(Z_2 - 1)[\sum_{k=1}^{n}(\frac{x_{2k}}{Z_k} + \frac{x_{1k}Z_k}{Z_1 Z_2})] = -\rho(Z_2 - 1)[p_1\frac{x_{21}}{x_{11}} + p_2\frac{x_{12}}{x_{22}}]$$

which shows that $Z_2$ (and therefore all $Z_k$) converge to 1 monotonically. In this case it follows from eqs. (13) and (14) that $D_{ij} \to 0$. □

We can employ this result on our original game: As $\mathbf{AllC}$ is weakly dominated by $\mathbf{O_C}$ and $\mathbf{AllD}$ by $\mathbf{O_D}$, we may confine ourselves on the two opportunistic strategies in the first stage. If we consider the options $\mathbf{P}$, $\mathbf{R}$ and $\mathbf{N}$ for the second stage we have a 2x3 role game. Then the previous proposition ensures that all orbits starting in the interior of the state space converge to the generalized Wright manifold where the dynamics coincides with the classical replicator equation.

# 2 Small mutation rates

## 2.1 Transition probabilities

The transition probabilities for the adiabatic case are calculated according to Nowak (2006), see also Sigmund (2010). The following tables $(a_{ij})$ show the fixation probabilities of a mutant $i$, invading a resident strategy $j$ for different values of $\mu$. For low mutation rates, $[\mathbf{O_D}, \mathbf{R}]$ is the only strategy which can invade the asocial $[\mathbf{O_D}, \mathbf{N}]$ with a higher probability than $1/M$ (the neutral fixation probability).
Parameter values: $M = 100, s = 1/10, b = 4, c = 1, \beta = \gamma = 2$.

| $\mu = 10\%$ | $[\mathbf{O_C}, \mathbf{N}]$ | $[\mathbf{O_C}, \mathbf{R}]$ | $[\mathbf{O_C}, \mathbf{P}]$ | $[\mathbf{O_D}, \mathbf{N}]$ | $[\mathbf{O_D}, \mathbf{R}]$ | $[\mathbf{O_D}, \mathbf{P}]$ |
|---|---|---|---|---|---|---|
| $[\mathbf{O_C}, \mathbf{N}]$ | 0.010 | 0.162 | 0.001 | **0.000** | 0.068 | 0.204 |
| $[\mathbf{O_C}, \mathbf{R}]$ | 0.000 | 0.010 | 0.000 | **0.000** | 0.088 | 0.229 |
| $[\mathbf{O_C}, \mathbf{P}]$ | 0.041 | 0.202 | 0.010 | **0.000** | 0.014 | 0.085 |
| $[\mathbf{O_D}, \mathbf{N}]$ | 0.094 | 0.074 | 0.003 | **0.010** | 0.003 | 0.138 |
| $[\mathbf{O_D}, \mathbf{R}]$ | 0.000 | 0.000 | 0.000 | **0.023** | 0.010 | 0.158 |
| $[\mathbf{O_D}, \mathbf{P}]$ | 0.004 | 0.001 | 0.000 | **0.000** | 0.000 | 0.010 |

| $\mu = 30\%$ | $[\mathbf{O_C}, \mathbf{N}]$ | $[\mathbf{O_C}, \mathbf{R}]$ | $[\mathbf{O_C}, \mathbf{P}]$ | $[\mathbf{O_D}, \mathbf{N}]$ | $[\mathbf{O_D}, \mathbf{R}]$ | $[\mathbf{O_D}, \mathbf{P}]$ |
|---|---|---|---|---|---|---|
| $[\mathbf{O_C}, \mathbf{N}]$ | 0.010 | 0.082 | 0.000 | **0.000** | 0.015 | 0.001 |
| $[\mathbf{O_C}, \mathbf{R}]$ | 0.000 | 0.010 | 0.000 | **0.005** | 0.069 | 0.078 |
| $[\mathbf{O_C}, \mathbf{P}]$ | 0.120 | 0.202 | 0.010 | **0.005** | 0.036 | 0.066 |
| $[\mathbf{O_D}, \mathbf{N}]$ | 0.073 | 0.018 | 0.000 | **0.010** | 0.002 | 0.022 |
| $[\mathbf{O_D}, \mathbf{R}]$ | 0.006 | 0.000 | 0.000 | **0.060** | 0.010 | 0.078 |
| $[\mathbf{O_D}, \mathbf{P}]$ | 0.177 | 0.049 | 0.000 | **0.004** | 0.000 | 0.010 |

| $\mu = 50\%$ | $[\mathbf{O_C}, \mathbf{N}]$ | $[\mathbf{O_C}, \mathbf{R}]$ | $[\mathbf{O_C}, \mathbf{P}]$ | $[\mathbf{O_D}, \mathbf{N}]$ | $[\mathbf{O_D}, \mathbf{R}]$ | $[\mathbf{O_D}, \mathbf{P}]$ |
|---|---|---|---|---|---|---|
| $[\mathbf{O_C}, \mathbf{N}]$ | 0.010 | 0.011 | 0.000 | **0.000** | 0.000 | 0.000 |
| $[\mathbf{O_C}, \mathbf{R}]$ | 0.009 | 0.010 | 0.000 | **0.048** | 0.049 | 0.001 |
| $[\mathbf{O_C}, \mathbf{P}]$ | 0.201 | 0.202 | 0.010 | **0.071** | 0.072 | 0.047 |
| $[\mathbf{O_D}, \mathbf{N}]$ | 0.052 | 0.001 | 0.000 | **0.010** | 0.000 | 0.000 |
| $[\mathbf{O_D}, \mathbf{R}]$ | 0.051 | 0.000 | 0.000 | **0.100** | 0.010 | 0.009 |
| $[\mathbf{O_D}, \mathbf{P}]$ | 0.246 | 0.137 | 0.001 | **0.102** | 0.011 | 0.010 |

### 2.2 Typical transitions between states

For different information levels, the following table displays the distribution of strategies which succeed in invading a homogeneous $[\mathbf{O_D}, \mathbf{N}]$ population. For small information levels especially, $[\mathbf{O_D}, \mathbf{R}]$ is best at overcoming this asocial state. The table shows the result of a simulation of the stochastic process with $10^7$ mutations, population size $M = 100$, selection strength $s = 1/10, b = 4, c = 1, \beta = \gamma = 2$.

| | $[\mathbf{O_C}, \mathbf{N}]$ | $[\mathbf{O_C}, \mathbf{R}]$ | $[\mathbf{O_C}, \mathbf{P}]$ | $[\mathbf{O_D}, \mathbf{N}]$ | $[\mathbf{O_D}, \mathbf{R}]$ | $[\mathbf{O_D}, \mathbf{P}]$ |
|---|---|---|---|---|---|---|
| $\mu = 0\%$ | 0.000 | 0.001 | 0.000 | – | **0.999** | 0.000 |
| $\mu = 10\%$ | 0.000 | 0.003 | 0.001 | – | **0.996** | 0.000 |
| $\mu = 20\%$ | 0.000 | 0.017 | 0.011 | – | **0.970** | 0.002 |
| $\mu = 30\%$ | 0.000 | 0.068 | 0.040 | – | **0.858** | 0.034 |
| $\mu = 40\%$ | 0.000 | 0.191 | 0.191 | – | **0.381** | 0.238 |

The next table displays the distribution of strategies which immediately preceded a homogeneous $[\mathbf{O_C}, \mathbf{P}]$ populations. If there is little information about co-players, the state of punishing cooperators is mostly reached via rewards (Simulation with $10^7$ mutations, $M = 100, s = 1/10, b = 4, c = 1, \beta = \gamma = 2$).

| | $[\mathbf{O_C}, \mathbf{N}]$ | $[\mathbf{O_C}, \mathbf{R}]$ | $[\mathbf{O_C}, \mathbf{P}]$ | $[\mathbf{O_D}, \mathbf{N}]$ | $[\mathbf{O_D}, \mathbf{R}]$ | $[\mathbf{O_D}, \mathbf{P}]$ |
|---|---|---|---|---|---|---|
| $\mu = 0\%$ | 0.228 | **0.644** | – | 0.000 | 0.128 | 0.000 |
| $\mu = 10\%$ | 0.394 | **0.457** | – | 0.002 | 0.144 | 0.004 |
| $\mu = 20\%$ | 0.234 | **0.479** | – | 0.014 | 0.212 | 0.062 |
| $\mu = 30\%$ | 0.098 | **0.419** | – | 0.018 | 0.211 | 0.254 |
| $\mu = 40\%$ | 0.014 | 0.169 | – | 0.008 | 0.205 | **0.604** |

### 2.3 Cheap incentives ($\gamma < \beta$)

In the main article we have usually assumed $\gamma = \beta$. The following figure shows the case of "cheap" incentives, $\gamma < \beta$. Note that the qualitative behavior remains unchanged; there is only a marginal difference to the case with $\gamma = \beta$ which is depicted in the right column of Fig. 3 in the main manuscript.

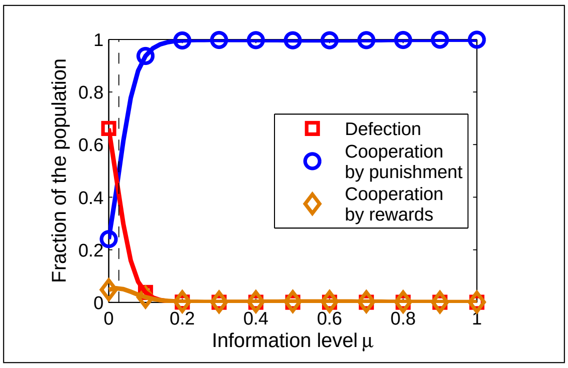


Figure 3: Stationary distribution in the case of "cheap" incentives. If $\gamma < \beta$, a homogeneous population of rewarding cooperators would obtain the maximum payoff. But evolution still leads to the stick instead of the carrot.
Parameter values: Population size $M = 100$, selection strength $s = 1/10$; $b = 4, c = 1, \beta = 2, \gamma = 3/4$.

## References

Gaunersdorfer, A., Hofbauer, J. & Sigmund, K. 1991 *On the Dynamics of Asymmetric Games.* Theor. Pop. Biol. **39:** 345-357.

Hofbauer, J. & Sigmund, K. 1998 *Evolutionary Games and Population Dynamics.* Cambridge UK: Cambridge University Press.

Nowak, M. A. 2006 *Evolutionary dynamics.* Cambridge MA: Harvard University Press.

Sigmund, K. 2010 *The calculus of selfishness.* Princeton NJ: Princeton University Press.