Direct reciprocity with costly punishment: Generous tit-for-tat prevails

David G. Rand, Hisashi Ohtsuki,

Martin A. Nowak

Journal of Theoretical Biology

256 (2009) 45–57

Evolutionary Game Theory

- Evolutionary game theory considers a population of players interacting in a game.
- Individuals have fixed strategies, they interact randomly with other individuals.
- The payoff of all these encounters are added up. Payoff is interpreted as fitness, and success in the game is translated into reproductive success.
- **Strategies that do well reproduce faster and strategies that do poorly are outcompeted. This is straightforward natural selection.**

Nash equilibrium

- **A situation in which neither of the players can improve his payoff by a unilateral change of strategy is a Nash equilibrium.**

- The Nash equilibrium for non repeated Prisoner's dilemma (1: defect/2: defect).

- Once a Nash equilibrium has been reached no player has a reason to deviate from his strategy- even if another state would provide a higher payoff for both players.

Tit-for-tat (TFT)

For the repeated prisoners dilemma the winning strategy was Tit-for-tat (TFT). TFT starts with cooperation and then does whatever the opponent did in previous round.

TFT will answer C for C and D for D. Playing against TFT is like playing the mirror image of yourself shifted by one round. TFT was invented by Anatol Rapoport.

Generous Tit-for-tat (GTFT)

- **GTFT strategy cooperates whenever the opponent has cooperated, but also cooperates one out of three times when the opponent has defected.**

- When one GTFT individual plays another, each receives an average payoff per round that is very close to the full reward for mutual cooperation, R. In contrast, two TFT players only obtain (R+P+T+S)/4.

- **GTFT can correct mistakes.** With the certain probability, a sequence of cooperation and defection leads back to mutual cooperation. The expected payoff for two GTFT players is higher than for two TFT players.

The costly punishment

- The standard model for direct reciprocity is the repeated Prisoners Dilemma, where in each round players choose between cooperation and defection.

- It is possibly to include third choice costly punishment, so at each round players have choice between cooperation, defection, and costly punishment.

- Analyzed in the paper were **reactive strategies: where behavior depends on what player did in the previous round.**

- All cooperative strategies which are Nash equilibria were identified and confirmed by numerical simualtions.

The key comparison between cost of cooperation and cost of punishment

- The essential is relation between **cost of cooperation c** and **cost of punishment α.**

- **If the cost of cooperation is greater then cost of punishment c >α** the Nash equilibrium is generous-tit-for-tat (GTFT), which does not use costly punishment.

- If the cost of cooperation is less then cost of punishment **c <α**, then there are infinitely many cooperative Nash equilibria and the response to defection can induce costly punishment.

Two key mechanisms of evolution of cooperation in humans

- **Direct reciprocity** *means there are repeated encounters between the same two individuals and behavior of the player depend on the actions of co-player.*

- **Indirect reciprocity** *means there are repeated encounters in a group of individuals, and behavior of the player also depends on the actions of co-players.*

- **Reciprocity** is an unavoidable consequence of small group size, given the cognitive abilities of humans.

Costly punishment is a form of direct or indirect reciprocity

- **It is <u>not possible</u> to consider costly punishment as an independent mechanism.**
- If I punish you because you have defected with me, then I use direct reciprocity.
- If I punish you because you have defected with others, then it is a case of indirect reciprocity.
- Most models of costly punishment use direct or indirect reciprocity.

Altruistic or costly punishment

- Costly punishment is sometimes called **“altruistic punishment”** *because some people use it in the second and last round of a game where they cannot directly benefit from this action in context of experiment.*

- Typically motives of the punishes are not ‘altruistic’ and the strategies instincts of people are mostly formed by situation of repeated games, where they could benefit from their action.

- Costly punishment makes no assumptions about the motive behind the action, so it is more precise term than ‘altruistic punishment’.

- The main idea of this paper is to examine a **hypothesis that costly punishment promote human cooperation.**

Payoff matrix for a repeated Prisoners Dilemma

Direct reciprocity is described by the repeated Prisoner's Dilemma. In each round of the game, two players can choose between cooperation, C, and defection, D. The payoff matrix is given by

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} a_2 & a_4 \\ a_1 & a_3 \end{pmatrix}. \end{array}$$

The game is a Prisoner's Dilemma if $a_1 > a_2 > a_3 > a_4$.

The cost of cooperation in prisoners dilemma

The cooperation means paying cost c for the other person to receive a benefit b.

Defection means either ‘doing nothing’ or gaining payoff d at the cost e for the other person.

In this formulation, the payoff matrix is given by

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} b-c & -c-e \\ d+b & d-e \end{pmatrix} \end{array}.$$

Here b>c>0 and d,e>=0.

The costly punishment in prisoners dilemma

- Including costly punishment means that we have to consider a **third strategy, P**, which has a **cost α** for the actor and a **cost β** for the recipient.
- The 3 x 3 payoff matrix is of the form

$$\begin{array}{c} \\ C \\ D \\ P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} b-c & -c-e & -c-\beta \\ d+b & d-e & d-\beta \\ -\alpha+b & -\alpha-e & -\alpha-\beta \end{pmatrix} \end{array}.$$

Punishment in the repeated Prisoners Dilemma

The classical 'punishment' using Tit-for-tat for defection is defection.

Two questions analyzed in the paper are:

- Is it advantageous to use costly punishment, P, instead of defection, D, in response to a co-player's defection?
- Does costly punishment allow cooperation to succeed in situations where tit-for-tat does not?

The probability of the game to be continued

The 3 x 3 payoff matrix with costly punishment is given in the form

$$\begin{array}{c} \\ C \\ D \\ P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} b-c & -c-e & -c-\beta \\ d+b & d-e & d-\beta \\ -\alpha+b & -\alpha-e & -\alpha-\beta \end{pmatrix} \end{array}.$$

Assumption: b,c, α, β>0 and d,e>=0.

Notation:
The probability w (0<w<1) that the game continues for another round.

(1-w) is the probability that game terminates.

The number of rounds has a geometrical distribution with mean 1/(1-w).

A probabilistic strategy of a player

A 'strategy' of a player is a behavioral rule that prescribes an action in each round.

- Each player has the following probabilistic strategy:
- In the first round, a player chooses an action (either C, D, or P) with probability p_0, q_0 and r_0, respectively.
- From the second round on, a player chooses an action depending on the opponent's action in the previous round.

- The probability that a player chooses C, D, or P, is given by p_i, q_i, and r_i, for each possible previous action (i=1,2,3 for C,D,P) of the opponent.

$$s = \begin{array}{c} \\ \text{Initial move} \\ \text{Response to } C \\ \text{Response to } D \\ \text{Response to } P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} p_0 & q_0 & r_0 \\ p_1 & q_1 & r_1 \\ p_2 & q_2 & r_2 \\ p_3 & q_3 & r_3 \end{pmatrix} \end{array}.$$

Since p_i, q_i, r_i are probabilities, our strategy space is

$$S_3^4 = \prod_{i=0}^{3} \{(p_i, q_i, r_i) \mid p_i + q_i + r_i = 1, p_i, q_i, r_i \geqslant 0\}.$$

Nash equilibrium for the repeated game

Let $u(s_1, s_2)$ represent the expected total payoff of an s_1-strategist against an s_2-strategist. Strategy s is a Nash equilibrium of the repeated game if the following inequality holds for any $s' \in S_3^4$:

$$u(s, s) \geqslant u(s', s).$$

This condition implies that no strategy s' can do better than strategy s against s.

Cooperative Nash equilibrium

- A Nash-equilibrium strategy, s is a cooperative if and only if two s- strategists always cooperate in the absence of errors.
- We search for Nash equilibria of the form:

$$s = \begin{array}{c} \\ \text{Initial move} \\ \text{Response to } C \\ \text{Response to } D \\ \text{Response to } P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ p_2 & q_2 & r_2 \\ p_3 & q_3 & r_3 \end{pmatrix} \end{array}.$$

A strategy cooperative Nash equilibrium without defection

$$s = \begin{array}{c} \\ \text{Initial move} \\ \text{Response to } C \\ \text{Response to } D \\ \text{Response to } P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1-r_2 & 0 & r_2 \\ 1-r_3 & 0 & r_3 \end{pmatrix} \end{array}.$$

The probabilities r_2 and r_3 must satisfy

$$r_2 > \frac{c+d}{w'(b+\beta)}$$

and

$$r_3 = \frac{c-\alpha}{w'(b+\beta)}.$$

w'= w(1-3ε)

A strategy cooperative Nash equilibrium without punishment

$$s = \begin{array}{c} \\ \text{Initial move} \\ \text{Response to } C \\ \text{Response to } D \\ \text{Response to } P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ 1-q_2 & q_2 & 0 \\ 1-q_3 & q_3 & 0 \end{pmatrix} \end{array}.$$

The probabilities q_2 and q_3 must satisfy

$$q_2 = \frac{c+d}{w'(b+e)}$$

and

$$q_3 > \frac{c-\alpha}{w'(b+e)}.$$

w'= w(1-3ε)

A strategy mixed cooperative Nash equilibrium

$$s = \begin{array}{c} \\ \text{Initial move} \\ \text{Response to } C \\ \text{Response to } D \\ \text{Response to } P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ p_2 & q_2 & r_2 \\ p_3 & q_3 & r_3 \end{pmatrix} \end{array}.$$

The probabilities p_i, q_i, and r_i $(i = 2, 3)$ must satisfy

$$bp_2 - eq_2 - \beta r_2 = b - \frac{c + d}{w'}$$

and

$$bp_3 - eq_3 - \beta r_3 = b - \frac{c - \alpha}{w'}.$$

w'= w(1-3ε)

Punishment and cooperation

The following relation for probability w (0<w<1) that the game continues for another round taking into account possible errors w'= w(1-3ε) hold

$$\frac{c+d}{b+\beta} \leqslant w' < \frac{c+d}{b+e}$$

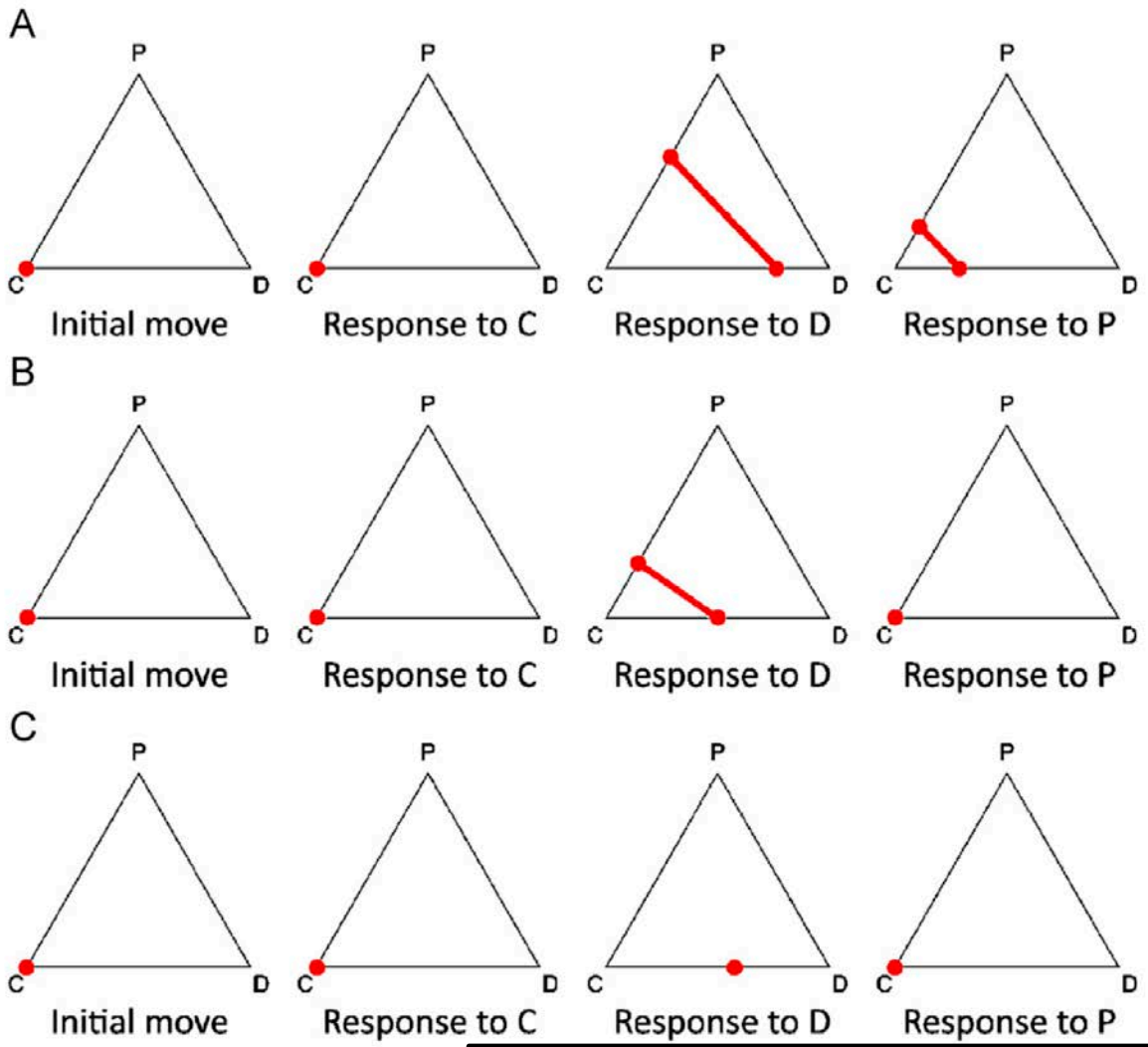
In the classical settings d=e=0 and ε tends to 0;
cost α for the actor and a **cost β** for the recipient.

Cooperation means paying a cost c, for the other person to receive a benefit b.

1. Punishment promote cooperation: α<=c
(Nash equilibrium exist even when cooperation is not beneficial b<c)

2. Punishment does not promote cooperation α> c

A space of strategies



(A) α<=c The best cooperative Nash equilibria when punishment is cheaper than cooperation.

(B) **α<=c** The best cooperative Nash equilibria when punishment is equal in cost to cooperation can use punishment in response to defection, but always cooperates in response to punishment.

(C) **α> c** The best cooperative Nash equilibrium when punishment is more expensive than cooperation is generous tit-for-tat. Only defection and cooperation are used in reaction to defection. Punishment is never used.

Cost of cooperation c and **cost of punishment α.**

Any pair of points from the line in the D-simplex and the line in the P-simplex is a payoff maximizing cooperative Nash equilibrium.

The highest payoff strategies with cooperative Nash equilibria

	Punishment is less costly than cooperation, $\alpha \leqslant c$	Punishment is more costly than cooperation, $\alpha > c$
Initial move	C	C
Response to C	C	C
Response to D	C or D or P Any (p_2, q_2, r_2) that satisfies $q_2 + \frac{b+\beta}{b+e} r_2 = \frac{c+d}{w(1-3\varepsilon)(b+e)}$	C or D $p_2 = 1 - \frac{c+d}{w(1-3\varepsilon)(b+e)}$, $q_2 = \frac{c+d}{w(1-3\varepsilon)(b+e)}$, $r_2 =$
Response to P	C or D or P Any (p_3, q_3, r_3) that satisfies $q_3 + \frac{b+\beta}{b+e} r_3 = \frac{c-\alpha}{w(1-3\varepsilon)(b+e)}$	C

The highest payoff strategies with cooperative Nash equilibrium: classical case d=e=0, ε→0

	Punishment is less costly than cooperation, $\alpha \leqslant c$	Punishment is more costly than cooperation, $\alpha > c$
Initial move	C	C
Response to C	C	C
Response to D	C or D or P Any (p_2, q_2, r_2) that satisfies $q_2 + \frac{b+\beta}{b} r_2 = \frac{c}{bw}$	C or D $p_2 = 1 - \frac{c}{bw}$, $q_2 = \frac{c}{bw}$, $r_2 = 0$
Response to P	C or D or P Any (p_3, q_3, r_3) that satisfies $q_3 + \frac{b+\beta}{b} r_3 = \frac{c-\alpha}{bw}$	C

Stochastic simulation methods

- A game between two players s_1 and s_2 can be described by a Markov process.
- For w = 1, the average payoff per round, $u(s_1, s_2)$, is calculated from the stationary distribution of actions.
- For w<1, the total payoff is approximated by truncating the series after the first 50 terms.
- In our simulations, each player s_i plays a repeated Prisoner's Dilemma with punishment against all other players.
- The average payoff of player s_i is given by

$$\pi_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} u(s_i, s_j).$$

The game between teacher and learner

We randomly sample two distinct players $s^{(T)}$ (Teacher) and $s^{(L)}$ (Learner) from the population, and calculate the average payoffs for each. The learner then switches to the teacher's strategy with probability

$$p = \frac{1}{1 + e^{-(\pi^{(T)} - \pi^{(L)})/\tau}}.$$

This is a monotonically increasing function of the payoff-difference, $\pi^{(T)} - \pi^{(L)}$, taking the values from 0 to 1. This update rule is called the 'pairwise comparison'

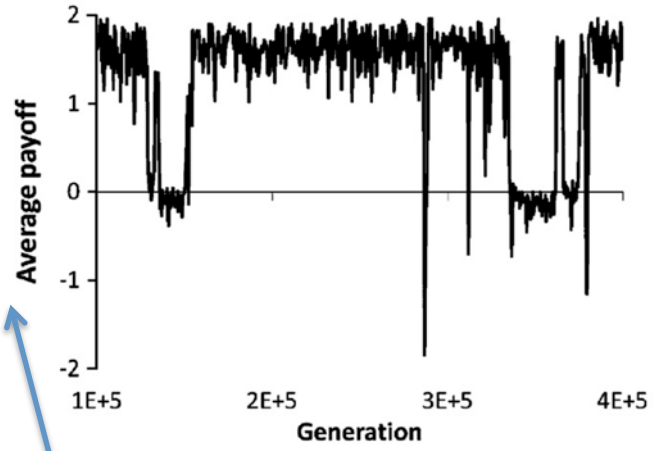
The parameter τ is called the 'temperature of selection'.
It is a measure of the intensity of selection. For very large τ we have weak selection.

In learning, there is a chance of 'mutation' (or 'exploration').
When the learner switches his strategy, then with probability μ he adopts a completely new strategy; hence μ can be interpreted as a mutation rate.

The simulation dynamics in finite populations

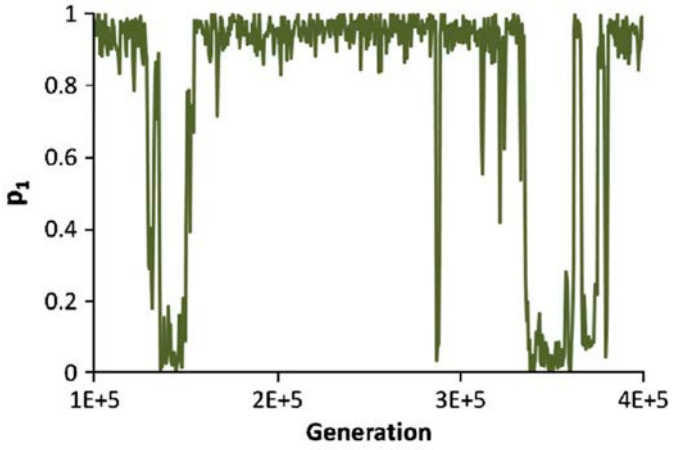
search for Nash equilibria of the form

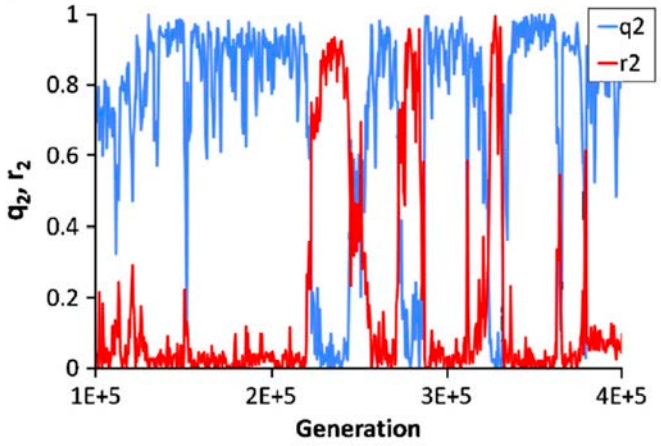
$$s = \begin{array}{c} \\ \text{Initial move} \\ \text{Response to } C \\ \text{Response to } D \\ \text{Response to } P \end{array} \begin{array}{c} \begin{array}{ccc} C & D & P \end{array} \\ \begin{pmatrix} 1 & 0 & 0 \\ 1 & 0 & 0 \\ p_2 & q_2 & r_2 \\ p_3 & q_3 & r_3 \end{pmatrix} \end{array}.$$



$$\pi_i = \frac{1}{N-1} \sum_{\substack{j=1 \\ j \neq i}}^{N} u(s_i, s_j).$$

Payoff values b=3;c=1;d=e=1; α=1, and β=4





Simulated frequencies of the move

Key parameters: **cost of cooperation c** and **cost of punishment α.**

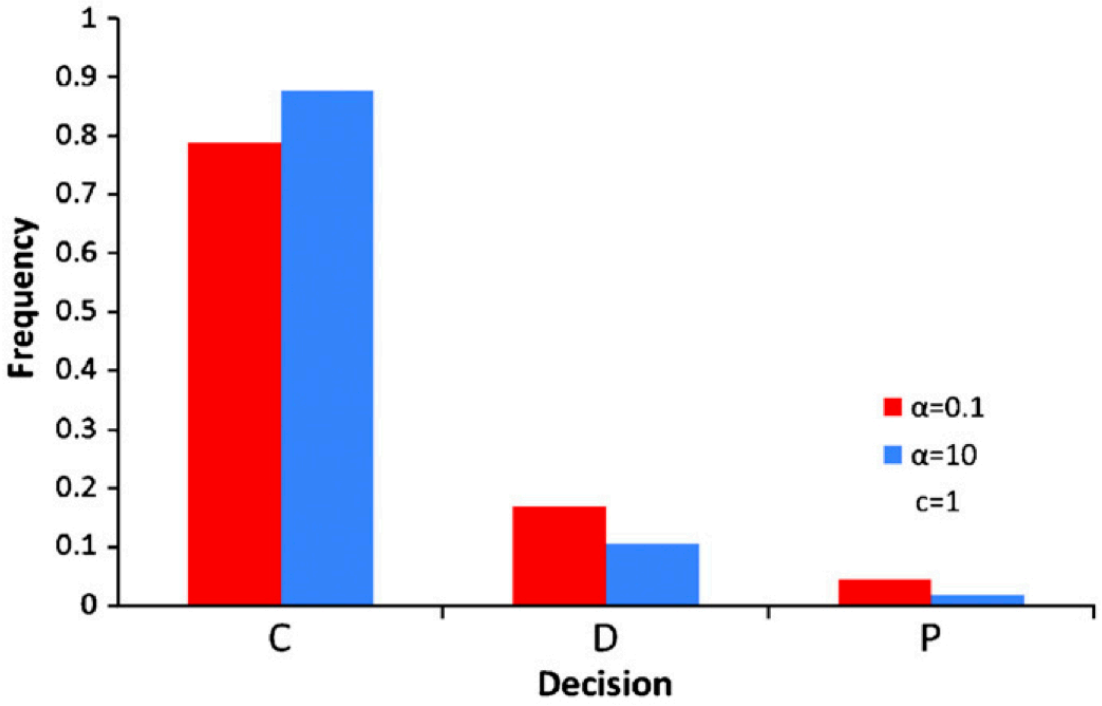


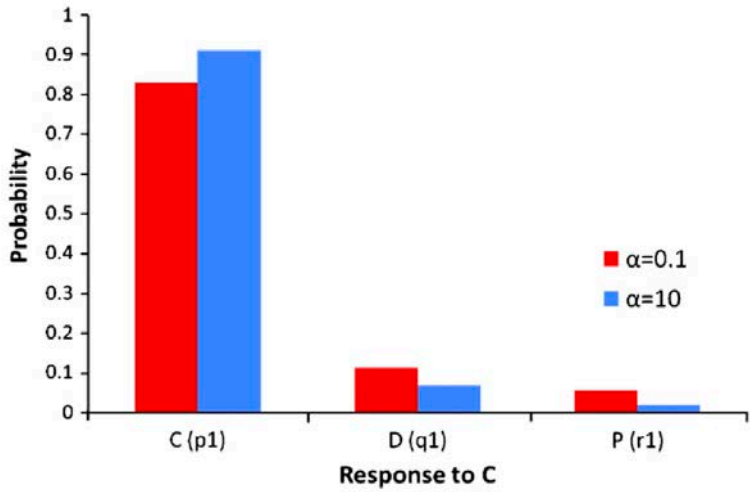
Fig. 3. The relative size of α and c has little effect on move frequencies. The time average frequency of cooperation, defection, and punishment are shown for $\alpha = 0.1$ and $\alpha = 10$, with $b = 3, c = 1, d = e = 1$, and $\beta = 4$. Simulation parameters $\mu = 0.1$, and $\tau = 0.8$ are used. Move use is time averaged over $N = 50$ players, playing for a total of 2×10^7 generations. Consistent with the Nash equilibrium analysis, there is a high level of cooperation in both cases, and the $\alpha > c$ simulation contains slightly more C, less D, and less P than the $\alpha < c$ simulation.

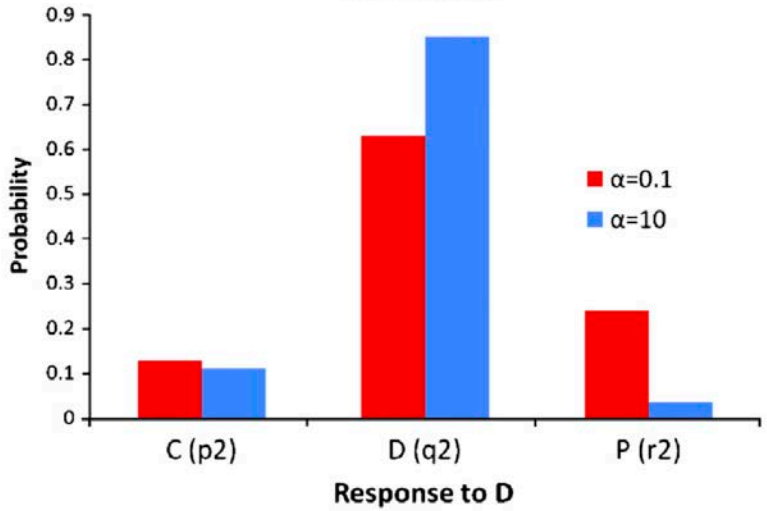
Consistency with the Nash equilibrium analysis:

there is a high level of cooperation for both values of α

then **α>c** simulation contains slightly more C, less D,
and less P than the **α<c** simulation.

Simulated time averages for each strategy

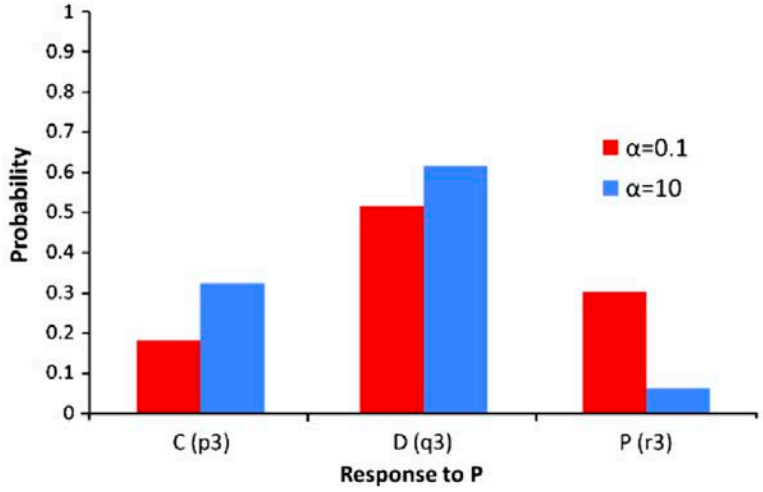




Shown an agreement between the Nash equilibrium analysis and the computer simulations:

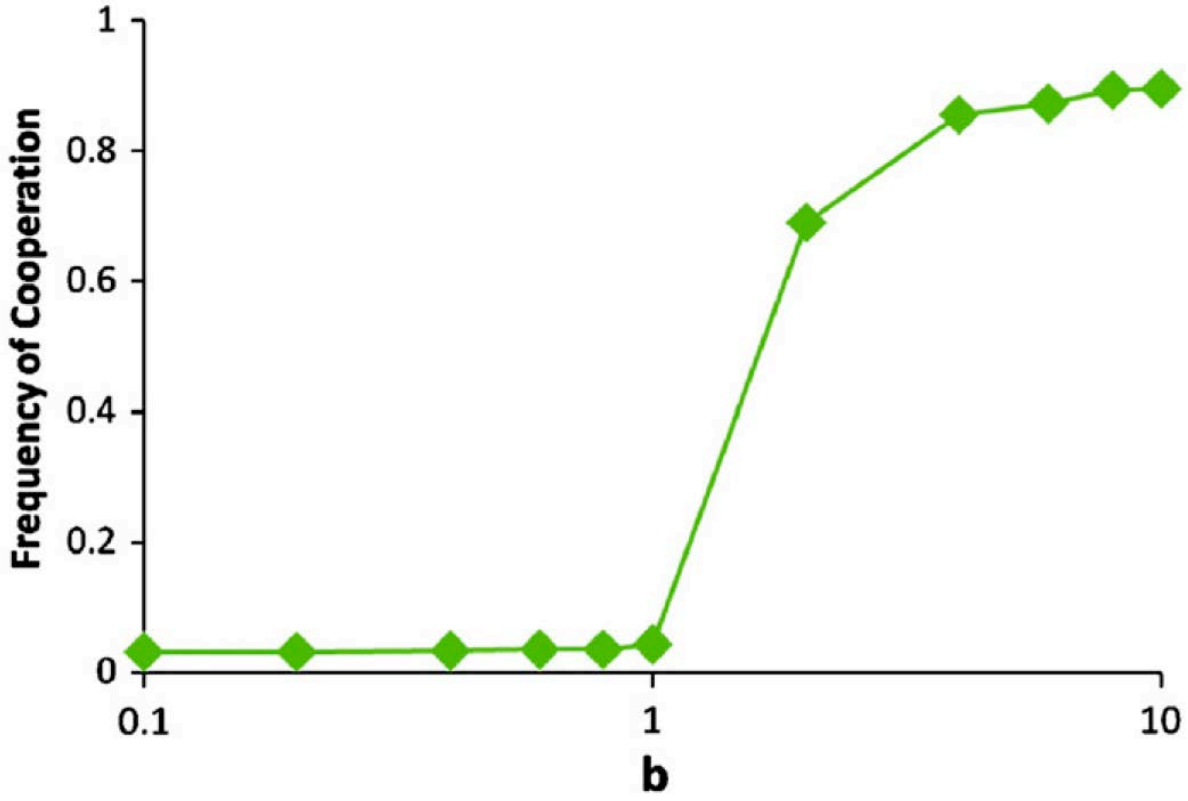
1)on the high level of mutual cooperation regardless of the value of a,
2) the low level of punishment when **α>c.**

The simulations find that even when **α>c**, the response to defection is much more likely to be defection than punishment.



Shown are strategy time averages for **α = 0.1** and **α = 10**, with b = 3; **c= 1**; d = e = 1; and β = 4. Simulation parameters μ = 0.1, and τ = 0.8 are used. Strategies are time averaged over N = 50 players, playing for a total of 2 x 10^7 generations.

Costly punishment does not promote the evolution of cooperation



Costly punishment does not promote the evolution of cooperation. Frequency of cooperation is shown as b is varied, with $c = 1, \alpha = 1, d = e = 1$, and $\beta = 4$. Simulation parameters $\mu = 0.1$, and $\tau = 0.8$ are used. Cooperation frequency is time averaged over $N = 50$ players, playing for a total of 2×10^7 generations. Cooperation is high when $b > 1$, and very low ($< 5\%$) when $b \leqslant 1$.

Contrary to Nash equilibrium analysis:

Costly punishment does not promote the evolution of cooperation

The cooperation means paying a cost, c, for the other person to receive a benefit b.

Cooperation only succeeds in classical direct reciprocity.

The summary of experimental effects of costly punishment on human cooperation

Evolution disfavors the use of costly punishment across a wide range of payoff and simulation parameter values.

The following conclusions come from these simulations:

(A) As punishment becomes more costly for the player who is punished, it becomes more effective to use costly punishment.

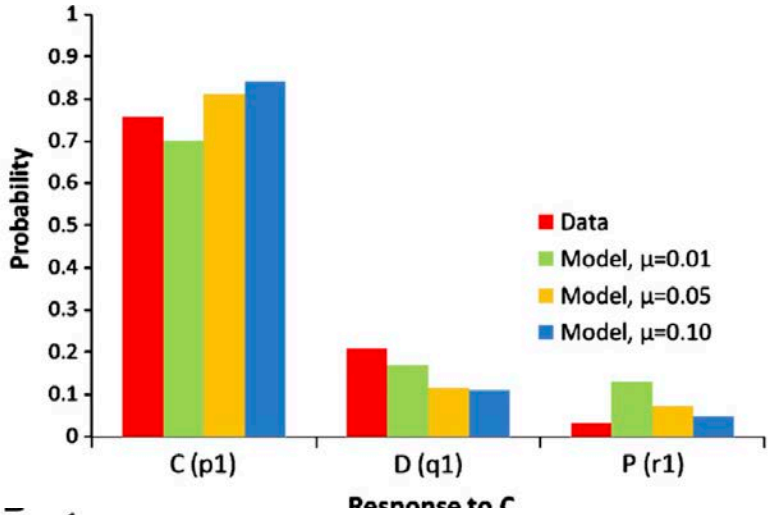
(B) As defection becomes more effective, it makes even less sense to punish.

(C) As punishment gets more expensive for the punisher, it becomes less effective to punish.

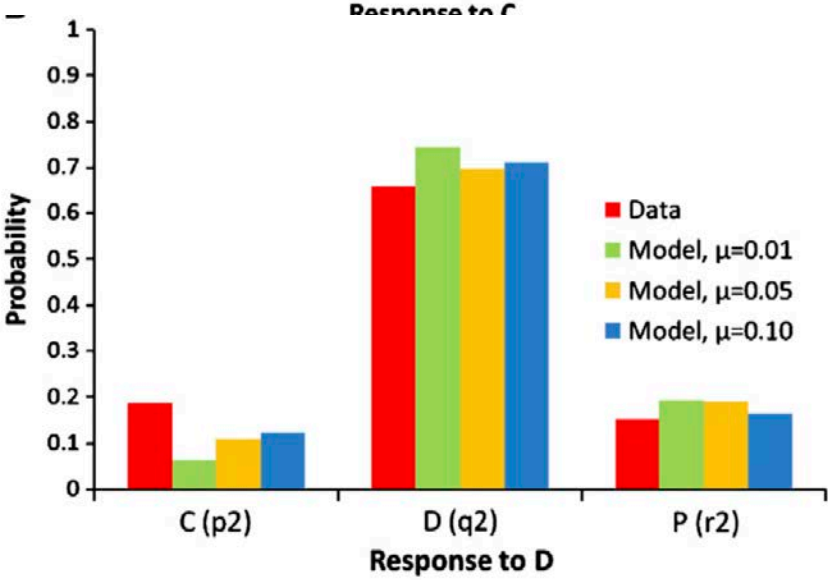
(D) As mutation rate m increases, mutation dominates selection and all move probabilities approach equal level 1/3.

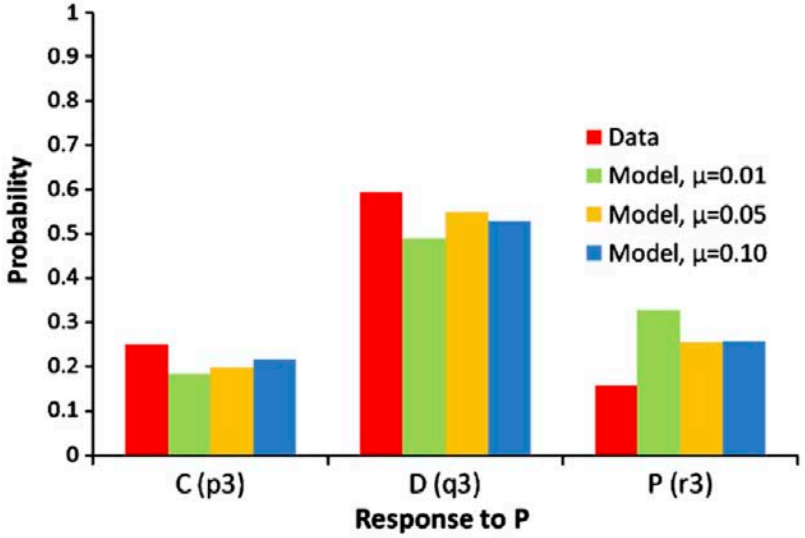
(F) Even in finitely repeated games defection is favored over punishment.

The simulations for effect of costly punishment on human cooperation



1. Defection is used much more often than punishment after the opponent defects or punishes.
2. Cooperation is reciprocated, but unlike in the Nash equilibrium analysis, the computer simulations find that it is uncommon to cooperate after the opponent has defected or punished.





Consider randomly chosen players: teacher and learner. The learner switches his strategy, then with probability μ he adopts a completely new strategy; hence μ can be interpreted as a mutation rate.

Summary: main hypothesis analyzed in the paper

- The main idea of this paper is to examine a **hypothesis that costly punishment promote human cooperation.**

- The approach for the study of costly punishment is to **extend cooperation games from two possible moves, C and D, to three possible moves, C, D, and** P and then study the consequences.

- In order to understand whether costly punishment can really promote cooperation, **one must examine the interaction between costly punishment and direct or indirect reciprocity.**

Summary: two questions for analysis of the extended prisoners dilemma

- Should costly punishment be a response to a co-player defection, instead of defection for defection as in classical direct reciprocity?

- Does the addition of costly punishment allow cooperation to succeed in simulations where direct or indirect reciprocity without costly punishment do not?

Summary: Nash equilibrium

- The essential is relation between **cost of cooperation c** and **cost of punishment α.**

- If **$c<\alpha$** then the only cooperative Nash equilibrium is generous-tit-for-tat, which does not use a costly punishment.

- If **$c>\alpha$** there are infinitely many Nash equilibria and response to defection can be mixture of cooperation, defection and costly punishment. The option for costly punishment allows such cooperative Nash equilibria to exist in parameter regions where there would have been no cooperation in classical direct reciprocity.

Summary: simulations of evolutionary dynamics in finite size populations

- For all parameter choices that were investigated, costly punishment, P, is used less often than defection, D, in response to a co-player's defection.

- Costly punishment fails to stabilize cooperation when cost of cooperation, c, is greater than the benefit of cooperation, b, i.e. c>b

- Therefore, in the context of repeated interactions

(1) natural selection opposes the use of costly punishment,

(2) costly punishment does not promote the evolution of cooperation.

Summary: main results of the paper

- Winning strategies tend to stick with **generous-tit-for-tat and ignore costly punishmen**t, even if the cost of punishment, α, is less than the cost of cooperation, c.

- In the framework of direct reciprocity, **selection does not favor strategies that use costly punishment.**

- **Costly punishment does not promote the evolution of cooperation.**