

# Retaliation and the role for punishment in the evolution of cooperation

Irenaeus Wolff\*

Thurgau Institute of Economics (TWI), University of Konstanz, Hauptstrasse 90, 8280 Kreuzlingen, Switzerland

---

## H I G H L I G H T S

- ▶ We allow for retaliation in a social dilemma game with punishment.
- ▶ Retaliating behaviour evolves endogenously in the model.
- ▶ Punishers may induce a mixed, partially cooperative equilibrium without punishers.
- ▶ Under positive mutation rates, punishment can increase the level of cooperation.
- ▶ When mutations occur, punishment cannot stabilise full cooperation.

---

## A B S T R A C T

Models of evolutionary game theory have shown that punishment may be an adaptive behaviour in environments characterised by a social dilemma situation. Experimental evidence closely corresponds to this finding but questions the cooperation enhancing effect of punishment if players are allowed to retaliate against their punishers. This study provides a theoretical explanation for the existence of retaliating behaviour in the context of repeated social dilemmas and analyses the role punishment can play in the evolution of cooperation under these conditions. We show a punishing strategy can pave the way for a partially cooperative equilibrium of conditional cooperators and defecting types and, under positive mutation rates, foster the cooperation level in this equilibrium by prompting reluctant cooperators to cooperate. However, when rare mutations occur, it cannot sustain cooperation by itself as punishment costs favour the spread of non punishing cooperators.

*Keywords:*  
Public goods  
Prisoner's dilemma  
Strong reciprocity  
Counter-punishment

## 1. Introduction

Recent laboratory experiments have cast serious doubt on the scope of earlier findings suggesting that punishment is a suitable solution for social dilemma situations. In particular, the very cooperation enhancing effect has been challenged: by relaxing the restriction to a single punishment stage and allowing for retaliation, Denant Boemont et al. (2007) and Nikiforakis (2008) demonstrate how sensitive the cooperative outcome of earlier studies is to changes in the experimental setup (e.g., Yamagishi, 1986 or Fehr and Gächter, 2000).<sup>1</sup> While the (exogenous) restriction to a single stage of punishment does not seem to be plausible in most real world situations, existing (evolutionary) models of cooperation have comfortably rested on this assumption. The present paper makes a first step in addressing the challenge posed

by the experimental results mentioned. In our model, we provide a possible explanation for why a retaliative strategy may be adaptive within the context of an iterated social dilemma game: if punishers stop sanctioning free riding behaviour after retaliation and conditional cooperators do not require full cooperation to maintain their cooperative attitude, retaliating defectors can reap the benefits from conditional cooperators' cooperation free of punishment without having to cooperate themselves. Within this framework, we show that punishing cooperators still can play two roles: they can break up an equilibrium of omnilateral defection and pave the way for a stable polymorphic equilibrium composed of conditional cooperators and a mix of defecting types; and they can increase the level of cooperation within a polymorphic equilibrium if mutation plays a role. What punishment cannot do, in the confines of our framework, is to maintain a full cooperation equilibrium if there is even a tiny mutation rate.

In the past, a variety of mechanisms have been proposed to account for the high level of cooperation among humans, such as kinship (Hamilton, 1964), reciprocal altruism (Trivers, 1971 or Axelrod and Hamilton, 1981), costly signalling (Zahavi, 1975 or Gintis et al., 2001), indirect reciprocity and reputation

---

\*Tel.: +41 716770515.

E-mail address: [wolff@twi-kreuzlingen.ch](mailto:wolff@twi-kreuzlingen.ch)

<sup>1</sup> For other studies looking at retaliation over several rounds of a social-dilemma game with a single punishment stage, cf., e.g., Dreber et al. (2008) or Herrmann et al. (2008).

(Alexander, 1979, 1987; Nowak and Sigmund, 1998, or Ohtsuki et al., 2009), 'culture' (Cavalli Sforza and Feldman, 1981 or Boyd and Richerson, 1985), norms (Sugden, 1986 or Sethi, 1996), group selection (Sober and Wilson, 1998), and strong reciprocity (Gintis, 2000 or Gintis et al., 2003).<sup>2</sup> The concept of strong reciprocity has received a lot of attention, as it is unique among the explanations proposed in the literature for the evolution of cooperation in that it can account for the substantial degree of cooperation observed in anonymous laboratory settings even when there is no repeated interaction between the same players. While the concept may have different readings, the general idea is that of a behavioural disposition "to sacrifice resources to bestow benefits on those who have bestowed benefits" and "to sacrifice resources to punish those who are not bestowing benefits in accordance with some social norm".<sup>3</sup> In a prisoner's dilemma or a public good game like the ones used in the aforementioned models on the evolution of cooperation, this translates into a cooperative action in settings without repeated interaction as well as with repeated interaction, unless in the latter case if the other player(s) is (are) in bad standing. Additionally, strong reciprocity stipulates the punishment of defectors whenever there are punishment stages and regardless of players' anonymity.

It has been shown time and again that unconditional cooperation cannot be stabilised easily in such settings; at the same time, costly punishment can be evolutionarily stable and thus support cooperation in societies even when there is no inter group conflict (Henrich and Boyd, 2001; Bowles and Gintis, 2004; Carpenter et al., 2004; Carpenter, 2007). Recent contributions have gone on to show that strong reciprocators can even invade a non cooperative society under certain conditions (Fowler, 2005; Hauert et al., 2007; Mathew and Boyd, 2009; Rand and Nowak, 2011, object, rejoined by García and Traulsen, 2012). However, virtually none of these studies has taken into account that in many real life situations, punished players will have the opportunity to get back at their punishers. If at all, models of cooperation have included additional punishment stages reserved to sanction enforcement (Henrich and Boyd, 2001).<sup>4</sup> The exception to the rule comes as a couple of papers proposed by Janssen and Bushman (2008) and Wolff (2009) who simply assume retaliative behaviour into existence, relying on the experimental evidence to justify this assumption. However, in light of the fact that they do not explain why retaliation may arise in the first place (and that it cannot arise in the situations they study), it remains unclear what we can learn from these studies about those situations in which retaliative behaviour may be adaptive. Rand et al. (2009) focus on whether replacing the defecting response in tit for tat by defection plus punishment increases cooperation. While retaliative behaviour is allowed to evolve within their model and does so to a certain degree in some of their simulations, it remains unclear what types do retaliate and whether and how often the according strategy part is effective in the sense of being triggered. The present paper aims to be clearer in these respects. Also, our focus lies on situations where punishment does not imply non cooperation.

In a recent contribution, Rand et al. (2010) remove the restriction that punishment be targeted at defectors to question

the punishment based explanation to cooperation altogether. However, they do so by showing that, under a certain form of learning, punishing cooperators can be invaded by cooperators who punish everybody else unconditionally or conditionally on making positive contributions as these invaders obtain a locally higher payoff than those they punish, and behaviour is copied on the basis of local payoff comparisons. In their model, only defection coupled with unconditional punishment of everybody else is evolutionarily stable. However, in evolutionary terms, their model suggests living in communities is not adaptive, as there are no gains from cooperation and agents merely destroy their own resources to destroy even more of others' resources. Note that it is the mere *possibility* of punishing others which in most relevant situations of life in groups will be given rather than the use of punishment to ensure cooperation that leads to the gloomy conclusion. The present paper proposes a different argument, suggesting that retaliation best be considered in the framework of repeated interactions<sup>5</sup>: retaliation may have a purpose in the pursuit of resources. If group members are generally willing to cooperate even if a small fraction of the group defects, and if punishers cease to punish when retaliated against, retaliators may have an evolutionary advantage when interactions are long enough.

Addressing the question in a repeated interaction perspective is not without its own problems. Especially when interactions take place in a group such as in the typical public good case the number of possible strategies is far too high for a comprehensive analysis.<sup>6</sup> Hence, we either have to drop the repeated game frame or restrict the number of admissible strategies. Given human interactions are often repeated and repeated game strategies such as tit for tat seem to be important, we choose the second avenue, restricting the set of included strategies first and performing a check of sensitivity to the restriction afterwards.<sup>7</sup> In restricting the number of strategies to a tractable number, we choose stylisations of strategies we encounter in laboratory experiments, as discussed in Section 2.1.<sup>8</sup> In this sense, our model says something about the interaction of those strategies we observe in the laboratory and why we may not observe other strategies; the evolution of the former when the latter would have been a possibility, however, is outside the scope of the presented model.

## 2. The game

At each moment in time, a large number of groups each consisting of  $N$  individuals are randomly drawn from a very large population. Groups remain constant for the duration of an indefinitely repeated game with continuation probability  $\pi$ . The stage game consists of three stages. In stage 0, the contribution stage, agents play a symmetric  $N$  person binary public good game. A cooperating player incurs a net cost of  $c$  to convey a benefit of  $b/(N-1)$  to every other member of the group, where  $0 < c < b$ , while a defecting player does not incur any cost nor

<sup>2</sup> The studies mentioned are only meant to indicate important early contributions. For a more detailed picture of the literature, the interested reader is kindly referred to, e.g., the works collected in Hammerstein (2003).

<sup>3</sup> Fehr and Henrich (2003, p. 57); for a discussion of the different readings of strong reciprocity depending on whether this is seen as a *behaviour* (or behavioural algorithm) or a *strategy*, cf. Stephens (2005).

<sup>4</sup> Recent laboratory evidence suggests additional punishment stages are used for additional punishment assignments and retaliation, but rarely for sanction enforcement, e.g., Nicklisch and Wolff (2011).

<sup>5</sup> Important articles within this framework are Fudenberg and Maskin (1986, 1990) and Binmore and Samuelson (1992). I thank an anonymous reviewer for pointing this out.

<sup>6</sup> Even without repetition, a game with two punishment stages already leads to a minimum of 16 possible strategies, as pointed out by an anonymous referee: (contribute or defect)  $\times$  (punish or not, contingent on the contribution decision)  $\times$  (retaliate or not).

<sup>7</sup> For a similar approach, see, e.g., Boyd and Richerson (1992); for the importance of tit-for-tat-like strategies, cf., e.g., Rand et al. (2009) and references cited therein.

<sup>8</sup> As another caveat, we restrict attention to those experiments in 'Western' countries, as we are lacking appropriate data for other regions.

convey any benefit. In the second stage, players may punish each other, incurring a cost of  $k$  and causing a damage of  $p$  for the player punished. In the third stage, finally, punished players may retaliate with the same  $k : p$  punishment technology. Reproduction, or learning, takes place between interactions and follows the standard discrete time replicator dynamics (Taylor and Jonker, 1978), and therefore, the fraction  $x_i$  of each agent type  $i$  evolves according to

$$x_i(\tau+1) = \frac{u_0 + u_i(\mathbf{x}(\tau)) \cdot x_i(\tau)}{u_0 + \sum_{\forall j} u_j(\mathbf{x}(\tau)) \cdot x_j(\tau)}, \quad (1)$$

where  $u_0$  is each agent's baseline fitness and  $u_i(\mathbf{x}(\tau))$  is an  $i$  agent's average additional fitness, or payoff, from the interaction given the type distribution in the population,  $\mathbf{x}(\tau)$ , at time  $\tau$ .

### 2.1. Strategies under consideration

Clearly, there is an infinite number of possible strategies and therefore, of potential agent types in a public good game that is repeated infinitely (albeit with a given ending probability), and that allows both for punishment of other agents after each contribution stage and for retaliation against such punishment.<sup>9</sup> Hence, we need to restrict our attention to a subset of all possible strategies. The following list presents a reduced set of agent type families, from which we will subsequently 'recruit' the four types making up our model population.

- $D$  never cooperates, never punishes, and does not react to punishment. The traditional defector type;
- $R$  never cooperates, always retaliates. The  $R$  type is a direct implementation of the idea that punished defectors may retaliate, as ample experimental evidence suggests many do (e.g., Nikiforakis, 2008);
- $T_n$  a nice tit for tat type who always cooperates if at least  $n$  other players cooperated in the preceding stage game, and defects, otherwise.  $T_n$ 's never punish and revert to cooperation when punished.<sup>10</sup>  $T_n$  types were introduced by Joshi (1987) and Boyd and Richerson (1988);
- $P$  always contributes and always punishes any non contributing player. Early treatments of punishment behaviour date back at least to Axelrod (1986);
- $DT_n$  starts by defecting, switching to  $T_n$  only once punished.  $DT_n$  types are straightforward generalisations of the *reluctant cooperator* type introduced by Boyd and Richerson (1992);
- $PT_n$  acts like a  $P$  type unless retaliated against; then immediately switches to behaving like a  $T_n$  type (if retaliated against in stage game  $\tau$ , the next move is conditioned on others' behaviour in stage game  $\tau$ , too);  $PT_n$  types (i) may be adaptive in the presence of  $R$ 's and more adaptive than  $P$ 's, as argued below and (ii) may be the reason for retaliating behaviour to appear, in the first place. In the spirit of Boyd and Richerson's (1992) model, they also may be seen as a kind of 'semi timid' punisher type;

$PD$  acts like a  $P$  type unless retaliated against; then immediately switches to behaving like a  $D$  type; the discussion of  $PT_n$ 's also applies to  $PD$ 's.

### 2.2. Additional assumptions and type choice

In the following, we will make two additional assumptions. First, restrict the attention to interaction groups of  $N=3$  individuals, to simplify the game enough to draw the main conclusions. Second, assume that  $((N-2)/(N-1))b > c$ , meaning that cooperating types who cooperate in the presence of a single defecting player still derive a strictly positive benefit from the interaction. This may be seen as a strong assumption given the restriction to groups of 3; however, there is no obvious reason why the main results presented in this paper should not carry over qualitatively to larger groups, in which case our second assumption seems no longer unrealistic. As an indication, we sketch a model variant with large  $N$  in Appendix B.

Similar to Boyd and Richerson (1992), we disregard pure defectors in favour of other types that start free riding but subsequently react to punishment in either of two ways: by cooperating in the next stage game or by retaliating. In terms of the type taxonomy above, we replace  $D$ 's by  $DT_n$ 's, on the one hand, and  $R$  types, on the other.<sup>11</sup> Furthermore, we discard the  $P$  type as it is dominated by  $PT_0$ , as well as the unconditionally cooperating (if only after punishment)  $T_0$  and  $DT_0$  which are dominated by  $T_1$  and  $DT_1$ , respectively.<sup>12</sup> We abstract from  $PT_0$ 's for a similar reason, even though there is a specific matching in which the  $PT_0$  does better than a  $PT_1$ , namely when matched with an  $DT_1$  and an  $R$  agent; in all other matchings, the  $PT_0$  does only equally well or worse. Finally, we can safely disregard the  $PD$  type as it is behaviourally equivalent to the  $PT_{N-1}$  type under the strategy set considered.

We are now left with seven agent types,  $R$ 's and three type pairs involving conditional cooperation somewhere in their strategy specification:  $T_1$  and  $T_2$ ,  $DT_1$  and  $DT_2$ , and  $PT_1$  and  $PT_2$ . To arrive, finally, at a tractable strategy set, let us eliminate three of the latter, one of each pair. For  $DT_n$ 's, eliminating  $DT_2$  seems like the natural choice, for a similar argument as presented for  $PT_0$ 's above: under many matchings,  $DT_1$  and  $DT_2$  do equally well; under a number of matchings,  $DT_1$  does better; and only when matched with a  $PT_2$  and an  $R$  agent, the  $DT_2$  does better, while in a group with a  $PT_1$  and an  $R$  agent, either type may do better, depending on the parameter choices (and the resolution of the coordination problem between the  $PT_1$  and the  $DT_1$ ). A sensible choice between the  $T_n$  and the  $PT_n$  types, respectively, is less obvious. The early works of Joshi (1987) and Boyd and Richerson (1988) have shown that in a world of only  $T_n$  types competing with  $D$ 's, the only stable conditional cooperator type is  $T_{N-1}$  (given all other types can be exploited by a single defector). However, both go on to show that the conditions for  $T_{N-1}$  to invade a defecting population are extreme. Two characteristics of the present model suggest different aspects may be important in our case: most importantly, we incorporate punishment and

<sup>9</sup> We do not include strategies using antisocial punishment as studies disentangling antisocial punishment and retaliation by design do not find evidence of 'unprovoked' antisocial punishment (e.g., Nicklisch and Wolff, 2011). Other studies such as Cheung (2012) and Cinyabuguma et al. (2006) also tend to attribute most antisocial punishment to what may be called "pre-emptive retaliation" (Cheung, 2012), while Herrmann et al. (2008) merely report received previous-round punishment to be a very strong predictor of antisocial punishment. Unfortunately, there has not been a study run with a design similar to Nicklisch and Wolff (2011) in any of the places known to exhibit high rates of antisocial punishment.

<sup>10</sup> The reaction to punishment was added for completeness as it is not part of the original models; it will be irrelevant in our main model.

<sup>11</sup> Boyd and Richerson (1992, p. 177) justify this choice in two ways, by an "abundant empirical evidence that organisms do respond to punishment," and pointing out that this merely is a best case for the evolution of cooperation. We briefly re-introduce pure defectors in Section 2.5 on equilibrium stability with respect to excluded types.

<sup>12</sup> Strictly speaking, all of the above dominance relationships referred to are weak, and hence, the corresponding strategies need not be driven to extinction. However, as long as we are focusing on interior solution paths, i.e., punishing respectively retaliating respectively free-riding types still occur,  $D$ 's,  $P$ 's,  $T_0$ 's, and  $DT_0$ 's are selected against more strongly than the corresponding  $DT_{N-1}$ ,  $PT_0$ ,  $T_1$ , and  $DT_1$ -types.

defector types may react to being sanctioned; second, we have posited above that there is a mutual benefit from cooperation even if only a subset of ( $N - 1$ ) players cooperate. The latter assumption did not play a role in the earlier studies, as defector types would still be better off; in the presence of punishing types, however, this may change, and the benefits forgone by  $T_{N-1}$ 's due to their highly demanding cooperation requirements may now play an important role. Finally, if we were to include errors which we refrain from for tractability  $T_{N-1}$ 's would always be invaded by  $T_{N-2}$ 's, for much the same reason as Tit for Tat players can be invaded in the two player case.<sup>13</sup> For the above reasons, this paper will focus on the less demanding  $T_{N-2}$  and  $PT_{N-2}$  (i.e.,  $T_1$  and  $PT_1$ ) types. In Section 2.5, we demonstrate that the  $(P)T_2$  types cannot invade any of the equilibria derived in the following section. To recapitulate, the agent types considered in the following model are the punishing cooperator  $PT_1$ , the conditional cooperator  $T_1$ , the reluctant cooperator  $DT_1$ , and the retaliating defector  $R$ . We will represent population states as  $\mathbf{x} = (x_p, x_t, x_d, x_r)$ , where indices  $p, t, d,$  and  $r$  indicate  $PT_1, T_1, DT_1,$  and  $R$  types, respectively.

### 2.3. Payoffs and equilibria

Having specified the game and the strategies to be considered, we next derive the expected payoffs corresponding to each strategy. Using Table 1, it is tedious but straightforward to determine the following payoff equations:

$$\begin{aligned} u(R, \mathbf{x}) &= b \left( x_t + x_p + \frac{\pi}{1-\pi} [(x_t + x_p)^2 + x_p x_d] \right) - 2(k+p)x_p, \\ u(DT_1, \mathbf{x}) &= b \left( x_t + x_p + \frac{\pi}{1-\pi} \left[ (x_t + x_p)^2 + x_p(2x_d + \frac{\pi}{1+\pi}x_r) \right] \right) - 2px_p \\ &\quad - \frac{\pi c}{1-\pi} \left( 2x_p - x_p^2 - \frac{\pi}{1+\pi} 2x_p x_r \right), \\ u(T_1, \mathbf{x}) &= \frac{b}{1-\pi} (x_t + x_p + \pi x_d x_p) - \frac{c}{1-\pi} (1 - \pi [1 - x_t(2 - x_t) \\ &\quad - x_p(2 - x_p) + 2x_p x_t]), \\ u(PT_1, \mathbf{x}) &= \frac{b}{1-\pi} \left( x_t + x_p + \pi \left[ x_d \left( x_t + x_p + x_d + \frac{x_r}{1+\pi} \right) \right] \right) - 2kx_d - 2(k+p)x_r \\ &\quad - \frac{c}{1-\pi} \left( 1 - \pi \left[ 1 - x_t(2 - x_t) - x_p(2 - x_p) + 2x_p x_t - x_d \left( x_d + 2 \frac{\pi}{1+\pi} x_r \right) \right] \right), \end{aligned} \quad (2)$$

where the time index  $\tau$  is dropped for legibility.

Focusing on monomorphic equilibria, we immediately see that none of the four strategies selected is evolutionarily stable:  $R$ 's can always be invaded by neutral drift by  $DT_1$ 's and *vice versa*; the same holds true for  $PT_1$ 's and  $T_1$ 's, respectively. Additionally, in the absence of punisher types,  $T_1$ 's can be invaded by either defecting type. At the same time,  $T_1$ 's can invade any of the defecting types as well as any mix thereof, as long as the continuation probability  $\pi$  is sufficiently close to unity; the lower the  $\pi$ , the higher the critical fraction of  $T_1$ 's needed to invade until invasion becomes impossible altogether. The defecting types' invasion barrier against  $T_1$ 's is given by

$$\hat{x}_t^{crit} = \frac{b - 2c \sqrt{b^2 - 4c[b - c]/\pi}}{2(b - c)}, \quad (3)$$

which exists only if the continuation probability  $\pi$  and the public good's benefit to cost ratio are sufficiently large.<sup>14</sup> Obviously, the ability of the conditional cooperator type to invade a defecting

population stems from the former's ability to concentrate the benefits of future cooperation predominantly on their likes, which can only pay if future interactions are sufficiently likely. On the other hand,  $PT_1$ 's cannot invade  $R$ 's while they can invade  $DT_1$ 's, as long as their investment into establishing cooperation within an interaction is outweighed by the expected long term benefits:

$$c + 2k < \pi \frac{b - c}{1 - \pi}. \quad (4)$$

The following proposition establishes an interesting result that is counterintuitive only at first sight.

**Proposition 1.** *If the continuation probability  $\pi$  exceeds  $\pi^{max} = (2p + 2k - c)/(2p + 2k)$ , neither cooperative strategy is stable against invasion by some defecting type.*

Proof is given in Appendix A.

The intuition for why a high continuation probability  $\pi$  allows retaliators to invade punishing cooperators is analogous to why punishers can invade a population of reluctant cooperators. Either type incurs punishment costs in the first stage game to reap future benefits: the punisher by inducing cooperation and the retaliator by curtailing being sanctioned. Conversely, punishers resist invasion by retaliators if and only if conforming to the prevailing cooperative behaviour in all interaction stage games is less costly than incurring the one time cost of being punished and retaliating; that is, if

$$\frac{c}{1 - \pi} < 2(p + k). \quad (5)$$

Finally,  $DT_1$ 's cannot invade a population of all  $PT_1$ 's as long as  $c < 2p$ .

This is straightforward: after the first stage game, punishers have induced  $DT_1$ 's to contribute, so there are no payoff differences from the second stage game on; further, if contributing is less costly than being punished by all other players, then it pays not to try free riding in the first stage game when all other players are punishers. Note that condition (6) may imply condition (5) if  $\pi < k/(k+p)$ , and *vice versa*, otherwise. This simply speaks of the fact that, when a second encounter is sufficiently unlikely, retaliating does not pay, as the probability of future gains from unpunished free riding are close to nil. In the remainder of this paper, we will concentrate on future encounters being sufficiently likely, so that (5) implies (6).

**Proposition 2.** *Define  $x_p^* \equiv \max\{c/[(1-\pi)(2k+2p)], c/2p\}$  and assume  $c < 2p$ . Then, the subset of points in the  $PT_1 - T_1$  continuum that are stable against invasion by both defecting types is nonempty and given by  $\mathbf{x}_c^{NE} = \{(1 - \theta(1 - x_p^*), \theta(1 - x_p^*), 0, 0) | \theta \in [0, 1)\}$ , if and only if*

$$\pi \begin{cases} < \frac{2p+2k-c}{2p+2k}, & b \leq \frac{2c}{\pi}, \text{ and} \\ \leq \frac{2p+2k-c}{2p+2k} & \text{otherwise.} \end{cases} \quad (7)$$

Proof is given in Appendix A.

Proposition 2 describes the set of 'fully cooperative' equilibria and delineates the conditions for their existence. As expected, the prospects for cooperation are better the lower the net contribution costs  $c$  and the stronger the impact of a punishment action. At the same time, punishment costs are negligible as selection acts on a global level and punishment actions are rare close to any 'fully cooperative' equilibria. Evidently, the 'cooperative'

<sup>13</sup> See Nowak and Sigmund (1993).

<sup>14</sup> For similar results, cf. Joshi (1987) and Boyd and Richerson (1988).

**Table 1**  
Matching table.

$i_1$	$i_2$	$i_3$	$prob(i_2, i_3   i_1)$	$u(i_1   i_2, i_3)$	$u(i_2   i_1, i_3)$	$u(i_3   i_1, i_2)$
$T_1$	$T_1$	$T_1$	$x_t^2$	$\frac{b-c}{1-\pi}$		
$T_1$	$T_1$	$R$	$2x_t x_r$	$\frac{b/2-c}{1-\pi}$		$\frac{b}{1-\pi}$
$T_1$	$T_1$	$DT_1$	$2x_t x_d$	$\frac{b/2-c}{1-\pi}$		$\frac{b}{1-\pi}$
$T_1$	$T_1$	$PT_1$	$2x_t x_p$	$\frac{b-c}{1-\pi}$		$\frac{b-c}{1-\pi}$
$T_1$	$R$	$R$	$x_r^2$	$-c$	$\frac{b}{2}$	
$T_1$	$R$	$DT_1$	$2x_r x_d$	$-c$	$\frac{b}{2}$	$\frac{b}{2}$
$T_1$	$R$	$PT_1$	$2x_r x_p$	$\frac{b/2-c}{1-\pi}$	$b-p-k+\pi\frac{b/2}{1-\pi}$	$\frac{b}{2}-c-k-p+\pi\frac{b/2-c}{1-\pi}$
$T_1$	$DT_1$	$DT_1$	$x_d^2$	$-c$	$\frac{b}{2}$	
$T_1$	$DT_1$	$PT_1$	$2x_d x_p$	$\frac{b}{2}-c+\pi\frac{b-c}{1-\pi}$	$b-p+\pi\frac{b-c}{1-\pi}$	$\frac{b}{2}-c-k+\pi\frac{b-c}{1-\pi}$
$T_1$	$PT_1$	$PT_1$	$x_p^2$	$\frac{b-c}{1-\pi}$	$\frac{b-c}{1-\pi}$	
$R$	$R$	$R$	$x_r^2$	$0$		
$R$	$R$	$DT_1$	$2x_r x_d$	$0$		$0$
$R$	$R$	$PT_1$	$2x_r x_p$	$\frac{b}{2}-p-k$		$-c-2k-2p$
$R$	$DT_1$	$DT_1$	$x_d^2$	$0$	$0$	
$R$	$DT_1$	$PT_1$	$2x_d x_p$	$\frac{b}{2}-p-k+\pi\frac{b/2}{1-\pi}$	$\frac{b}{2}-p+\pi\frac{\pi b/2-c}{1-\pi^2}$	$-c-2k-p+\pi\frac{b/2-\pi c}{1-\pi^2}$
$R$	$PT_1$	$PT_1$	$x_p^2$	$b-2p-2k+\pi\frac{b}{1-\pi}$	$\frac{b}{2}-p-k+\pi\frac{b/2-c}{1-\pi}$	
$DT_1$	$DT_1$	$DT_1$	$x_d^2$	$0$		
$DT_1$	$DT_1$	$PT_1$	$2x_d x_p$	$\frac{b}{2}-p+\pi\frac{b-c}{1-\pi}$		$-c-2k+\pi\frac{b-c}{1-\pi}$
$DT_1$	$PT_1$	$PT_1$	$x_p^2$	$b-2p+\pi\frac{b-c}{1-\pi}$	$\frac{b}{2}-c-k+\pi\frac{b-c}{1-\pi}$	
$PT_1$	$PT_1$	$PT_1$	$x_p^2$	$\frac{b-c}{1-\pi}$		

Note: Probabilities with  $i_2$  and  $i_3$  being focal are omitted for brevity. Payoffs for  $i_2$  (and  $i_3$ ) are only given where different from  $i_1$  ( $i_2$ ).

equilibrium component presented in Proposition 2 is not the only set of (neutrally) stable states: there is also a ‘defecting component’  $\mathbf{x}_D^{NE} = \{(0, 0, \delta\bar{x}_d, 1 - \delta\bar{x}_d) | \delta \in [0, 1]\}$ , where  $\bar{x}_d$  is the highest fraction of  $DT_1$ ’s under which  $PT_1$ ’s do not obtain a positive payoff, and therefore, cannot invade the mix of retaliators and reluctant conditional cooperators.<sup>15</sup>

Unless the continuation probability is close to unity (and hence,  $T_1$ ’s could invade defectors), both components are also dynamically stable in the sense that a small perturbation away from any point but the end points of these components results in the system reverting to a point on the same component that is not too far away. The reason for this stability is that near the defecting component  $\mathbf{x}_D^{NE}$ , retaliation is not costly because there are so few punishers (i.e.,  $u(R, \mathbf{x}) \approx u(DT_1, \mathbf{x})$ ) while punishment is because of the large fraction of retaliators (i.e.,  $u(PT_1, \mathbf{x}) < u(R, \mathbf{x})$ ;  $T_1$ ’s cannot invade, anyway); near the cooperative component  $\mathbf{x}_C^{NE}$ , little defection happens, which means being a punisher is not costly (i.e.,  $u(PT_1, \mathbf{x}) \approx u(T_1, \mathbf{x})$ ), while defection is costly due to the high fraction of punishers (i.e.,  $u(R, \mathbf{x}) < u(DT_1, \mathbf{x}) < u(PT_1, \mathbf{x})$ ). The induced adaptations by unsuccessful invaders and the agents

using the pure strategy being slightly more costly than the other pure strategy contained in the equilibrium take place at comparable time rates in the replicator dynamic. Therefore, the system moves back towards the respective equilibrium component, unless when the population approaches the respective component’s end point.<sup>16</sup>

Finally, the examination of monomorphic equilibria at the outset suggests there may be another stable equilibrium component. If the conditions are such that some critical mass of  $T_1$ ’s can invade defecting strategies and conversely, defecting strategies can invade conditional cooperators, then there must be at least one other set of steady states  $\mathbf{x}_M = \{(0, \hat{x}_t^{NE}, \rho(1 - \hat{x}_t^{NE}), (1 - \rho)(1 - \hat{x}_t^{NE})) | \rho \in [0, 1]\}$ , in which a fraction of  $T_1$ ’s coexists with a possibly degenerate mix of defecting types.<sup>17</sup> This set is defined by

$$\hat{x}_t^{NE} = \frac{b - 2c + \sqrt{b^2 - 4c[b - c]}/\pi}{2(b - c)}. \tag{8}$$

**Proposition 3.** Assume  $c < 2p$ . If  $\hat{x}_t^{NE} = (b - 2c + \sqrt{b^2 - 4c[b - c]}/\pi) / 2(b - c)$  has a solution within  $[0, 1]$ , the subset of steady states on the  $T_1 - DT_1 - R$  plain that is stable against invasion by  $PT_1$ ’s is nonempty and given by  $\mathbf{x}_3^{NE} = \{(0, \hat{x}_t^{NE}, \rho(1 - \hat{x}_t^{NE}), (1 - \rho)(1 - \hat{x}_t^{NE})) | \rho \in [0, \hat{\rho}]\}$ , where

<sup>15</sup> This fraction is given by the solution to

$$\bar{x}_d = \frac{p(1-\pi^2)+\pi(b/2-\pi c)}{\pi^2 b-\pi c(1-\pi)} + \sqrt{\left(\frac{p(1-\pi^2)+\pi(b/2-\pi c)}{\pi^2 b-\pi c(1-\pi)}\right)^2 + \frac{(1-\pi^2)(c+2k+2p)}{\pi^2 b-\pi c(1-\pi)}}.$$

This solution exists: the first term is negative because  $p(1-\pi^2)$ ,  $b/2-\pi c$ , and  $\pi^2 b-\pi c(1-\pi)$  are positive: obviously,  $p(1-\pi^2) > 0$ ;  $b/2-\pi c > 0$  because, by assumption,  $(N-2)b/(N-1) > c$ ; and  $\pi^2 b > \pi c(1-\pi)$  because, following from (4),  $\pi^2 b > \pi c + 2\pi k(1-\pi)$ , and, in turn,  $\pi c + 2\pi k(1-\pi) > \pi c(1-\pi)$ . From there, it is obvious that  $\bar{x}_d > 0$ ; on the other hand, condition (4) also ensures  $\bar{x}_d < 1$ .

<sup>16</sup> Cf. Weibull and Salomonsson (2006), and references cited therein, for an analogous argument on very similar dynamic phenomena.

<sup>17</sup> If  $\hat{x}_t^{crit}$  as defined in (3) is larger than 0, its combination with any mix of defecting strategies also constitutes a set of—necessarily unstable—steady states.

$$\hat{\rho} = \frac{2p + \pi(b - 2c) + \pi^2(b - 2p) \sqrt{[2p + \pi(b - 2c) + \pi^2(b - 2p)]^2 - 8c(k+p)(1 - \pi - \pi^2 + \pi^3)}}{2c(1 - \pi)}$$

Proof is given in Appendix A.

Proposition 3 simply states that whenever  $\hat{x}_t^{NE}$  as defined by (8) exists, then there will also be a subset of the above defined set  $\mathbf{x}_M$  that is stable against invasion by  $PT_1$ 's. This simply follows from the fact that  $PT_1$ 's will be worse off than  $T_1$ 's whenever all defectors within  $\mathbf{x}_M$  retaliate.

#### 2.4. Dynamics

For a better understanding of the model dynamics, we provide exemplary model trajectories under different conditions in Fig. 1. As initial population states, we choose strategy distributions such that initially equal small shares of invading strategies compete with one incumbent strategy. We do not depict  $R$  as the incumbent strategy, as it is always stable against invasion (and would always look like the upper left hand panel in Fig. 1, only with retaliator fractions near one). Instead, we add the uniform mix of types as an initial state, for an illustration on the dynamics under more mixed initial states.

The first column represents the case in which all preconditions for the emergence of cooperation are favourable; in particular, the continuation probability  $\pi$  fulfils condition (7), and  $p > c$ , so that a single punisher is enough to make contributing worthwhile. The second condition is less restrictive than it may sound: punishment only has to be higher than the *net costs* of contributing, that is, contribution costs net of the benefit derived from the public good.<sup>18</sup> The panels in this column confirm what we have derived above: there are two types of equilibrium sets to which the population converges, a set in which cooperation prevails and one that leads to universal defection.

In the second column, we depict the case when condition (5) is violated, that is,  $\pi$  is high enough that it pays to be a retaliator in a population of only  $PT_1$ 's. In line with Propositions 1 and 2, a population with too high a fraction of punishers is readily invaded by retaliators. The latter may be followed by conditional cooperators if public good benefit and continuation probability are large enough (so that the solution to (8) exists as in the depicted case). However, under some initial population states, conditional cooperators will only be able to invade a defecting population if the initial presence of  $PT_1$ 's allows the fraction of  $T_1$ 's to surpass the critical mass: if the initial punisher fraction is set to zero in the third central column panel, the population is stuck in a defecting equilibrium (not depicted). In this sense, we can state:

**Proposition 4.** Fix  $\pi$  such that  $(c + 2k)/(b + 2k) < \pi < 1$ . If  $\mathbf{x}_3^{NE}$  is nonempty (i.e., the  $b$  to  $c$  ratio and  $\pi$  are sufficiently high), then  $c \geq 2k/3$  is sufficient (but not necessary) to ensure the existence of initial states such that a punishing strategy is needed to reach the 'partially cooperative' equilibrium component  $\mathbf{x}_3^{NE}$ . Notwithstanding, the punishing strategy may be driven to extinction in the ensuing equilibrium.

Proof is given in Appendix A.

Proposition 4 does not mean being a punisher is a dumb thing to do: the third central column panel shows the punishing strategy may be highly successful in the initial, defecting state of the world, provided defectors are likely to be reluctant cooperators. What the extinction of punishers does mean is that it pays to switch strategies once retaliators become sufficiently common.

<sup>18</sup> For example, in a linear laboratory setting with endowments of 10 experimental-currency units and a marginal per-capita return of 0.6, this would mean that a punisher can harm any other player at least by 4 units.

The right hand column represents the case that  $p < c$ , the case in which the costs a single punisher can inflict on a non contributing agent are lower than the net costs of contributing; at the same time, the continuation probability is both too high for  $PT_1$  to be stable against invasion by retaliators and too low for an equilibrium in which conditionally cooperative and defecting types coexist. Therefore, the only attractor left is the 'defecting' equilibrium component  $\mathbf{x}_D^{NE}$ , in this case comprising all possible mixes of reluctant defectors and retaliators, even though punishers may initially spread and persist for some time before being driven out by retaliating players. If we reduce the continuation probability so that  $PT_1$  is stable again, we obtain horizontally stretched versions of the upper three first column panels (not depicted here). However, when starting from an uniformly mixed initial population state, the population goes to pure defection straight away. This clearly indicates what should be clear from the beginning: if the expected potential benefits from future interaction are small because these interactions are unlikely to happen, and if punishers have rather limited punishment possibilities, the likelihood of a cooperative equilibrium being reached is small.

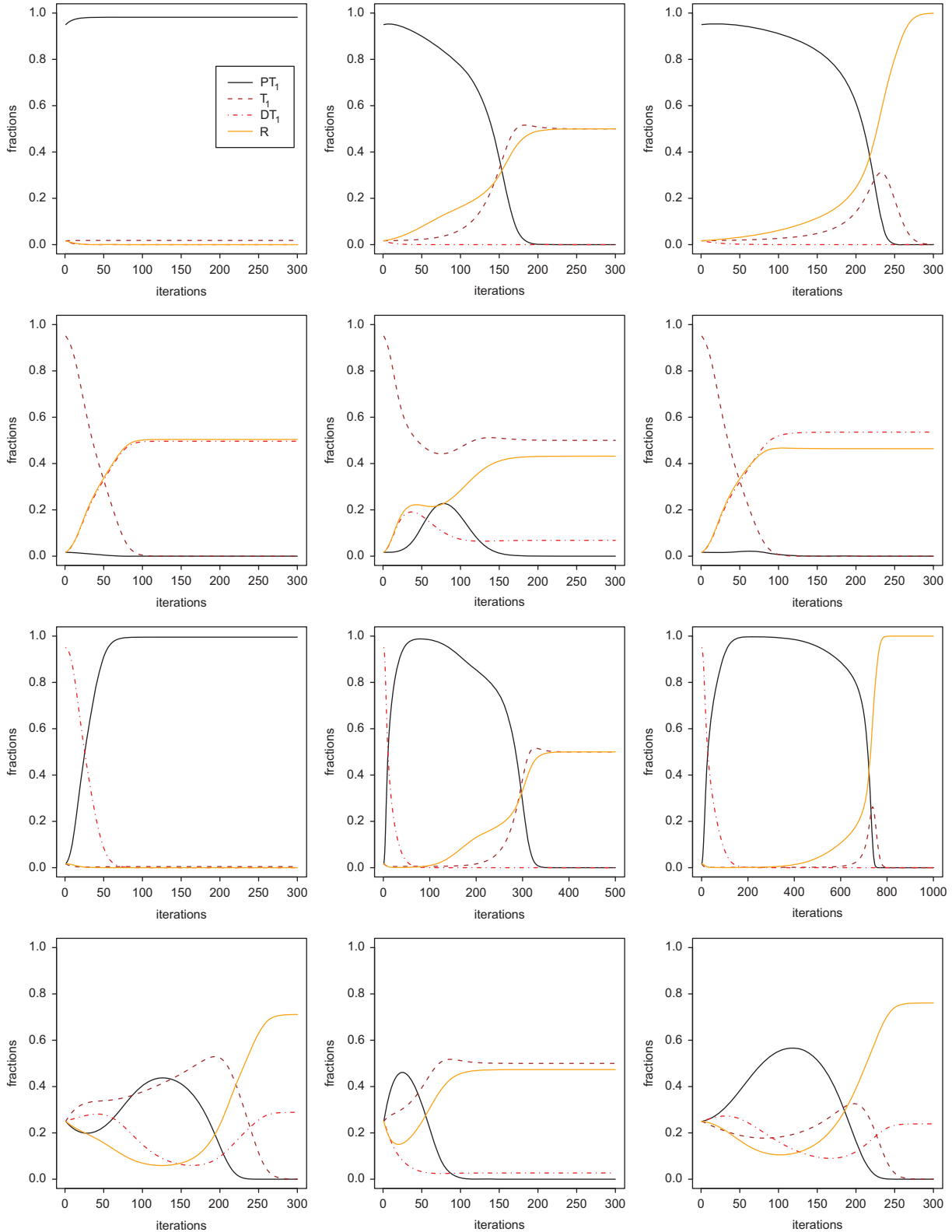
#### 2.5. Sensitivity to strategy exclusion and mutation

In this section, we will touch briefly on two things: (i) we will illustrate that neither of the equilibria derived above would be invaded by strategies that we excluded from the analysis and (ii) we will review what happens to the prevalence of cooperation if we allow for mutations. For the first purpose, Fig. 2 is provided. It depicts trajectories for the cases when small fractions of  $PT_2$ ,  $T_2$ ,  $DT_2$ , and  $D$  types are injected into the long term equilibria (in order not to tilt the results in any direction, we chose equal fractions of  $DT_1$ 's and  $R$ 's in the center and right panels). What Fig. 2 shows is that the equilibria derived above are stable against invasion by the undominated types disregarded in our analysis. If at all, they may help  $PT_1$ 's to invade a polymorphic equilibrium temporarily in the high continuation probability case (right hand panel).

The remainder of this section is devoted to the cooperative equilibrium cases when mutations occur. Given Eqs. (2), it is easy to show that in a world of omnilateral defection, retaliators are almost always selected against under our assumptions when the cooperating types are maintained in small fractions by mutation.<sup>19</sup> The intuition is clear: the defecting types do equally well when matched with conditional cooperators; but a retaliator meeting a punishing cooperator incurs additional costs to leash back at the punisher while substantially lowering the chances of benefitting from the public good in the future. On the other hand, with mutation keeping the defecting types alive, punishing cooperators will be selected against strongly in a world in which they are omnipresent while they would increase in numbers under the parameters chosen in this paper when everybody else is a conditional cooperator.<sup>20</sup> The intuition for these facts is that, when everybody is a punisher, it is costless to free ride on punishment without losing the benefits in terms of future

<sup>19</sup> The condition is given by  $(\pi/(1-\pi))[b(1-2\epsilon-(1/(1+\pi))x_r) - c(2-\epsilon - (2\pi/(1+\pi))x_r)] + 2k > 0$ , where  $\epsilon$  is the fraction of entrants per type that are produced by mutation. The left-hand side decreases in  $x_r$  when, as assumed,  $b > 2c$ . If we set  $x_r \equiv 1 - 2\epsilon$  and take the limit for  $\epsilon \rightarrow 0$ , the left-hand condition reduces to  $\pi b - 2c + 2k(1 - \pi^2)/\pi > 0$ , which will be fulfilled for most relevant parameter combinations  $(k, \pi)$ .

<sup>20</sup> Disregarding any terms with squared mutant fractions, i.e., setting  $\epsilon^2 \equiv 0$ , the comparison  $u(T_1, \mathbf{x}) - u(PT_1, \mathbf{x}) > 0$  is easily transformed into  $(\pi b/(1-\pi))x_t - 4k - 2p < 0$ , which is negative for  $x_t = 0$  but positive for  $x_t = 1$  under the parameter values chosen above.

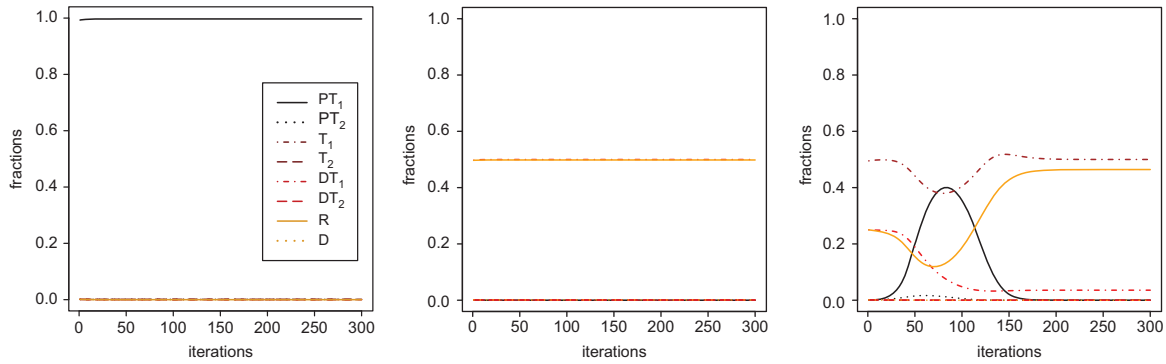


**Fig. 1.** Evolution of types under  $\pi = 3/5, k = 1/2, p = 3/2$  (first column);  $\pi = 4/5, k = 1/2, p = 3/2$  (second column);  $\pi = 3/5, k = 2/7, p = 6/7$  (third column); in all cases,  $c = 1, b = 4$  (i.e.,  $MPCR = 2/3$ ). Note that in the central and right-hand third-row panels, the time-scale was adapted.

interactions, while when nobody else punishes, a punisher reaps higher benefits compared to a conditional cooperator from sporadic encounters with reluctant cooperators.

Combining the above, we would expect a circle: retaliators would be taken over by reluctant cooperators, followed suit by

punishers and conditional cooperators who are, in turn, invaded by retaliators as soon as the fraction of punishers starts to fade. This is, indeed, what happens when the conditions for cooperation are favourable (i.e., when condition (7) holds) and as long as mutation is weak. We illustrate this in the center and right hand



**Fig. 2.** Stability of equilibria under  $\pi = 3/5$  (left, center), and  $\pi = 4/5$  (right), when formerly excluded strategies are included; in all cases,  $c = 1$ ,  $b = 4$  (i.e.,  $MPCR = 2/3$ ),  $k = 1/2$ , and  $p = 3/2$ . (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

first row panels of Fig. 3. When mutation is stronger, it can induce a polymorphic equilibrium as in the upper left hand panel of Fig. 3.<sup>21</sup>

In the second row, we depict the other case in which a positive level of cooperation was attainable, namely, when punishment was not constrained to being weak and the likelihood of future interactions is high. In the absence of mutation the population approached a polymorphic equilibrium component comprising conditional cooperators and a mix of defectors that was stable against invasion by punishers. With mutation, the equilibrium is fully mixed, as mutation counteracts the selective forces working against punishers, with the corresponding repercussions on the equilibrium levels of retaliators, conditional and reluctant cooperators. A numerical analysis suggests that the equilibrium level of punishers and cooperation, for that matter, increases monotonically in the mutation rate in the relevant range, while the equilibrium converges smoothly towards an equilibrium in the above mentioned component as the mutation rate approaches zero, as indicated also by the according second row panels. Note that the effects of mutation on the cooperation level may be substantial: already for the case depicted in the central panel (i.e., for a mutation rate of 1%), first and further iteration cooperation rates are 60.4% and 52.9%, respectively, compared to 50% and 37.5% when no mutations occur.

Summarising, we find that, for intermediate continuation probabilities, moderate mutation puts the population on a cycle of cooperation enacted by punishment, free riding on punishment, defection supported by retaliation, decline of retaliation, and cooperation by punishment again, where the periods of defection tend to be longer than those of cooperation. In turn, under high continuation probabilities as well as under high mutation rates, the population approaches a stable polymorphic equilibrium in which the mutation rate determines the cooperation level; the lower the mutation rate, the lower the level of cooperation, and the less punishing cooperators in the population. We restate the above in the following result:

**Result 1.** In the presence of moderate mutation rates, punishment plays two roles: under intermediate continuation probabilities fulfilling condition (7), it induces short periods of high cooperation levels between longer periods of omnilateral defection; when continuation probabilities are high, it is able to shift up notably the existing cooperation level in the polymorphic steady state.

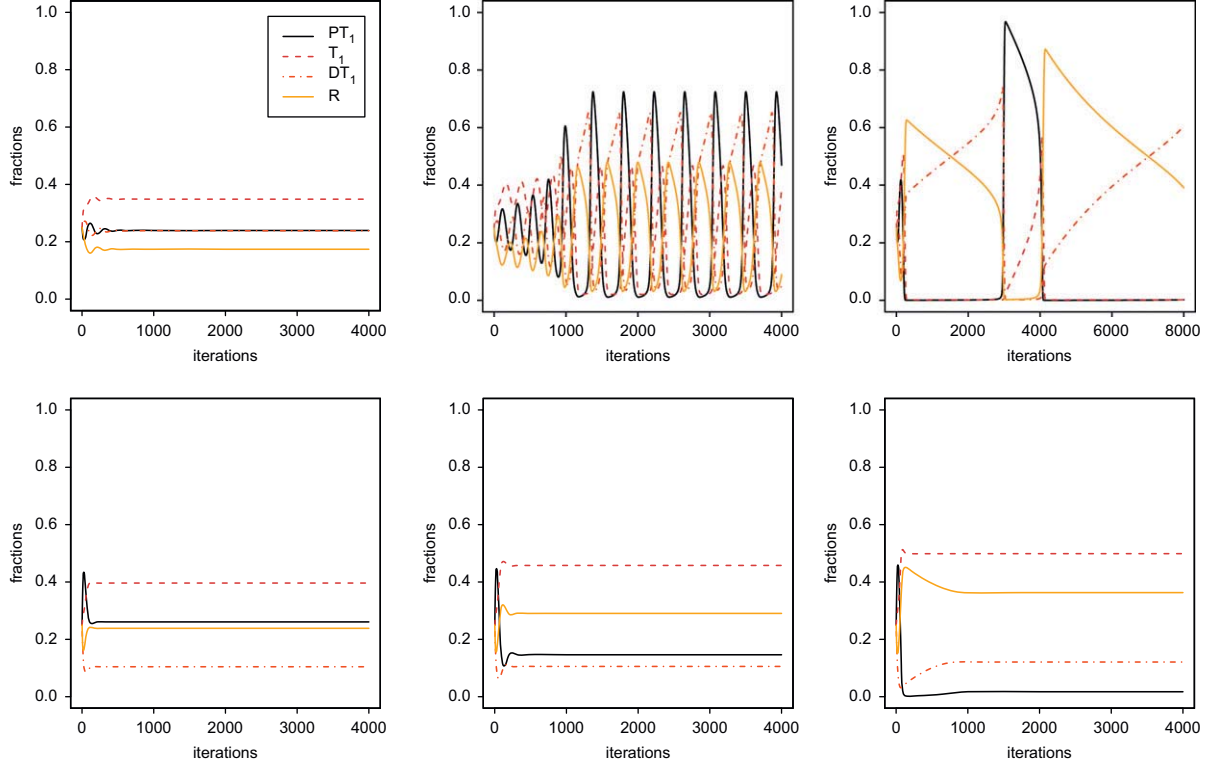
### 3. Summary and discussion

Laboratory experiments suggest retaliation may be an important factor in human behaviour, and that this may have detrimental effects on cooperation. Yet, virtually all models of strong reciprocity and sanction enforced cooperation disregard the possibility of defectors getting back at their punishers. In this paper, we provide a model that accounts for the spread of retaliation as part of a strategy within the context of a repeated social dilemma, and examine its effects on the achievable cooperation level, paying particular attention to the role of punishment in thwarting defection. In the model presented, retaliators may thrive in a society with punishers as they can discourage punishment actions while free riding on future cooperative effort. Depending on the specific parameters, the population is predicted to end up in a cooperative state supported by punishment, in a polymorphic state in which a stable fraction of conditional cooperators coexists with defecting strategies, or a defecting state guarded against invasion by punishers through retaliation.

When mutation plays no role or players do not experiment on their strategy even sporadically, under a different model interpretation a punishing cooperator strategy may be evolutionarily stable when the continuation probability is sufficiently small; when it surpasses a critical value, it pays for defecting types to invest into discouraging future punishment actions by retaliating (Proposition 1). In this case, mirroring the earlier results by Joshi (1987) and Boyd and Richerson (1988), there will be a polymorphic equilibrium component composed of conditional cooperators and a mix of defecting types, in which there are enough retaliating players to rule out invasion by punishing cooperators (Proposition 3). Yet, even though in this equilibrium there are no punishers present, they may play an important role: in a world of omnilateral defection, in which sufficiently many players are reluctant cooperators, punishers may break the invasion barrier conditional cooperators face and thereby jump start cooperation (Proposition 4). In this sense, the temporary presence of punishers may be an unexpected solution to the equilibrium selection problem in the literature on cooperation in infinitely repeated games.

The limited role for punishment is accentuated when looking at the case that mutation plays a role. In this case, punishment merely induces short periods of cooperation that interrupt comparatively long periods of almost omnilateral defection when future interactions are not too likely; under a high continuation probability, it helps bring up the level of cooperation to a limited but non negligible extent in an already partially cooperative population (Result 1). In this sense, punishment seems to be a facilitating device that can pave the way for and boost the power

<sup>21</sup> Note that the graphs are produced under the assumption of a very large population, so that stochastic sampling of mutants plays no role.



**Fig. 3.** Dynamics for a model including mutation rates of  $\mu = 0.02$  (left),  $\mu = 0.01$  (center), and  $\mu = 0.001$  (right), for  $\pi = 3/5$  (first row) and  $\pi = 4/5$  (second row); in all cases,  $c = 1$ ,  $b = 4$ ,  $k = 1/2$ , and  $p = 3/2$ . In the upper right-hand panel, the time scale was adapted.

of conditional cooperation, rather than being an explanation for the evolution of cooperation of itself. A retaliation strategy, in contrast, has a place in this scenario as it conveys a benefit to its carrier organism as soon as punishers become sufficiently wide spread.

### Acknowledgements

I am grateful to an anonymous referee of an earlier study for inciting this work. Furthermore, I would like to thank the lively research group of the Thurgau Institute of Economics (TWI) as well as two anonymous reviewers for helpful comments on earlier drafts.

### Appendix A. Proofs for Propositions 1–4

**Proposition 1.** *If the continuation probability  $\pi$  exceeds  $\pi^{\max} = (2p+2k-c)/(2p+2k)$ , neither cooperative strategy is stable against invasion by some defecting type.*

**Proof.** By (2), and as noted above,

$$u(T_1, (0, 1-\varepsilon, \varepsilon, 0)) - u(DT_1, (0, 1-\varepsilon, \varepsilon, 0)) = \frac{(1-\varepsilon^2\pi)c - \varepsilon(1-\varepsilon)\pi b}{1-\pi},$$

which is negative under any value of  $\pi$  for small enough  $\varepsilon$ .  $PT_1$ , on the other hand, is stable against invasion by  $R$  if and only if

$$u(PT_1, (1-\varepsilon, 0, 0, \varepsilon)) - u(R, (1-\varepsilon, 0, 0, \varepsilon)) = 2(k+p)(1-2\varepsilon) \frac{(1-\varepsilon^2\pi)c - \varepsilon(1-\varepsilon)\pi b}{1-\pi} > 0. \quad (9)$$

Letting  $\varepsilon$  go to 0 and rearranging, we obtain the critical value from the proposition.  $\square$

**Proposition 2.** *Define  $\underline{x}_p \equiv \max\{c/[(1-\pi)(2k+2p)], c/2p\}$  and assume  $c < 2p$ . Then, the subset of points in the  $PT_1 - T_1$  continuum that are stable against invasion by both defecting types is nonempty and given by  $\mathbf{x}_c^{NE} = \{(1-\theta(1-\underline{x}_p), \theta(1-\underline{x}_p), 0, 0) | \theta \in [0, 1)\}$ , if and only if*

$$\pi \begin{cases} < \frac{2p+2k-c}{2p+2k}, & b \leq \frac{2c}{\pi}, \text{ and} \\ \leq \frac{2p+2k-c}{2p+2k} & \text{otherwise.} \end{cases} \quad (10)$$

**Proof.** First of all, along the  $PT_1 - T_1$  continuum, all cooperating strategies obtain the same payoff of  $(b-c)/(1-\pi)$ . Equating  $u(R, (x_p', 1-x_p', 0, 0))$  as defined by (2) with  $(b-c)/(1-\pi)$  gives  $x_p' = c/[(1-\pi)(2k+2p)]$ ; for  $u(DT_1, (x_p'', 1-x_p'', 0, 0)) = (b-c)/(1-\pi)$ , we obtain  $x_p'' = c/2p$ . Eq. (10) and  $c < 2p$  ensure the existence of  $\mathbf{x}_c^{NE}$ . By the monotonicity of defector payoffs along the  $PT_1 - T_1$  vertex, any fraction of punishers above  $\underline{x}_p$  implies a lower payoff for the defecting entrants.

If  $\pi > (2p+2k-c)/(2p+2k)$ , Proposition 1 ensures  $\mathbf{x}_c^{NE} = \emptyset$ . Finally, if  $\pi = (2p+2k-c)/(2p+2k)$ , evolutionary stability of  $\mathbf{x}_c^{NE} = \{(1, 0, 0, 0)\}$  depends on the exact  $b$  to  $c$  ratio: if  $\varepsilon = 0$ ,  $u(PT_1, \mathbf{x}_c^{NE}) - u(R, \mathbf{x}_c^{NE}) = 0$ ; at the same time, the left hand side of (9) decreases in  $\varepsilon$  if and only if its derivative is negative:

$$2c + \pi b - 2\varepsilon\pi(b-c) < 0, \quad (11)$$

obtained by substituting  $c/(1-\pi) = 2(p+k)$  and rearranging. Hence, only if  $b > 2c/\pi$  is there an  $\bar{\varepsilon}$  such that any  $\varepsilon$ ,  $0 < \varepsilon < \bar{\varepsilon}$ , yields  $u(PT_1, (1-\varepsilon, 0, 0, \varepsilon)) - u(R, (1-\varepsilon, 0, 0, \varepsilon)) > 0$ .  $\square$

**Proposition 3.** Assume  $c < 2p$ . If  $\hat{x}_t^{NE} = (b - 2c + \sqrt{b^2 - 4c(b - c)/\pi}) / 2(b - c)$  has a solution within  $[0, 1]$ , the subset of steady states on the  $T_1$   $DT_1$   $R$  plain that is stable against invasion by  $PT_1$ 's is nonempty and given by  $\mathbf{x}_3^{NE} = \{(0, \hat{x}_t^{NE}, \rho(1 - \hat{x}_t^{NE})), (1 - \rho)(1 - \hat{x}_t^{NE}) | \rho \in [0, \hat{\rho}]\}$ , where

$$\hat{\rho} = \frac{2p + \pi(b - 2c) + \pi^2(b - 2p) \sqrt{[2p + \pi(b - 2c) + \pi^2(b - 2p)]^2 - 8c(k + p)(1 - \pi - \pi^2 + \pi^3)}}{2c(1 - \pi)}.$$

**Proof.** Existence of stable states in  $\mathbf{x}_3^{NE}$  follows from the fact that for  $\rho = 0$ ,  $u(PT_1, \mathbf{x}_3^{NE}) - u(T_1, \mathbf{x}_3^{NE}) = (1 - \hat{x}_t^{NE})2(k + p) < 0$ . To obtain the boundary of  $\mathbf{x}_3^{NE}$ , we equate  $u(T_1, \mathbf{x}_3^{NE})$  to  $u(PT_1, \mathbf{x}_3^{NE})$  and rearrange to obtain

$$\frac{(x_d^2 + x_d x_r) \pi b}{1 - \pi} (x_d^2 + 2x_d x_r + x_r^2) \cdot 2k \left( x_d^2 + \frac{2x_d x_r \pi}{1 + \pi} \right) - \frac{c}{1 - \pi} (x_d x_r + x_r^2) \cdot 2p = 0. \quad (12)$$

Recalling that  $x_d = \rho(1 - \hat{x}_t^{NE})$  and  $x_r = (1 - \rho)(1 - \hat{x}_t^{NE})$ , (12) simplifies to

$$\rho^2 \frac{c}{1 + \pi} + \rho \left( \frac{\pi b}{1 - \pi} - \frac{2\pi c}{1 - \pi^2} + 2p \right) - 2(k + p) = 0. \quad (13)$$

Solving (13) for  $\rho$  yields  $\hat{\rho}$  from the proposition.<sup>22</sup> Monotonicity of  $PT_1$  and  $T_1$  payoffs along  $\rho$  ensures  $u(PT_1, \mathbf{x}_3^{NE}) < u(T_1, \mathbf{x}_3^{NE})$  for all  $\mathbf{x} \in \mathbf{x}_3^{NE}$ .  $\square$

**Proposition 4.** Fix  $\pi$  such that  $(c + 2k)/(b + 2k) < \pi < 1$ . If  $\mathbf{x}_3^{NE}$  is nonempty (i.e., the  $b$  to  $c$  ratio and  $\pi$  are sufficiently high), then  $c \geq 2k/3$  is sufficient (but not necessary) to ensure the existence of initial states such that a punishing strategy is needed to reach the 'partially cooperative' equilibrium component  $\mathbf{x}_3^{NE}$ . Notwithstanding, the punishing strategy may be driven to extinction in the ensuing equilibrium.

**Proof.** Note that  $\hat{x}_t^{crit}$  defined by (3) exists whenever  $\mathbf{x}_3^{NE}$  is nonempty. Moreover, when  $PT_1$ 's are absent, any mix of  $DT_1$  and  $R$  is stable against invasion by  $T_1$ : by  $b > c$ ,  $\hat{x}_t^{crit} > 0$  for any  $\pi < 1$ . Particularly, this holds also for mixes of defector types in which  $DT_1$  is very prevalent. Consider the special case  $\mathbf{x}' = (0, 0, 1, 0)$ . By (4),  $PT_1$ 's can invade  $\mathbf{x}'$  if  $\pi > (c + 2k)/(b + 2k)$ . Existence of  $\mathbf{x}_3^{NE}$  obviously requires  $b^2 - 4c(b - c)/\pi > 0$ , which implies  $\pi > (c + 2k)/(b + 2k)$  if  $c > 2k/3$ .<sup>23</sup> The final claim from the proposition is evident from the example given in the third central column panel of Fig. 1.  $\square$

## Appendix B. Supplementary data

Supplementary data associated with this article can be found in the online version at <http://dx.doi.org/10.1016/j.jtbi.2012.09.012>.

## References

Alexander, Richard D., 1979. Darwinism and Human Affairs. University of Washington Press, Seattle.

<sup>22</sup> Obviously, there may be another solution,  $\hat{\rho}'$ , to (13), replacing the minus in front of the square root by a plus. Yet, if  $\hat{\rho}'$  exists,  $\hat{\rho}' > 1$ : for existence, the term under the square root must be positive, i.e.,  $[2p + \pi(b - 2c) + \pi^2(b - 2p)]^2 > 8c(k + p)(1 - \pi - \pi^2 + \pi^3)$ .  $\hat{\rho}' < 1$  obviously requires  $2p + \pi(b - 2c) + \pi^2(b - 2p) < 2c(1 - \pi)$ . However,  $[2c(1 - \pi)]^2 < 8c(k + p)(1 - \pi - \pi^2 + \pi^3)$  under  $c < 2p$  for any  $\pi$ . Hence, either  $\hat{\rho}'$  does not exist or  $\hat{\rho}' > 1$ .

<sup>23</sup> To see this, rearrange  $b^2 - 4c(b - c)/\pi > 0$  to  $\pi > (4cb - 4c^2)/b^2$ , implying  $\pi > (c + 2k)/(b + 2k)$  if and only if  $\omega \equiv 3b^2c + 8bck - 4bc^2 - 8c^2k - 2b^2k > 0$ .  $\omega$  increases in  $b$  if  $6bc + 8ck - 4c^2 - 4bk > 0$ , which is fulfilled for any  $b > 2c$  if  $c \geq 2k/3$ . Hence,  $\omega > 12c^3 + 16c^2k - 8c^3 - 8c^2k - 8c^2k$  by  $b > 2c$ , and consequently,  $\omega > 0$ .

Alexander, Richard D., 1987. The Biology of Moral Systems. Aldine de Gruyter, New York.  
Axelrod, Robert, 1986. An evolutionary approach to norms. Am. Political Sci. Rev. 80, 1095–1111.  
Axelrod, Robert, Hamilton, D. William, 1981. The evolution of cooperation. Science 211, 1390–1396.

Binmore, Kenneth G., Samuelson, Larry, 1992. Evolutionary stability in repeated games played by finite automata. J. Econ. Theory 57, 278–305.  
Bowles, Samuel, Gintis, Herbert, 2004. The evolution of strong reciprocity: cooperation in heterogeneous populations. Theor. Popul. Biol. 65, 17–28.  
Boyd, Robert, Richerson, Peter J., 1985. Culture and the Evolutionary Process. University of Chicago Press, Chicago.  
Boyd, Robert, Richerson, Peter J., 1988. The evolution of reciprocity in sizable groups. J. Theor. Biol. 132, 337–356.  
Boyd, Robert, Richerson, Peter J., 1992. Punishment allows the evolution of cooperation (or anything else) in sizable groups. Ethology Sociobiology 13, 171–195.  
Carpenter, Jeffrey P., 2007. Punishing free-riders: How group size affects mutual monitoring and the provision of public goods. Games Econ. Behav. 60, 31–51.  
Carpenter, Jeffrey P., Matthews, Peter H., Ong'ong'a, Okomboli, 2004. Why Punish? Social reciprocity and the enforcement of prosocial norms. J. Evol. Econ. 14, 407–429.  
Cavalli-Sforza, Luigi L., Feldman, Marcus W., 1981. Cultural Transmission and Evolution. Princeton University Press, Princeton, NJ.  
Cheung, Stephen L., 2012. New insights into conditional cooperation and punishment from a strategy method experiment. Working paper.  
Cinyabuguma, Matthias, Page, Talbot, Putterman, Louis, 2006. On perverse and second-order punishment in public goods experiments with decentralized sanctioning. Exp. Econ. 9, 265–279.  
Denant-Boemont, Laurent, Masclot, David, Noussair, Charles, 2007. Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. Econ. Theory 33, 145–167.  
Dreber, Anna, Rand, David G., Fudenberg, Drew, Nowak, Martin A., 2008. Winners don't punish. Nature 452, 348–351.  
Fehr, Ernst, Gächter, Simon, 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90 (4), 980–994.  
Fehr, Ernst, Henrich, Joseph, 2003. Is strong reciprocity a maladaptation? On the evolutionary foundations of human altruism. In: Hammerstein, Peter (Ed.), Genetic and Cultural Evolution of Cooperation. MIT Press, Cambridge, MA.  
Fowler, James H., 2005. Altruistic punishment and the origin of cooperation. Proc. Natl. Acad. Sci. USA 102, 7047–7049.  
Fudenberg, Drew, Maskin, Eric, 1986. The Folk theorem in repeated games with discounting or with incomplete information. Econometrica 54, 533–556.  
Fudenberg, Drew, Maskin, Eric, 1990. Evolution and cooperation in noisy repeated games. Am. Econ. Rev. 80 (2), 274–279.  
García, Julián, Traulsen, Arne, 2012. Leaving the loners alone: evolution of cooperation in the presence of antisocial punishment. J. Theor. Biol. 307, 168–173.  
Gintis, Herbert, 2000. Strong reciprocity and human sociality. J. Theor. Biol. 206, 169–179.  
Gintis, Herbert, Bowles, Samuel, Boyd, Robert, Fehr, Ernst, 2003. Explaining altruistic behavior in humans. Evol. Hum. Behav. 24, 153–172.  
Gintis, Herbert, Smith, Eric A., Bowles, Samuel, 2001. Costly signaling and cooperation. J. Theor. Biol. 213, 103–119.  
Hamilton, William D., 1964. The genetical evolution of social behaviour. I and II. J. Theor. Biol. 7, 1–52.  
Hammerstein, Peter (Ed.), 2003. Genetic and Cultural Evolution of Cooperation. MIT Press, Cambridge, MA.  
Hauert, Christoph, Traulsen, Arne, Brandt, Hannelore, Nowak, Martin A., Sigmund, Karl, 2007. Via freedom to coercion: the emergence of costly punishment. Science 316, 1905–1907.  
Henrich, Joseph, Boyd, Robert, 2001. Why people punish defectors: weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. J. Theor. Biol. 208, 79–89.  
Herrmann, Benedikt, Thöni, Christian, Gächter, Simon, 2008. Antisocial punishment across societies. Science 319, 1362–1367.  
Janssen, Marco A., Bushman, Clint, 2008. Evolution of cooperation and altruistic punishment when retaliation is possible. J. Theor. Biol. 254, 541–545.  
Joshi, N.V., 1987. Evolution of cooperation by reciprocation within structured demes. J. Genet. 66, 69–84.  
Mathew, Sarah, Boyd, Robert, 2009. When does optional participation allow the evolution of cooperation? Proc. R. Soc. B 276, 1167–1174.  
Nicklisch, Andreas, Wolff, Irenaues, 2011. Cooperation norms in multiple-stage punishment. J. Public Econ. Theory 13, 791–827.  
Nikiforakis, Nikos, 2008. Punishment and counter-punishment in public good games: Can we really govern ourselves?. J. Public Econ. 92, 91–112.

- Nowak, Martin A., Sigmund, Karl, 1993. A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. *Nature* 364, 56–58.
- Nowak, Martin A., Sigmund, Karl, 1998. Evolution of indirect reciprocity by image scoring. *Nature* 393, 573–577.
- Ohtsuki, Hisashi, Iwasa, Yoh, Nowak, Martin A., 2009. Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. *Nature* 457, 79–82.
- Rand, David G., Armao IV, Joseph J., oNakamaru, Mayuk, Ohtsuki, Hisashi, 2010. Anti-social punishment can prevent the co-evolution of punishment and cooperation. *J. Theor. Biol.* 265, 624–632.
- Rand, David G., Nowak, Martin A., 2011. The evolution of antisocial punishment in optional public goods games. *Nat. Commun.* 2, 434.
- Rand, David G., Ohtsuki, Hisashi, Nowak, Martin A., 2009. Direct reciprocity with costly punishment: generous tit-for-tat prevails. *J. Theor. Biol.* 256, 45–57.
- Sethi, Rajiv, 1996. Evolutionary stability and social norms. *J. Econ. Behav. Organ.* 29, 113–140.
- Sober, Elliott, Wilson, David S., 1998. *Unto Others: The Evolution and Psychology of Unselfish Behavior*. Harvard University Press, Cambridge, MA.
- Stephens, Christopher, 2005. Strong reciprocity and the comparative method. *Anal. Kritik* 27, 97–105.
- Sugden, Robert, 1986. *The Economics of Rights, Co-operation, and Welfare*. Blackwell Publishing Limited, Oxford.
- Taylor, Peter D., Jonker, Leo B., 1978. Evolutionary stable strategies and game dynamics. *Math. Biosci.* 40, 145–156.
- Trivers, Robert L., 1971. The evolution of reciprocal altruism. *Q. Rev. Biol.* 46, 35–57.
- Weibull, Jörgen W., Salomonsson, Marcus, 2006. Natural selection and social preferences. *J. Theor. Biol.* 239, 79–92.
- Wolff, Irenaeus, 2009. Punishers under fire: retaliation and the evolution of cooperation. Working Paper.
- Yamagishi, Toshio, 1986. The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol. Rev.* 51, 110–116.
- Zahavi, Amotz, 1975. Mate selection—a selection for a handicap. *J. Theor. Biol.* 53, 205–214.