Refinement of defection strategies stabilizes cooperation

Neelesh Dahanukar* and Milind Watve

Department of Microbiology, Abasaheb Garware College, Karve Road, Pune 411 004, India and
Indian Institute of Science Education and Research, Sai Trinity, Central Tower, Garware Circle, Pune 411 021, India

Cooperation among unrelated individuals is an enduring evolutionary riddle and a number of possible solutions have been suggested. Most of these suggestions attempt to refine cooperative strategies, while little attention is given to the fact that novel defection strategies can also evolve in the population. Especially in the presence of punishment to the defectors and public knowledge of strategies employed by the players, a defecting strategy that avoids getting punished by selectively cooperating only with the punishers can get a selective benefit over non-conditional defectors. Furthermore, if punishment ensures cooperation from such discriminating defectors, defectors who punish other defectors can evolve as well. We show that such discriminating and punishing defectors can evolve in the population by natural selection in a Prisoner's Dilemma game scenario, even if discrimination is a costly act. These refined defection strategies destabilize unconditional defectors. They themselves are, however, unstable in the population. Discriminating defectors give selective benefit to the punishers in the presence of non-punishers by cooperating with them and defecting with others. However, since these players also defect with other discriminators they suffer fitness loss in the pure population. Among the punishers, punishing cooperators always benefit in contrast to the punishing defectors, as the latter not only defect with other punishing defectors but also punish them and get punished. As a consequence of both these scenarios, punishing cooperators get stabilized in the population. We thus show ironically that refined defection strategies stabilize cooperation. Furthermore, cooperation stabilized by such defectors can work under a wide range of initial conditions and is robust to mistakes.

Keywords: Discrimination, evolution of cooperation, evolutionary game theory, punishment, reputation.

IN Prisoner's Dilemma (PD), despite the fact that mutual cooperation always yields more benefits than mutual defection, personal interests forbid the evolution of cooperation. A number of possible mechanisms which have been claimed to bolster cooperation, include kin selection[1], direct[2,3] and indirect[4,5] reciprocity. These mechanisms fail to account for cooperation in humans throughout the range of realistic conditions. Humans are known to cooperate with genetically unrelated individuals, with people they will never meet again and when reputation gains are negligible[6,7]. Recent ethnographic[8,9], empirical[6] and theoretical[10–16] studies advocate the idea that cooperation can evolve if cooperating individuals have an inclination towards punishing the defectors. This suggestion, however, displaces rather than resolves the puzzle[17]. Since non-punishing cooperators can free-ride on punishers' contribution to punishment, punishment itself is destabilized. The second-order free riders can invade punishing cooperators and in the absence of punishment, cooperation cannot be stable. Mechanisms including conformism[16], meta-punishment[18], meta-reward[19], signalling[20], reputation damage[21,22], group selection[12,15] and voluntary participation[23–25] have been suggested as possible solutions to the second-order free-rider problem. However, they work only under specific conditions.

The focus so far has been on sophistication of cooperative and punishing strategies, while little consideration has been given to the possibility that greater sophistication in the defection strategies will also evolve. We show here that sophistication in defection ironically stabilizes punishment and cooperation in turn.

If public knowledge of strategies is possible, then punishers can build a reputation for punishing. This can enable the defectors to discriminate between punishing and non-punishing individuals. We propose that if such discriminating defectors defect with both non-punishing cooperators and other defectors, but cooperate with punishers to avoid punishment, then such defectors will have a selective benefit over unconditional defectors. Furthermore, if discriminating defectors cooperate to avoid punishment, then a defector strategy that punishes other defectors can exploit more cooperation by threatening the discriminating individuals. We show that such discriminating and punishing defectors will evolve under natural selection and in turn destabilize unconditional defection. However, these defectors confront severe fitness loss in pure populations as they keep defecting with similar strategists or punish and get punished from similar strategists, or both. Nonetheless, these strategies give selective benefit to the punishing cooperators and thus stabilize punishment and in turn cooperation in the population.

*For correspondence. (e-mail: neeleshdahanukar@rediffmail.com)

Our synthesis buds from earlier work on the evolution of cooperation in the presence of punishment[2,18,26]. Trivers[2], while discussing the evolution of cooperation through reciprocal altruism, anticipated that punishment (what he called as moralistic aggression) will evolve to protect against defection, which in turn would lead to the evolution of discriminating strategies. In a repeated game, Boyd and Richerson[18] considered a strategy called 'reluctant cooperators' who defect in all round before getting punished and after getting punished continue cooperating forever. A more direct reference to strategy similar to discriminating defectors comes in Sigmund *et al.*[26] who, in a mini-public goods game, have considered, among other behavioural options, a small probability that defectors cooperate with punishers. They, however, failed to consider discriminating defector as a distinct behavioural strategy and they do not discuss the contribution of this strategy in the evolution of punishment in an extensive manner. The concept of defecting punishers (also called selfish punishers) has been advocated as a strategy that maintains polymorphic equilibrium, where both cooperators and defecting punishers coexist[27–29]. We, however, argue that only defecting punishers do not resolve the problem of cooperation, but discriminating and punishing defectors together can destabilize defective equilibrium of PD and pave the way for the evolution of punishment and cooperation in the system.

In this article, we consider two refinements in defection strategies: (1) discriminating ability and (2) punishing ability. We build our model in two stages. In the first stage we have considered a model in which only cooperators can punish and the strategies are deterministic, i.e. a player can either be an unconditional cooperator, punishing cooperator, unconditional defector or discriminating defector. We show that punishing cooperators are stable in the presence of discriminating defectors and the results are robust to errors in discrimination. In the second stage we relax our assumption that only cooperators can punish and that players use only one discrete strategy, by building a stochastic strategy game. We show that punishment evolves in the presence of discriminating defectors even in the stochastic strategy game. We further show that it is possible to evolve stable cooperation with refined defection strategies and without classical cooperators.

Public knowledge and the evolution of discriminating defectors

Model

In a two-person PD game both players have two behavioural options, either to cooperate or to defect. Let b be the benefit of a cooperative act and c be the cost of cooperation, such that $b > c > 0$. In a PD interaction the benefit of a cooperative act is only to the opponent, while the cooperator pays the cost of cooperation. Thus, if both the players cooperate mutually, then both achieve the payoff $b–c$; while if both defect, both get nothing. If one player cooperates while the other defects, then the cooperator gets $–c$ and the defector gets b. As defection always pays better, irrespective of the behavioural choice of the opponent, defection is the only Nash Equilibrium (NE) in this game.

We considered an extended PD game with four behavioural options for each player (Table 1). Cooperators (C) unconditionally cooperate with all other strategies without punishing a defector. A defector (D) unconditionally defects with all other strategies. Punishing cooperators (PC) cooperate with all other strategies, but punish a defector with a penalty y to the defector and at a cost x to itself ($y > x > 0$). If we consider a scenario where public knowledge of strategies is known, then punishers can build a reputation for punishing a defector. In such conditions, discriminating defector (DD) can discriminate between players and can selectively cooperate with a punisher to avoid punishment. A player adopting DD strategy pays cost d for discrimination. We have assumed that $d < c$, so that players will prefer to discriminate and defect rather than cooperate with a cooperator. It would be also logical to assume that $y > c + d$, so that players prefer to cooperate with a punisher rather than defect and get punished.

If p, q, r and s are the frequencies of players adopting strategies C, D, PC and DD respectively, such that $p + q + r + s = 1$, then based on the payoff matrix (Table 1), the payoffs of individuals adopting different strategies can be given as:

$$E_{\mathrm{C}} = (p + r)b - c, \tag{1}$$

$$E_{\mathrm{D}} = (p + r)b - ry, \tag{2}$$

$$E_{\mathrm{PC}} = (p + r + s)b - qx - c, \tag{3}$$

$$E_{\mathrm{DD}} = (p + r)b - rc - d, \tag{4}$$

Table 1. Payoff matrix of the extended Prisoner's Dilemma game

	C	D	PC	DD
C	$b–c$	$–c$	$b–c$	$–c$
D	b	0	$b–y$	0
PC	$b–c$	$–c–x$	$b–c$	$b–c$
DD	$b–d$	$–d$	$b–c–d$	$–d$

Each player can adopt one of the four different strategies – cooperator (C), defector (D), punishing cooperator (PC) and discriminating defector (DD). Payoffs of only the row players are given in the matrix, where b is the benefit of a cooperative act, c the cost of cooperation, x the cost of punishment, y the penalty paid by the defector who is punished and d is the cost of discrimination between players adopting different strategies ($d < c$ and $y > c + d$).

where E_I indicates average payoff of an individual adopting strategy I. The frequency of the players using the given strategy changes according to the replicator equation[30,31],

$$\dot{x}_i = x_i(f_i - \bar{f}). \tag{5}$$

The replicator equation describes deterministic but frequency-dependent selection dynamics. The fitness, f_i, of type i is a function of the frequencies of all strategies (phenotypes) $\vec{x} = (x_1, x_2, ..., x_n)$. For pairwise interactions, we can consider linear fitness functions, $f_i = \Sigma_j x_j a_{ij}$, where the a_{ij} values denote the payoff matrix of the game (i.e. payoff of an individual using strategy i when in contest with an individual using strategy j). The average payoff of the population is given by $\bar{f} = \Sigma_i x_i f_i$. The replicator dynamics ensures that even if the population starts from any initial frequency of players adopting different strategies, it will eventually change and will converge to NE. Once the population converges to NE, it will stay there forever[31]. We will indicate the payoff of strategy I when in contest with a strategy J as $E[I|J]$.

In our model, players interact in pairs and the same two players need not meet again (or in other words, need not keep a memory of previous interactions). Thus, the individuals are concerned only with their current gains. We have assumed that the discriminators can keep track of the strategies of other players by eavesdropping, gossiping or through a network of communication. Irrespective of the mode of information gathering, we always assume that discrimination incurs some cost to the discriminator.

Analysis

In the absence of PC and DD, D is the only NE in the game and even though initially rare, D can invade a population of C and get stabilized. However, if we add PC players to the population of C and D, then the dynamics of the game changes drastically. There is a mixed strategy NE at $r^* = (c + x)/(y + x)$, where if $p = 0$ and $q^* = (y - c)/(y + x)$, then such a population is bistable. That is if $r > r^*$, then cooperation gets apparently stable, while if $r < r^*$, D invades the system and becomes stable (Figure 1). However, r^* is a weak equilibrium. Even when $r > r^*$, it cannot be assured that cooperation will be stable, because if there is even a single defector in the system then PC players will pay some cost of punishing, while C will not and so selection will favour C over PC players. Also, when selection is neutral for C and PC, natural drift can bring down the frequency of PC below r^*. If this happens, defectors can invade the population and become stable. Similar arguments are advocated for strategies regarding indirect reciprocity[5] and for punishment in 'n-person' situations[11].

If we add DD players to the heterogeneous population containing C, D and PC, then PC can become stabilized (Figure 2). We have studied the dynamics of PC in the presence of DD in two different scenarios. In the first scenario, the cost of discrimination is assumed negligibly small. In this game, DD players will replace D players in the population. This is because DD players do not get punished as they cooperate with a PC player. In a population comprising of PC, C and DD, PC is the only NE predicted by the replicator dynamics (Figure 3 *a*). Success of PC in the presence of DD is due to the selective benefit that it gets over a C player, as E[PC|DD] > E[C|DD]. Furthermore, a PC player gets cooperation from PC, C as well as DD players and a PC player does not have to pay the cost of punishment as no strategy defects with it. In a population of C players, when DD and PC are rare, DD players invade C players as E[DD|C] > E[C|C]. Nonetheless, DD players cannot sustain their population against PC players as E[PC|DD] > E[DD|DD]. The heterogeneous population of C, PC and DD, therefore, has a unique NE at PC.

In the second scenario, the cost of discrimination is considered significant. In this game both DD and D players will coexist. Thus, we will consider a population containing only PC, D and DD strategy as in the presence of two defecting strategies, D and DD; C anyway gets eliminated in the initial few rounds of the iterated PD (Figure 2). In a population comprising only PC, D and DD players, PC can become stable over a wide set of initial frequencies (Figure 3 *b*), or in other words, has a bigger basin of attraction. The success of PC lies in the selective benefit to the PC players over D players in the presence of DD players as E[PC|DD] > E(D, DD) and in the ability of PC players to invade the population of DD players. The point $r^* = (c + x)/(y + x)$ on the PC and D edge of the

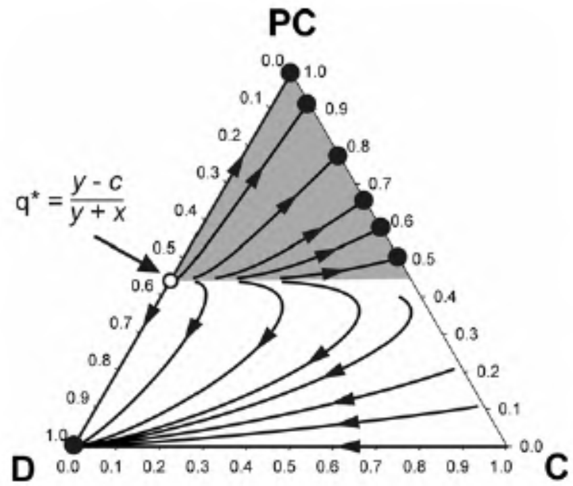

Figure 1. Ternary plot of change in frequency of players using PC, C and D strategies. Filled circles are stable Nash Equilibria (NEs), while unfilled circles are unstable NEs. Arrows indicate direction of changing frequencies of individuals using a particular strategy, from their initial frequencies. The point q^* is the threshold frequency of D above which D will be stable. Cooperation evolves in the grey area. Parameters used here are: $b = 1$, $c = 0.3$, $x = 0.1$ and $y = 0.8$.

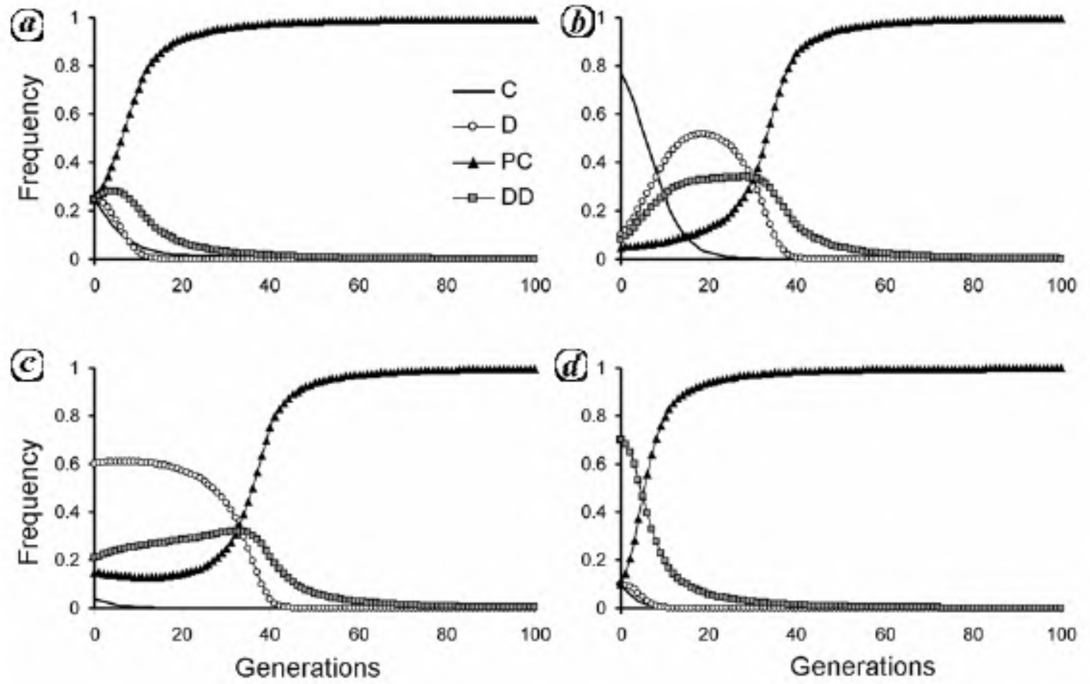

Figure 2. Change in frequency and average payoffs of non-punishing cooperators (C), defectors (D), punishing cooperators (PC) and discriminating defectors (DD), for different initial frequencies of the strategies. (***a***) $p = 0.25$, $q = 0.25$, $r = 0.25$, $s = 0.25$. (***b***) $p = 0.77$, $q = 0.10$, $r = 0.05$, $s = 0.08$. (***c***) $p = 0.04$, $q = 0.60$, $r = 0.15$, $s = 0.21$. (***d***) $p = 0.1$, $q = 0.1$, $r = 0.1$, $s = 0.7$. Parameters used here are: $b = 1$, $c = 0.3$, $x = 0.1$, $y = 0.8$ and $d = 0.05$.

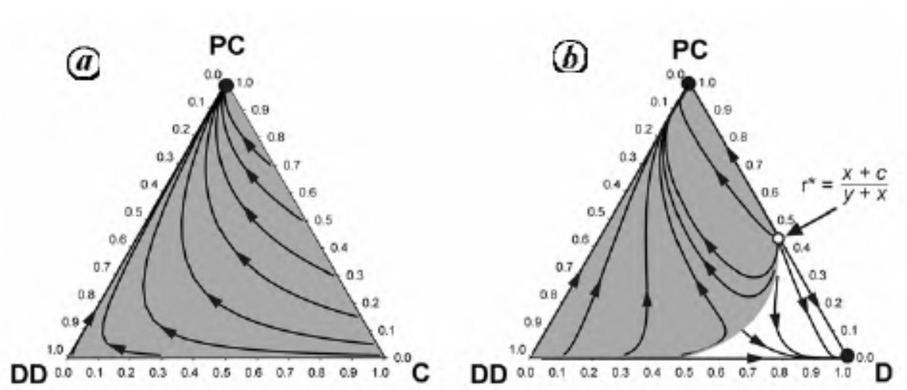

Figure 3. Evolutionary dynamics between C, D, PC and DD players when the cost of discrimination is negligibly small ($d = 0.001$) (***a***) and is substantial ($d = 0.05$) (***b***). The point r^* is the threshold frequency of PC above which PC or alternatively cooperation will be stable. Cooperation evolves in the grey area. Filled circles are stable NEs, while empty circles are unstable NEs. Arrows indicate direction of changing frequencies of individuals using a particular strategy, from their initial frequencies. Other parameters used here are: $b = 1$, $c = 0.3$, $x = 0.1$ and $y = 0.8$.

simplex given in Figure 3*b* is an unstable equilibrium and PC players are fixed in the population only if $r > r^*$.

Mistakes in discrimination

Discrimination by DD players is prone to mistakes. DD players can defect with PC players by mistake, either by error in discrimination or because of insufficient knowledge about the strategy of the opponent. It is logical to think that in the case of uncertainty, DD players will always defect with the opponent, as defection is the only NE in the PD. If μ is the probability that a DD player defects even with a PC player, then the payoff of PC and DD players will be as follows:

$$E_{\mathrm{PC}} = [p + r + s(1-\mu)]b - [q + s(\mu)]x - c, \quad (6)$$

$$E_{\mathrm{DD}} = (p + r)b - r(1-\mu)c - r\mu y - d. \quad (7)$$

Note that the payoffs of C and D players (eqs (1) and (2)) will remain unaltered.

In a heterogeneous population consisting of C, D, PC and DD strategies, punishment becomes stable even when the error in discrimination approaches 60%. The tolerance to mistakes, however, is influenced by the initial frequencies of the strategies and the cost of discrimination (d). There is a general tendency towards increase in tolerance to mistakes as the frequency of C and D is less, while frequency of PC and DD is more. This is because of the instability of PC, in the presence of C and D, when DD is either rare or absent.

Also, tolerance to mistakes in discrimination increases as the cost of discrimination decreases. The reason for this trend can be attributed to the fact that PC gets a selective benefit over C and D only in the presence of DD players, as $E[\mathrm{PC}|\mathrm{DD}] > E[\mathrm{C}|\mathrm{DD}]$ and $E[\mathrm{PC}|\mathrm{DD}] > E[\mathrm{D}|\mathrm{DD}]$. So establishment of DD players in the population is an obligatory first step for the success of PC. Thus, as the cost of discrimination d gets smaller, DD players can establish their population rapidly. The decrease in tolerance to mistakes with increase in d is nonlinear and is best described by a polynomial function (Figure 4).

There is always a threshold level of tolerance to errors above which punishment, and in turn cooperation, is unstable in the system. This threshold behaviour can be attributed to the following fact. PC players are always evolutionary stable against DD players because $E[\mathrm{PC}|\mathrm{PC}] > E[\mathrm{DD}|\mathrm{PC}]$ for any $c < y$. However, PC players can invade the population of DD players only if $E[\mathrm{PC}|\mathrm{DD}] > E[\mathrm{DD}|\mathrm{DD}]$, which is fulfilled when $\mu < (b - c + d)/(b + x)$. Thus, for sufficiently small errors, PC players can invade the population of DD players and become stable. However, if the last condition is reversed then both PC and DD are bistable, with an unstable mixed strategy NE, where both types of players coexist. If the errors are high and DD players are common in the population, PC players cannot establish their population. However, because DD players have to pay the cost of discrimination, their population is vulnerable to invasion by D players. As a result, above the threshold level of errors, the population yields to defection.

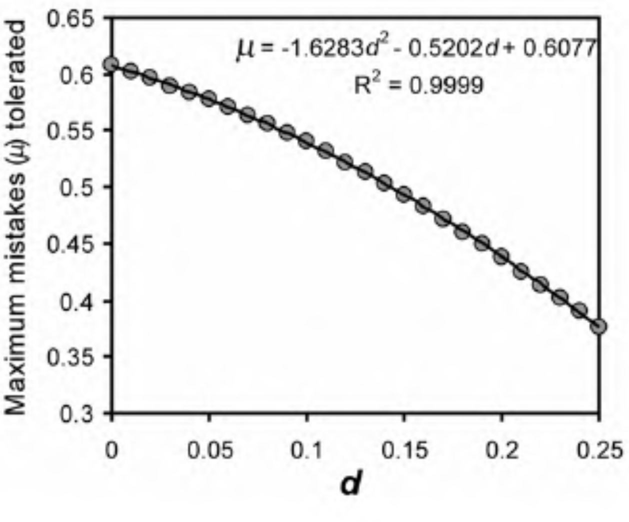

Figure 4. Maximum mistakes (μ) tolerated for different values of cost of discrimination (d) when a DD player defects with a PC player by mistake. The initial frequencies of all the strategies were considered equal. Other parameters were: $b = 1$, $c = 0.3$, $x = 0.1$ and $y = 0.8$.

We will look at the change in the dynamics of the game with increase in mistakes for two scenarios, as we did before. In the first scenario, the cost of discrimination is assumed negligible so that in a population DD will replace D players. Thus, we will consider only three types of players, C, PC and DD. In this game cooperation is still stable when there are 60% mistakes (Figure 5). Nonetheless, PC is not the only NE and as the mistakes increase both PC and C will be stable. The reason for the shift in NE from pure PC to a mixture of PC and C is because of the following reason. When some of the DD players defect with PC players by mistake, PC will pay the cost of punishment. This will give C players a selective benefit over PC players, as they do not punish.

In the second scenario we will consider the cost of discrimination to be significant, so both D and DD players will coexist. However, we can eliminate C players from this population, as they anyway get extinct in the initial few rounds of the game in the presence of two defecting strategies. The dynamics of the game with D, PC and DD strategies in the presence of mistakes depicts that as the mistakes increase, the domain in which punishment and in turn cooperation is stable decreases (Figure 6). However, this decrease is slow with increase in mistakes and the basin of attraction for PC is still bigger than for D players.

In the above analysis we assumed two conditions: (1) DD players have to pay a cost of discrimination even though they fail to discriminate sometimes, and (2) DD players only make mistakes by defecting with a PC player and never by cooperating with other players. We will relax these assumptions below. We can relax the first assumption by replacing d with $d(1 - \mu)$ in eq. (7). The second assumption can be rectified by considering that with the same error rate (μ), a DD player cooperates with C, D and other DD players. The resultant dynamics is qualitatively similar to that discussed above. However, tolerance to mistakes is significantly lesser in this case (Figure 7). If the cost of discrimination is small, the tolerance to mistakes can be up to 30%. Under the modified conditions, PC players will be evolutionary stable against invasion by DD players if $y > c + d$, the basic condition of the model, which is essential for the evolution of the discriminators. Furthermore, PC players will invade the population of DD players if $\mu < (b - c + d)/(2d - c + x + d)$. If μ is large, PC players are unable to invade DD players and owing to the cost of discrimination, they would yield to D players. This demonstrates that both defection and discrimination need to achieve perfection for the stability of cooperation.

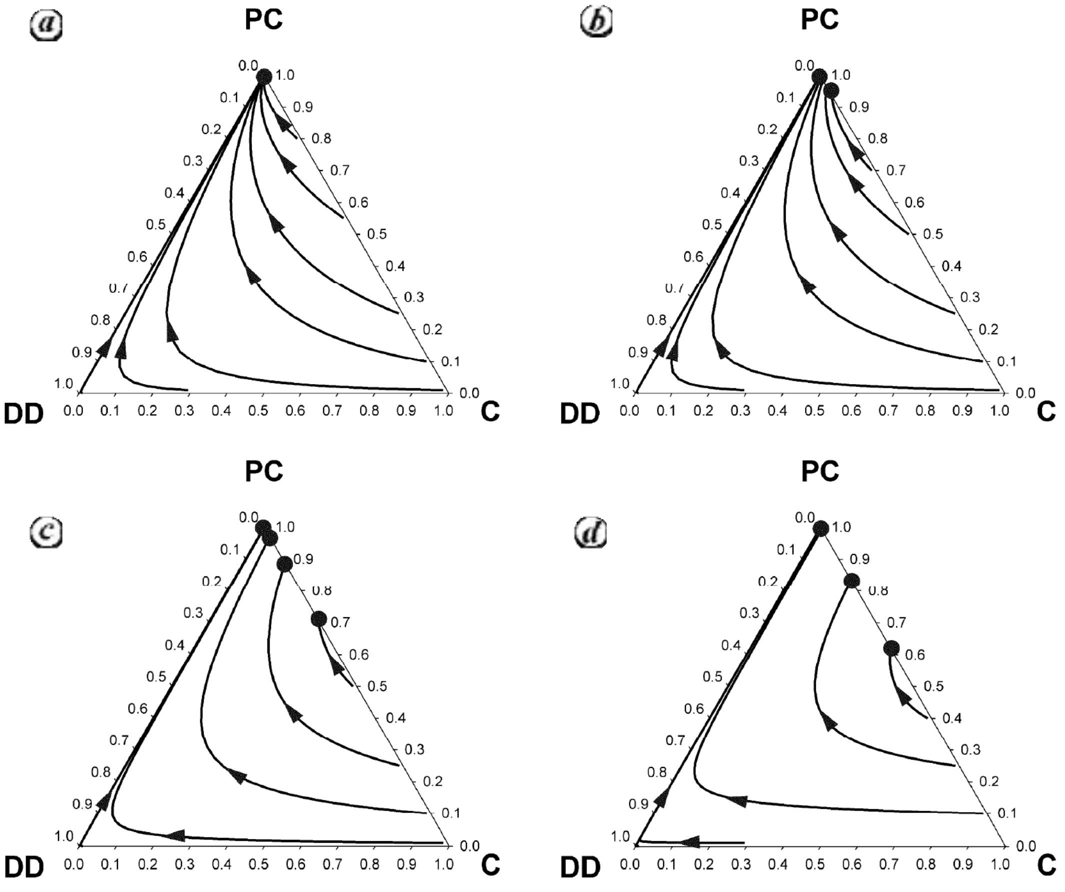

Figure 5. Ternary plot of change in the frequency of players using PC, C and DD strategies, when the cost of discrimination (d) is negligibly small, for different probabilities of mistakes (μ). (***a***) $\mu = 0.01$, (***b***) $\mu = 0.1$, (***c***) $\mu = 0.4$ and (***d***) $\mu = 0.6$. Filled circles are stable NEs. Arrows indicate direction of changing frequencies of individuals using a particular strategy, from their initial frequencies. The dynamics depicts that PC is the only NE when mistakes are rare, while as mistakes increase, both PC and C are stable. Parameters used here are: $b = 1$, $c = 0.3$, $x = 0.1$, $y = 0.8$ and $d = 0.001$.

Discriminating and punishing defectors

So far, we have considered that players use discrete strategies like complete cooperation, discrimination or punishment with certainty. Furthermore, we also considered that only cooperators punish the defectors and norm violators. There is no a priori reason why only cooperators should punish. We will relax both these assumptions by constructing stochastic models.

Consider each player as having a strategy given by a pair of probabilities $(t, w) \in [0, 1]^2$, where t is the probability that the player cooperates (so the player defects with probability $1 - t$) and w is the probability that the player punishes a defector. We can plot all the possible combinations of strategies in a unit square with vertices (0, 0) for defectors who never punish, (0, 1) for defector who punishes, (1, 0) for cooperators who do not punish and (1, 1) for cooperators who punishes. These four vertices are the pure strategies of the game, while all other points in the interior of the unit square and on the borders are mixed strategies. If the population is using a strategy (t_0, w_0), then the payoff of a player adopting a different strategy (t_1, w_1) can be given as:

$$E[(t_1, w_1) \mid (t_0, w_0)] = t_1 t_0 (b - c) + (1 - t_1) t_0 b + t_1 (1 - t_0)(-c) + w_0 (1 - t_1)(-y) + w_1 (1 - t_0)(-x). \quad (8)$$

Keeping one probability in a strategy the same as the population, we can derive differential equations that govern the dynamics of the other probability. Thus, the differential equations that depict the change in t and w can be given as:

$$\dot{t} = -c + wy, \quad (9)$$

$$\dot{w} = -x + tx. \quad (10)$$

A vector field plot of eqs (9) and (10) is shown in Figure 8 *a*. If $w < c/y$, then the derivative of t (eq. (9)) is negative and so t decreases; when t is less than 1, the derivative of w (eq. (10)) is also negative and so the strategy (0, 0) is a

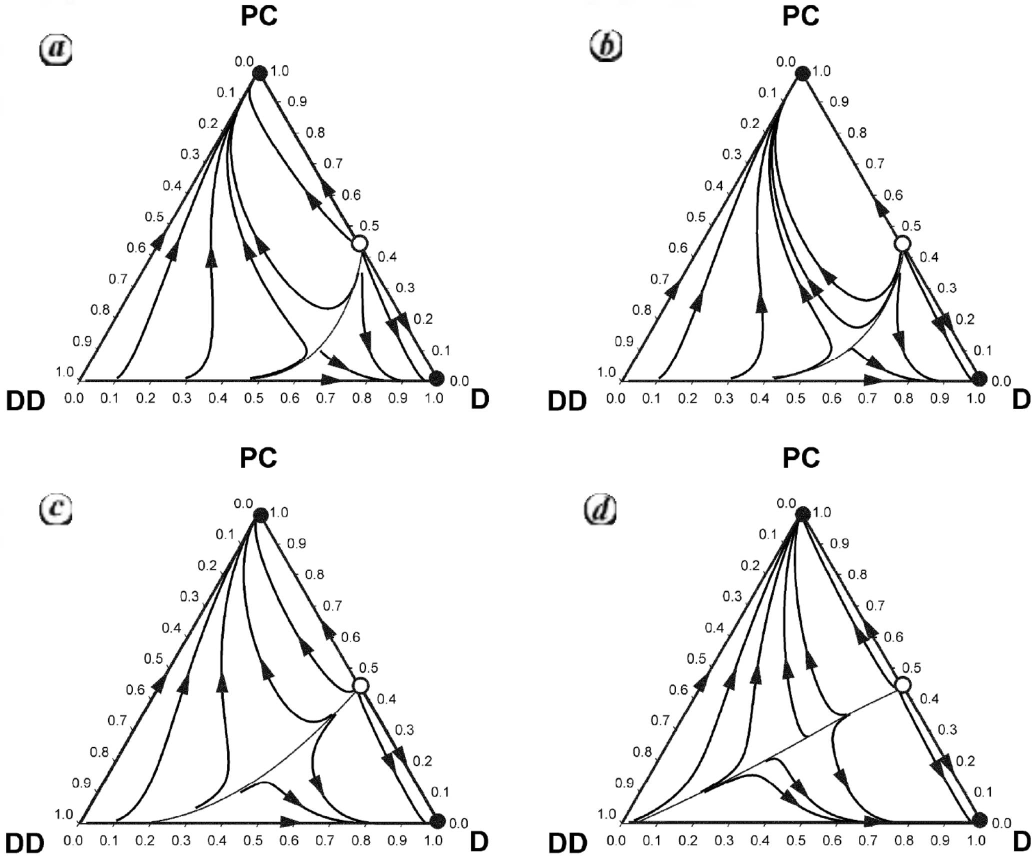

Figure 6. Ternary plot of change in frequency of the players using PC, D and DD strategies, when the cost of discrimination (d) is significant, for different probabilities of mistakes (μ). (***a***) $\mu = 0.01$, (***b***) $\mu = 0.1$, (***c***) $\mu = 0.4$ and (***d***) $\mu = 0.6$. Filled circles are stable NEs, while unfilled circles are unstable NEs. Arrows indicate direction of changing frequencies of individuals using a particular strategy, from their initial frequencies. Parameters used here are: $b = 1$, $c = 0.3$, $x = 0.1$, $y = 0.8$ and $d = 0.05$.

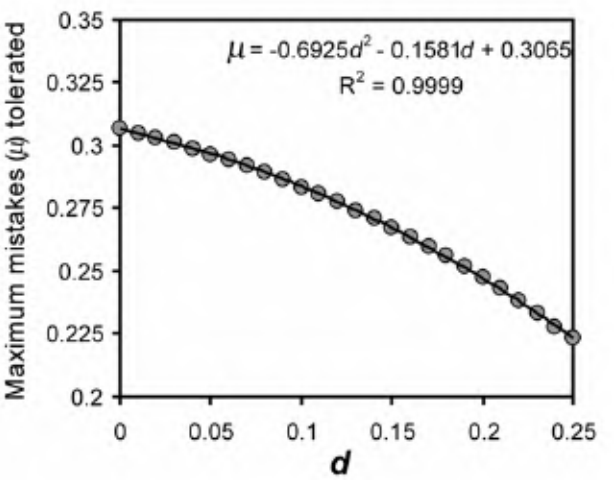

Figure 7. Maximum mistakes (μ) tolerated for different values of cost of discrimination (d). The initial frequencies of all the strategies were considered equal. Other parameters were: $b = 1$, $c = 0.3$, $x = 0.1$ and $y = 0.8$.

stable NE of the game. If $w > c/y$, then the edge defined by the points $(1, c/y)$ and $(1, 1)$ is stable; however, this stability is weak. This is because in the absence of defectors natural drifts can decrease the punishing frequency less than c/y and under these conditions $(0, 0)$ will be the only stable strategy. This argument is similar to that raised in Figure 1, except for the fact that we have considered both cooperators and defectors as being able to punish norm violators.

We will now consider two scenarios similar to what we considered for the deterministic model. In the first scenario, the cost of discrimination is negligibly small and so all the defectors exist as discriminating defectors, while in the second scenario the cost of discrimination is significant and so both unconditional defectors and discriminating defectors coexist.

For the first scenario, let us consider each player to have a strategy given by a pair of probabilities $(u, w) \in [0, 1]^2$, where u is the probability that the player cooperates, $(1 - u)$ the probability that the player discriminates between individuals and cooperates only with punishers and w the probability that the player punishes a defector. The pure strategies at the vertices will be $(0, 0)$ for discriminating defectors who never punish, $(0, 1)$ for discriminating defectors who punish, $(1, 0)$ for cooperators who do not punish, and $(1, 1)$ for cooperators who punish. All other points in the interior of the unit square and on the borders are mixed strategies. If the population is using a strategy (u_0, w_0), then the payoff of a player adopting a different strategy (u_1, w_1) can be given as:

$$\begin{aligned}E[(u_1, w_1) \mid (u_0, w_0)] &= u_1 u_0 (b-c) + u_1(1-u_0)[w_1(b-c) \\ &+ (1-w_1)(-c)] + (1-u_1)u_0[w_0(b-c) \\ &+ (1-w_0)b] + (1-u_1)(1-u_0) \\ &\times [w_1 w_0 (b-c) + (1-w_1)w_0(-c) \\ &+ w_1(1-w_0)b] + (1-u_1)(-d).\end{aligned} \quad (11)$$

The derivates which describe the change in the probabilities u and w can be given as:

$$\dot{u} = -c + wc + d, \quad (12)$$

$$\dot{w} = b - ub. \quad (13)$$

A vector field plot of eqs (12) and (13) is shown in Figure 8 *b*. The edge described by the points $(1, 1 - d/c)$ and $(1, 1)$ is an evolutionary stable NE of the game and is a global attractor. When $d \to 0$, there is only one point $(1, 1)$, which is evolutionary stable. In any case, discriminating defectors stabilize punishment in the system.

In the second scenario, let us consider that each player has a strategy given by a pair of probabilities $(v, w) \in [0, 1]^2$, where v is the probability that the player is a discriminating defector $(1 - v)$ the probability that the player is a unconditional defector and w the probability that the player punishes a defector. The pure strategies at the vertices will be (0, 0) for defectors who never punish, (0, 1) for defectors who punish, (1, 0) for discriminating defectors who do not punish, and (1, 1) for discriminating defectors who punish. All other points in the interior of the unit square and on the borders are mixed strategies. If the population is using a strategy (v_0, w_0), then the payoff of a player adopting a different strategy (v_1, w_1) can be given as:

$$\begin{aligned}E[(v_1, w_1) \mid (v_0, w_0)] &= (1-v_1)w_0(-y) + (1-v_0)w_1(-x) \\ &+ (1-v_1)v_0 w_1 b + v_1(1-v_0)w_0(-c) \\ &+ v_1 v_0 [w_1 w_0 (b-c) + w_1(1-w_0)b \\ &+ (1-w_1)w_0(-c)] + v_1(-d).\end{aligned} \quad (14)$$

The derivates which describe the change in the probabilities v and w can be given as,

$$\dot{v} = wy - wc - d, \quad (15)$$

$$\dot{w} = -x + vx + vb. \quad (16)$$

A vector field plot of eqs (15) and (16) is shown in Figure 8 *c*. There are two evolutionary stable NEs at (1, 1) and (0, 0) and two unstable NEs at $(0, d/(y-c))$ and $((x/(b+x)), 0)$. If punishment frequency is less then $d/(y-c)$ and discrimination frequency is less then $x/(b+x)$, then defection without punishment (0, 0) is the only stable strategy. However, as cost of discrimination (d) decreases and cost of cooperation (x) decreases, and the area in which defection is stable and punishment does not evolve becomes negligibly small. Thus, the equilibrium (1, 1)

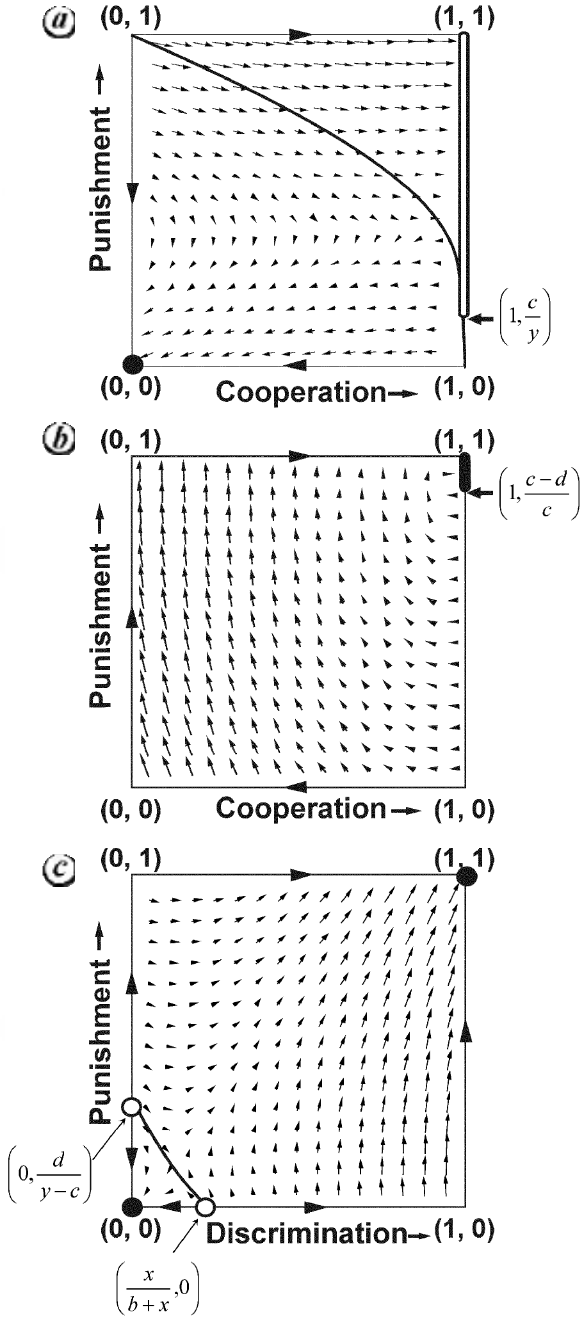

Figure 8. Vector field plots of stochastic strategy game. Plot of eqs (9) and (10) (***a***), eqs (12) and (13) (***b***) and eqs (15) and (16) (***c***). For further details see text. Unfilled bars and circles are unstable NEs, while filled bars and circles are stable NEs. Parametric values are the same as in Figure 3.

has a bigger basin of attraction. An interesting finding of this analysis is that, as all the discriminators punish, all discriminators cooperate with each other to avoid punishment at the equilibrium (1, 1). As a result (1, 1) has stable cooperation in the system, but ironically no individuals are primarily cooperators.

Discussion

Punishment in PD is unstable in the presence of non-punishing cooperators. In models where only cooperators are allowed to punish as well as models where even defectors can punish, both punishers and cooperators are stable above a threshold value of punishers (Figures 1 and 8 *b*). However, this stability is weak since at this stage selection is neutral for both punishers and cooperators and natural drift can drive the punishment below this threshold level. Similar arguments have been raised for cooperation based on direct[32] and indirect[5] reciprocity. The only possible solution to this problem will be a selective advantage to punishers over cooperators and defectors. We recognize that discriminating individuals can give such selective advantage to the punishers. These discriminating defectors can spontaneously emerge in a population of cooperators, punishers and defectors, if the cost of discrimination is low and the cost of cooperation as well as the penalty paid by punished defector is sufficiently large.

Our discriminating defectors are quite different from other discriminators considered in the evolutionary game theory earlier. In the earlier models of indirect reciprocity or reputation, cooperators discriminate and selectively cooperate with a reputed cooperator[33]. This is paradoxical because first, in terms of rational game theory, why should players care about the reputation of others beyond their own payoff, and secondly, why should they decrease their own reputation by withholding help from those who are less reputed? Also, cooperating with a cooperator is not a rational choice (or a NE) in a PD interaction. Our discriminator, on the contrary, is self-concerned and while he defects with other strategies, he only cooperates with the punishers to avoid getting punished.

A major essence of our model is that all the strategies have purely selfish motivations and yet the population evolves cooperation through stability of punishment. In our game, discriminating defectors make the act of punishment non-altruistic. Punishers punish a defector at a cost to themselves, but in return gain reputation and cooperation from discriminating defectors who are threatened by the punishment. This is in contrast to altruistic punishment discussed by other workers[6,12,24].

Public knowledge of the strategies is a key factor of our model. A discriminating defector can gather information in a variety of ways, including eavesdropping, gossiping or a network of communication, but irrespective of the way we have assumed that discrimination is costly. A number of models so far that have incorporated discriminating strategies[4,5,21,33] have neglected such a cost. In the present model the mode of gathering information is not trivial. Rather, the limitations of such information and mistakes in discrimination are of more concern, especially in the cases of gossiping and a network of communication, where the trustworthiness of the orator is in question. We have shown that the model is fairly robust against limitations and mistakes in discrimination (Figures 4 and 7).

If DD players cooperate with players other than PC players by mistake, up to a critical error rate, the dynamics of our game is not qualitatively affected. If they cooperate too often, PC players cannot invade the population of DD players, and since the population of DD players is always vulnerable to invasion by unconditional defectors, cooperation is fleeting under this condition. It is ironical that cooperating too often leads to the collapse of cooperation. This scenario further bolsters our argument that refinement in defection and discrimination is essential for the evolution of cooperation. Our results further suggest that indirect reciprocity[4,5] alone is an unlikely candidate for the evolution of cooperation.

At a second level we have relaxed the assumptions that players can use only discrete strategies and that only cooperators can punish. Even when these conditions are relaxed, we show that discriminating defectors help in the evolution of punishment behaviour (Figure 8 *b* and *c*). Recent studies advocated the maintenance of stable cooperation in the population in the presence of selfish punishers who defect in the main game, but punish those who have defected[26–28]. However, just giving the defector a chance of punishment did not stabilize punishment in our game dynamics (Figure 8 *a*). We observed that punishment was only weakly stable and defection without punishment was the only stable NE of the game. Nonetheless, addition of discriminating defectors stabilized punishment in all the scenarios. Furthermore, in a subgame where we considered the presence of only defectors, discriminating defectors and punishers were present, and discriminating defectors who punished were stable with a bigger basin of attraction (Figure 8 *c*). This situation is the most interesting because there are no cooperators in the system. However, since discriminating defectors cooperate with the punishers and since all the discriminating defectors are punishers at this equilibrium, stable cooperation is maintained in the system without classical 'cooperators'.

Conclusion

The human social system, which is based on cooperation in genetically unrelated individuals and often in a large population, is better explained by strong reciprocity in terms of punishment to a defector as a more promising mechanism than kin selection and direct reciprocity. However, how costly punishment becomes stable is an enduring evolutionary conundrum. Contemporary work that explains the stability of altruistic punishment requires special conditions. Our argument based on discriminating and punishing defectors is a simple, yet practical mechanism that has the potential to explain the stability of punishment and cooperation in human social systems. The essence of our system is that all the strategies have a selfish foundation and yet it evolves cooperation for a wide set of conditions. In our extended PD game punishment is not altruistic, since in the presence of discriminating defectors, punishers have a high probability of getting cooperation. Another attractive feature of our findings is that the punishment (and in turn cooperation) is highly robust against mistakes in discrimination. Thus, in human interactions where public knowledge of strategies employed by other players is possible, even at a small cost to self, discriminating defectors can stabilize punishment in a wide set of conditions.

1. Hamilton, W. D., The genetic evolution of social behavior I and II. *J. Theor. Biol.*, 1964, **7**, 1–52.
2. Trivers, R. L., The evolution of reciprocal altruism. *Q. Rev. Biol.*, 1971, **46**, 35–57.
3. Axelrod, R. and Hamilton, W. D., The evolution of cooperation. *Science*, 1981, **211**, 1390–1396.
4. Nowak, M. and Sigmund, K., Evolution of indirect reciprocity by image scoring. *Nature*, 1998, **393**, 573–577.
5. Panchanathan, K. and Boyd, R., A tale of two defectors: the importance of standing for evolution of indirect reciprocity. *J. Theor. Biol.*, 2003, **224**, 115–126.
6. Fehr, E. and Gachter, S., Altruistic punishment in humans. *Nature*, 2002, **415**, 137–140.
7. Fehr, E. and Fischbacher, U., The nature of human altruism. *Nature*, 2003, **425**, 785–791.
8. Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H. and Mcelreath, R., In search of *Homo economicus*: behavioral experiments in 15 small-scale societies. *AEA Papers Proc.*, 2001, **91**, 73–78.
9. Henrich, J. *et al.*, Costly punishment across human societies. *Science*, 2006, **312**, 1767–1770.
10. Fehr, E., Fischbacher, U. and Gächter, S., Strong reciprocity, human cooperation and the enforcement of social norms. *Human Nat.*, 2002, **13**, 1–25.
11. Bowles, S. and Gintis, H., The evolution of strong reciprocity: cooperation in heterogeneous populations. *Theor. Popul. Biol.*, 2004, **65**, 17–28.
12. Boyd, R., Gintis, H., Bowles, S. and Richerson, P. J., The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. USA*, 2003, **100**, 3531–3535.
13. Dahanukar, N. and Ganguly, B., Punish but don't tit: an evolutionary stable strategy in iterated Prisoners' Dilemma. *Logic. Biol.*, 2004, **4**, 60–67.
14. Gardner, A. and West, S. A., Cooperation and punishment, especially in humans. *Am. Nat.*, 2004, **164**, 753–764.
15. Gintis, H., Strong reciprocity and human sociality. *J. Theor. Biol.*, 2000, **206**, 169–179.
16. Henrich, J. and Boyd, R., Why people punish defectors. *J. Theor. Biol.*, 2001, **208**, 79–89.
17. Bowles, S. and Gintis, H., *Homo reciprocans*. *Nature*, 2002, **415**, 125–128.
18. Boyd, R. and Richerson, P., Punishment allows the evolution of cooperation (or anything else) in sizable groups. *Ethol. Sociobiol.*, 1992, **13**, 171–195.
19. Kendal, J., Feldman, M. W. and Aoki, K., Cultural coevolution of norm adoption and enforcement when punishers are rewarded or non-punishers are punished. *Theor. Popul. Biol.*, 2006, **70**, 10–25.
20. Gintis, H., Smith, E. A. and Bowles, S., Costly signalling and cooperation. *J. Theor. Biol.*, 2001, **213**, 103–119.
21. Panchanathan, K. and Boyd, R., Indirect reciprocity can stabilize cooperation without the second-order free rider problem. *Nature*, 2004, **432**, 499–502.
22. Milinski, M., Semmann, D. and Krambeck, H. J., Reputation helps solve the 'tragedy of the commons'. *Nature*, 2002, **415**, 424–426.
23. Hauert, C., De Monte, S., Hofbauer, J. and Sigmund, K., Volunteering as red queen mechanism for cooperation in public goods games. *Science*, 2002, **296**, 1129–1132.
24. Fowler, J. H., Altruistic punishment and the origin of cooperation. *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 7047–7049.
25. Hauert, C., Traulsen, A., Brandt, H., Nowak, M. A. and Sigmund, K., Via freedom to coercion: the emergence of costly punishment. *Science*, 2007, **316**, 1905–1907.
26. Sigmund, K., Hauert, C. and Nowak, M. A., Reward and punishment. *Proc. Natl. Acad. Sci. USA*, 2001, **98**, 10757–10762.
27. Nakamaru, M. and Iwasa, Y., The coevolution of altruism and punishment: role of the selfish punisher. *J. Theor. Biol.*, 2006, **240**, 475–488.
28. Eldakar, O. T., Farrell, D. L. and Wilson, D. S., Selfish punishment: altruism can be maintained by competition among cheaters. *J. Theor. Biol.*, 2007, **249**, 198–205.
29. Eldakar, O. T. and Wilson, D. S., Selfishness as second-order altruism. *Proc. Natl. Acad. Sci. USA*, 2008, **105**, 6982–6986.
30. Hofbauer, J. and Sigmund, K., *Evolutionary Games and Population Dynamics*, Cambridge University Press, 1998.
31. Gintis, H., *Game Theory Evolving*, Princeton University Press, Princeton, NJ, 2000.
32. Boyd, R. and Lorberbaum, J. P., No pure strategy is evolutionary stable in a repeated Prisoner's dilemma game. *Nature*, 1987, **327**, 58–59.
33. Nowak, M. A. and Sigmund, K., Evolution of indirect reciprocity. *Nature*, 2005, **437**, 1291–1298.

ACKNOWLEDGEMENTS. We thank Richa Batra for initial discussions during the development of the model. We also thank the two anonymous reviewers for their critical comments on an earlier draft of the manuscript.

Received 28 July 2008; revised accepted 5 February 2009