Social Norm, Costly Punishment and the Evolution to Cooperation

Tongkui, Yu and Shu-Heng, Chen and Honggang, Li

1 February 2011

Social Norm, Costly Punishment and the Evolution to Cooperation

Tongkui Yu[a,b], Shu-heng Chen[b], Honggang Li[a,*]

[a]*Department of systems science, School of Management, Beijing Normal University, Beijing 100875*

[b]*AI-ECON Research Center, Department of Economics, National Chengchi University, Taipei 116*

Abstract

Both laboratory and field evidence suggest that people tend to voluntarily incur costs to punish non-cooperators. While costly punishment typically reduces the average payoff as well as promotes cooperation. Why does the costly punishment evolve? We study the role of punishment in cooperation promotion within a two-level evolution framework of individual strategies and social norms. In a population with certain social norm, players update their strategies according to the payoff differences among different strategies. In a longer horizon, the evolution of social norm may be driven by the average payoffs of of all members of the society. Norms differ in whether they allow or do not allow for the punishment action as part of strategies, and, for the former, they further differ in whether they encourage or do not encourage the punishment action. The strategy dynamics are articulated under different social norms. It is found that costly punishment does contribute to the evolution toward cooperation. Not only does the attraction basin of cooperative evolutionary stable state (CESS) become larger, but also the convergence speed to CESS is faster. These two properties are further enhanced if the punishment action is encouraged by the social norm. This model can be used to explain the widespread existence of costly punishment in human society.

Keywords:
social norm, costly punishment, cooperative evolutionary stable state, attraction basin, convergence speed
JEL: C02, C73, D64

*Corresponding to: hli@bnu.edu.cn

1. Introduction

Cooperation is of utmost importance to human society, and our civilization is based upon the cooperation between genetically unrelated individuals in large groups (Axelrod, 1984). This is obviously true for modern societies with large organizations and nation states, but it also holds for hunter-gathers society with sophisticated forms of hunting, warfare, and food sharing (Fehr and Fischbacher, 2003). While cooperation leads to a tension between what is best for the individual and what is best for the group. A group does better if everyone cooperates, but each individual is tempted to defect. Neither the naive natural selection assumption in biology nor the pure self-interested individual assumption in economics can lead to cooperation directly (Nowak, 2006; Olson, 1965; Ostrom, 2000; Henrich et al., 2005). There must be some specific mechanisms for the emergence of cooperation in a population (Taylor and Nowak, 2007).

Recently, the effect of costly punishment on cooperation has received considerable attention from various disciplines [1]. Costly punishment, which is also called altruistic punishment (Fehr and Gächter, 2002) or sanctioning (Falk et al., 2005) in some literature, means that people have the propensity to incur a cost in order to punish social norm violator (Henrich et al., 2006). It is also a part of strong reciprocity which is a combination of voluntarily cooperation to cooperative, norm-abiding behaviors and punishment to non-cooperative, norm-violating behaviors (Gintis, 2000; Fehr et al., 2002).

In the light of the behavioral experiments (Fehr and Simon, 2000; Fehr and Gächter, 2002; Gürerk et al., 2006; Rockenbach and Milinski, 2006; Henrich et al., 2010) and ethnographic evidence (Knauft et al., 1991; Boehm, 1993), it is no longer the question whether there is costly punishment (Falk et al., 2005). And cross-cultural evidence in complex large-scale and small-scale societies around the globe (Oosterbeek et al., 2004; Henrich et al., 2005,

[1] There are great amount of literature on this topic, for instance: Fehr and Gächter (2002), Boyd et al. (2003), Fowler (2005), Rockenbach and Milinski (2006), Henrich et al. (2006), Henrich (2006), Gürerk et al. (2006) and Ohtsuki et al. (2009) in general journals; Henrich and Boyd (2001), Gintis (2000) and Bowles and Gintis (2004) in biological journals; Fehr and Fischbacher (2004) in cognitive science journals; Ostrom et al. (1992), Fehr and Simon (2000), Andreoni et al. (2003), Falk et al. (2005) and Bochet et al. (2006) in economic and other social science journals.

2006; Marlowe et al., 2008) suggests that the punishment to selfish behavior is a "human universal" (Gächter and Herrmann, 2009).

But the role of costly punishment in promoting cooperation is ambiguous. In behavioral experiments, costly punishment has been shown to effectively enforce cooperation (Fehr and Gächter, 2002; Fehr and Fischbacher, 2003; Gürerk et al., 2006; Rockenbach and Milinski, 2006). While some other experiments indicate that punishment is less efficient for the pay-off destroyed through punishment exceeds the gains from increased cooperation (Dreber et al., 2008; Milinski and Rockenbach, 2008; Egas and Riedl, 2008; Wu et al., 2009). And a theoretical work of Ohtsuki et al. (2009) argues that costly punishment can not lead to an efficient equilibrium in most situations, and efficient strategy is to withhold help for defectors rather than punishing them.

Then another question arises naturally: what does the costly punishment exist for if it provides little efficiency? Similar questions are proposed by Dreber et al. (2008) that "*costly punishment ... require a mechanism for its evolution*" and Milinski and Rockenbach (2008) that "*costly punishment remains one of the most thorny puzzles in human social dilemmas*".

One possible source of such a puzzle may be that these studies only focus on a short period of experience and omit the long history of culture evolution. Obviously, the experimental works can only get the spot performance of subjects with cultivated culture, but not the process of culture cultivation of subjects. And the analytical work of Ohtsuki et al. (2009) also only analyzes the equilibrium (i.e. the Cooperative Evolutionary Stable State, CESS) but not the route to the equilibrium to give the conclusion that costly punishment is mostly less efficient. If we turn our attention to the states far away from the equilibrium and study the route of co-evolution of the social norms and individual strategies, the costly punishment may play a different role in promoting cooperation.

We try to give a possible explanation within a two-level evolution framework of individual strategies and social norms, by extending the model in Ohtsuki et al. (2009). And we pay more attention to the evolution route instead of the stationary state to investigate the role of punishment. Consider a world with many societies competing for resource, territory, dominance or even the opportunity to survive. Each society has a social norm which is used to assign reputations to individuals based on their employed actions. Norms may allow or do not allow for the punishment action as part of strategies, and, for the former, they may encourage or do not encourage the punishment action. Within a society, individuals interact to each other and they have three action choices: cooperation, defection or punishment. An individual can take an action according to the opponent's reputation. At the same time, based on the applied social norm, a new reputation is assigned to this individual. This reputation will determine what action others will take to him.

Individuals in each society interact to each other under the applied social norm, and they learn and update their strategies to get higher individual payoff. This within-society competition determines how often different strategies are used in the society.

In a longer horizon, due to the between-society competition, societies may evolve their social norms by comparing the average payoff of all the social members that different social norms can provide. Such social norm evolution may take the form of social transformation, civil war, external war, colonization, etc. This two-level evolution framework shares the same idea with the culture group selection of Bergstrom (2002) and Henrich (2004), for social norm can be regarded as a kind of culture (Young, 2008).

In such a two-level evolution framework, individual strategy is the adaptation to the social norm, and ultimately the survived social norm determines how often the punishment actions are taken. If the fittest social norm is the one with punishment option or even that encouraging punishment, individuals from such culture background will naturally exhibit the tendency of punishment. This argument is also supported by the cross-culture experiments which demonstrate that punishment are substantially shaped by the cultural background across a range of diverse societies (Gächter and Herrmann, 2009).

Instead of combining two levels of evolution into one equation as in Henrich (2004), we only model the dynamics of the evolution of individual strategies under several fixed social norms. This is because that there are totally 64 (or 16) types for social norms with (or without) punishment option, and only a few of them can foster cooperation (Henrich, 2004). We select three typical social norms including non-punishment, punishment-optional and punishment-provoking social norm, and explicitly model the evolutionary dynamics of individual strategies under these three social norms. By comparing the cooperation ratio and average payoff in the dynamics under different social norms, we can get more clear insight of the driving force of the evolution of the social norm.

It is found that costly punishment does contribute to the evolution toward cooperation. Once individuals has the choice of punishment, not only does the attraction basin of cooperative evolutionary stable state (CESS) become larger, but also the convergence speed to CESS is faster in the social norms with punishment option. These two properties are further enhanced if the punishment action is encouraged in the punishment-provoking social norm.

This result implies that costly punishment is necessary in at least two situations.The first is that when there are too many defectors in a society, it will be stuck in a social dilemma that defection is the best choice for each individual in non-punishment or even punishment-optional social norm and it can only struggle out of the social dilemma by encouraging punishment to provide individuals the incentive to punish and cooperate. The second is that when the society is not patient enough and wishes to reach the highly cooperative state quickly, a social norm with punishment or even encouraging punishment can increase the speed approaching to cooperative evolutionary stable state.

The remainder of the paper is organized as follows. In the next section, we give the model of evolutionary donor-recipient game, which differs from the work of Ohtsuki et al. (2009) in that we explicitly model the dynamics of the strategy evolution. In the third section, we compare the attraction basin of CESS and the converge speed to the CESS in three social norms. The last section contains the conclusion and discussions.

2. Model

2.1. Donor recipient game

At each small time interval Δt, a fraction $2\Delta t$ of players is randomly sampled from an society of large population to form pairs. In each pair, one player acts as a donor and the other player as a recipient. The donor has two basic behavioral choices: cooperation (C), defection (D). Cooperation involves a cost c for the donor and a benefit b for the recipient. Defection has no cost and yields no benefit. A donor may also have the choice of punishment (P) in some social norms. Punishment has cost α for the donor and cost β for the recipient. Here c, b, α and β are all positive real number. Each individual is endowed with a binary reputation, which is either good (G) or bad (B). The donor can base his decision on the recipient's reputation. After each interaction, the reputation of the donor is updated according to the 'social norm' of the population, while the reputation of the recipient remains the same. The reputation update process is susceptible to errors. With probability μ, where $0 \leq \mu \leq 0.5$, an incorrect reputation is assigned.

With probability $1-\mu$ the correct reputation is assigned. All individuals come to the same conclusion; there are no private lists of reputation.

2.1.1. Strategies

Each player has an action rule (or strategy), s, which depends on the recipient's reputation. A player with an action rule s takes the action $s(G)$ toward a good recipient, and the action $s(B)$ toward a bad one. Each of $s(G)$ and $s(B)$ can be either 'C', 'D', or 'P'. For social norms without punishment option, there are $2^2 = 4$ possible action rules: $s(G)s(B) = CC, CD, DC, DD$. For social norms with punishment available, there are $3^2 = 9$ possible action rules: $s(G)s(B) = CC, CD, CP, DC, DD, DP, PC, PD$, and PP. In the present work, we only study four of these strategies, 'CC', 'CD', 'CP' and 'DD', but not DC, DP, PC, PD and PP, because they are odd and not feasible for study the emergence of cooperation.

2.1.2. Social norms

A social norm n is used for updating the reputations of players. A donor who has taken the action $X(X = C, D, P)$ toward a recipient whose reputation is $J(J = G, B)$ is assigned the new reputation $n(J, X)(= G, B)$ by the social norm n. Social norms of this type are based on 'second-order assessment', and they depend on both the action of the donor and the reputation of the recipient Nowak and Sigmund (2005). Figure 1 gives three typical social norms we will study in this work with the related ordinary strategies.

In non-punishment social norm 'GGBG' as figure 1(a), individuals have no choice of punishment, and a donor can only cooperate or defect. Cooperators to both good and bad recipients are assigned a good reputation. Defectors to bad recipient are also assigned a good reputation. Defectors to good recipient are assigned a good reputation.

In punishment-optional social norm 'GGBGBG' as figure 1(b), individuals have the choice of punishment, and the punishment to a bad recipient will gain a good reputation. But the punishment is just an optional action to bad recipient, because donors defect to a bad recipient without any cost is also assigned a good reputation.

In punishment-provoking social norm 'GGBBBG' as figure 1(c), a donor who defects to a bad recipient is assigned a bad reputation, this encourage a donor to take either cooperation or defection action to bad recipient. So this norm is a more punishment-provoking.

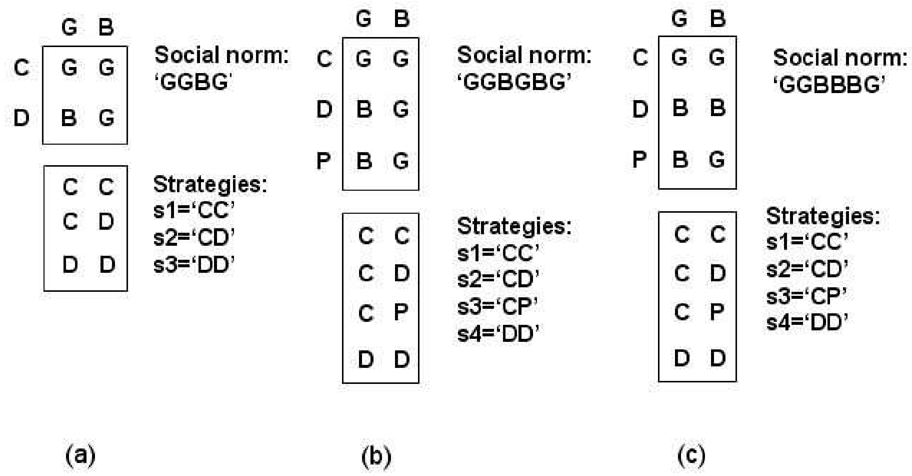


Figure 1: Typical social norms (a. non-punishment, b. punishment-optional and c. punishment-provoking social norm) with the related ordinary strategies. A social norm is used to update the donor's reputation taking into account both the donor's action (cooperation, C, defection, D, or punishment, P) and the recipient's reputation (good, G, or bad, B). A strategy specifies for the donor to cooperate, C, defect, D, or punish, P against a recipient, whose reputation is either good, G, or bad, B. In non-punishment (a) norm, individuals have no choice of punishment. In punishment-optional (b) norm, individuals have the choice of punishment, but punishment is not encouraged because donors can also get a good reputation by defecting to a bad recipient. In punishment-provoking (c) norm, defection to a bad recipient leads to a bad reputation, individuals can only cooperate or punish to bad recipients to get a good reputation, so punishment is more encouraged.

We will explicitly model the dynamics of individual strategies under these three social norms to study the role of punishment in promoting cooperation.

2.2. Evolutionary dynamics of strategies

In a society with social norm n, individuals interact to each other. Each of them has his own strategy that specifies what action he will take to recipients with good or bad reputation. Once a donor takes an action, a new reputation is assigned to him according to the social norm n. And this reputation will determines what action others may take to him.

Individuals will learn and update their strategies to get a higher individual payoff by imitating the better strategy. Sometimes a player is given an opportunity to change his strategy. He randomly samples a player and compares the difference in payoffs. If a sampled player has a greater payoff then the sampling player imitates the sampled player's strategy with probability proportional to the difference in payoffs. Otherwise a sampling player remains the same strategy. So the expected payoff of a strategy can be interpreted as its fitness and strategies with higher fitness have more chance to reproduce.

In this model, the payoff of a strategy rely on not only the relative abundance of the strategies but also the fraction of individuals with good reputation in the society. Because the reputation of individuals is ever changing, so it is hard to give a proper calculation of the payoff of a strategy. In the similar situation, Ohtsuki and Iwasa (2007) calculate the expected payoff of a strategy as the discounted total payoff along the infinitely long future of reputation evolution with the initial reputation of all individuals is good. But there are two problem for this method: the first is that the calculation of the payoff along the infinitely long future is based on the fixed strategies frequency, while individual strategy is also evolving although relatively slowly; the second is that arbitrarily assigning a good initial reputation to all individuals is not suitable, for individual reputation should be inherited from one period to the next period.

Fortunately, we find that for a fixed relative abundance of the strategies in a society, the reputation distribution (the frequencies of individuals with good and bad reputation) will converge to a stable state quickly, because that the reputation is the instantaneous result of a donor's action and the distribution of the action rules (strategies) are fixed.

Because the pace of agents' strategy updating is much lower than reputation dynamics, we assume that individuals will update their strategy only

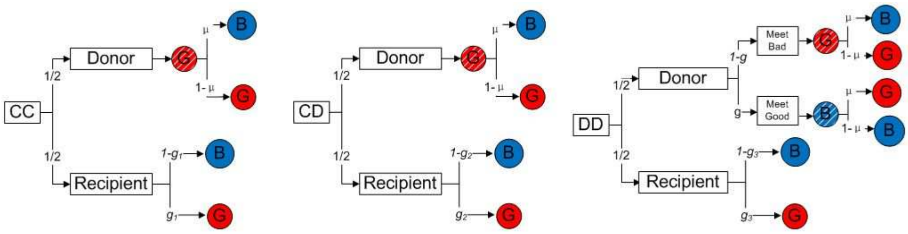


Figure 2: Reputation dynamics of individuals taking different strategies in punishment-provoking (GGBBBG) social norm.

after they can actually conceive the payoff of different strategies by a sufficiently long time during which individual reputation distribution converges to the stable state. So we calculate a strategy's expected payoff in the stable reputation distribution as the fitness measurements. The following two subsections provide the calculation of the stable reputation distribution and the expected strategy payoff in the fixed strategy distribution.

2.2.1. Stable reputation distribution

A stable reputation distribution can be derived given that the frequency of strategies taken by all players in the whole society is fixed. In 'GGBG' norm, a fraction x_1 ($0 \leq x_1 \leq 1$) of players take strategy 'CC', and x_2, x_3 ($0 \leq x_2, x_3 \leq 1$) of players take 'CD' and 'DD' strategy. Here $x_1+x_2+x_3 = 1$. And the ratios of players with good reputation in 'CC', 'CD' and 'DD' players are denoted by g_1, g_2 and g_3 respectively. Thus the ratio of players with good reputation in entire population is $g = x_1 g_1 + x_2 g_2 + x_3 g_3$.

Figure 2 gives the reputation dynamics of 'CC', 'CD' and 'DD' players. A 'CC' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action no matter what reputation the recipient has, and this tends to make him a good reputation. Due to the assignment error, he gets a good reputation with a probability $1-\mu$ and bad reputation with probability μ. The 'CC' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as the current frequency g_1. So the new frequency of good reputation 'CC' players is $g_1^{'} = \frac{1}{2}g_1 + \frac{1}{2}(1-\mu)$.

Similarly we get the new frequency of good reputation 'CD' and 'DD' players are $g_2^{'} = \frac{1}{2}g_2 + \frac{1}{2}(1-\mu)$ and $g_3^{'} = \frac{1}{2}g_1 + \frac{1}{2}(1-g)(1-\mu) + \frac{1}{2}g\mu$. Since $g = x_1 g_1 + x_2 g_2 + x_3 g_3$, we can solve the linear recursion and get the stable reputation frequency of each strategy $g_1^* = g_2^* = 1-\mu$, $g_3^* = (1-\mu)[1-$

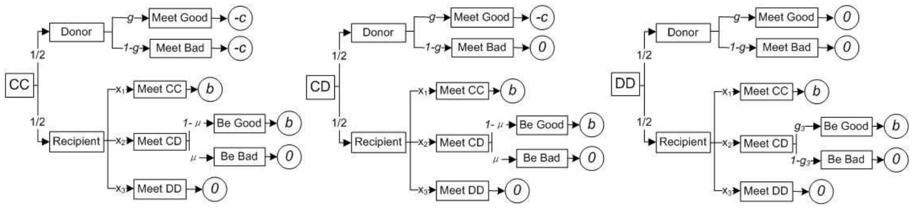


Figure 3: The calculation of expected payoff of strategies in non-punishment social norm.

$\frac{1-2\mu}{1+(1-2\mu)x_3}]$. And the total good reputation frequency is $g^* = \frac{1-\mu}{1+(1-2\mu)x_3}$.

For 'GGBGBG' norm, we can also get the stable reputation frequency of 'CC' 'CD' and 'CP' players $g_1^* = g_2^* = g_3^* = 1 - \mu$, and stable reputation frequency of 'DD' players is $g_4^* = (1-\mu)[1 - \frac{1-2\mu}{1+(1-2\mu)x_4}]$. The total good reputation frequency is $g^* = \frac{1-\mu}{1+(1-2\mu)x_4}$.

For 'GGBBBG' norm, the stable reputation frequency of 'CC' and 'CP' players are $g_1^* = g_3^* = 1-\mu$, and that of 'CD' players is $g_2^* = \mu+(1-s\mu)g^*$ and 'DD' players $g_4^* = \mu$. The total good reputation frequency is $g^* = \frac{(1-\mu)(x_1+x_3)}{1-(1-2\mu)x_2}$.

Detailed process to get the stable reputation frequency of given strategies frequency in the three norms is provided in the appendix A.

2.2.2. Fitness measurements of strategies

We calculate a strategy's expected payoff in the stable reputation distribution as the fitness measurements driving the strategy evolvement.

The calculation expected payoff of 'CC', 'CD' and 'DD' strategy in non-punishment (GGBG) social norm is illustrated in figure 3.

For a 'CC' player, he has $\frac{1}{2}$ chance to be a donor and cooperate with a cost c. With another $\frac{1}{2}$ chance being a recipient, he meets a 'CC' , 'CD' and 'DD' player with probability x_1, x_2 and x_3 and is expected to get b, $(1-\mu)b$ and 0 revenue respectively. So the expected revenue of strategy 'CC' is $p_1 = \frac{1}{2}(-c) + \frac{1}{2}[bx_1 + bx_2(1-\mu)]$. Similarly, the expected revenue of strategy 'CD' and 'DD' can also be calculated. The expected revenue of all three strategies in 'GGBG' social norm are

$$\begin{cases} p_1 = \frac{1}{2}(-c) + \frac{1}{2}[bx_1 + bx_2(1-\mu)] \\ p_2 = \frac{1}{2}g(-c) + \frac{1}{2}[bx_1 + bx_2(1-\mu)] \\ p_3 = \frac{1}{2}(0) + \frac{1}{2}(bx_1 + bx_2 g_3) \end{cases} \tag{1}$$

.

For 'GGBGBG' norm, we can also get the expected revenue of strategy 'CC', 'CD', 'CP' and 'DD' as

$$\begin{cases} p_1 = \frac{1}{2}(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1-\mu)] \\ p_2 = \frac{1}{2}g(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1-\mu)] \\ p_3 = \frac{1}{2}g(-c) + \frac{1}{2}(1-g)(-\alpha) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1-\mu)] \\ p_4 = \frac{1}{2}[bx_1 + b(x_2 + x_3)g_4] + \frac{1}{2}x_3(1-g_4)(-\beta) \end{cases} \quad (2)$$

.

For ‘GGBBBG’ norm, we can also get the expected revenue of strategy ‘CC’, ‘CD’, ‘CP’ and ‘DD’ as

$$\begin{cases} p_1 = \frac{1}{2}(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1-\mu)] \\ p_2 = \frac{1}{2}g(-c) + \frac{1}{2}x_3(1-g_2)(-\beta) + \frac{1}{2}[bx_1 + bg_2(x_2 + x_3)] \\ p_3 = \frac{1}{2}g(-c) + \frac{1}{2}(1-g)(-\alpha) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1-\mu)] \\ p_4 = \frac{1}{2}[bx_1 + b(x_2 + x_3)\mu] + \frac{1}{2}x_3(1-\mu)(-\beta) \end{cases} \quad (3)$$

.

Detailed process to get the expected revenue of strategies in three norms is provided in the appendix B.

2.2.3. Replicator dynamics of strategies frequency

We model the strategies frequency dynamics as the replicator equation (Hofbauer and Sigmund, 1998), given by $\dot{x}_i = x_i(p_i - \bar{p})$, where $\bar{p}$ is the average payoff in the entire society, defined as $\bar{p} = \sum(x_i p_i)$. Here $i = 1, 2, 3$ for ‘GGBG’ norm and $i = 1, 2, 3, 4$ for ‘GGBGBG’ and ‘GGBBBG’ norm. These differential equations are defined on the simplex $S_3 = \{(x_1, x_2, x_3)|x_1 + x_2 + x_3 = 1, x_i \geq 0\}$ for ‘GGBG’ norm and $S_4 = \{(x_1, x_2, x_3, x_4)|x_1 + x_2 + x_3 + x_4 = 1, x_i \geq 0\}$ for ‘GGBGBG’ norm and ‘GGBBBG’ norm. Each corner of the simplex is an equilibrium of the dynamics corresponding to a monomorphic population. Because only the relative size of payoff matters in this dynamics, so we can shift the expected payoff value additively without altering the dynamics at all.

For non-punishment (GGBG) social norm, we define $p_1^{'} = p_1 - p_3$, $p_2^{'} = p_2 - p_3 =$ and $\bar{p} = x_1 p_1^{'} + x_2 p_2^{'}$ using the corresponding p_i $(i = 1, 2)$ in equation 1.

So the dynamical system of strategies frequency in non-punishment (G-

GBG) social norm is

$$\begin{cases} \dot{x}_1 = x_1(p_1^{'} - \bar{p}) = -cx_1 + cx_1^2 \\ \qquad + \dfrac{[(1-2\mu)b + c]x_1x_2 - (1-2\mu)bx_1x_2(x_1+x_2)}{2 - \frac{1-2\mu}{1-\mu}(x_1+x_2)} \\ \dot{x}_2 = x_2(p_2^{'} - \bar{p}) = cx_1x_2 \\ \qquad + \dfrac{-cx_2 + [(1-2\mu)b + c]x_2^2 - (1-2\mu)bx_2^2(x_1+x_2)}{2 - \frac{1-2\mu}{1-\mu}(x_1+x_2)} \end{cases} \tag{4}$$

.

For punishment-optional (GGBGBG) and punishment-provoking (GG-BBBG) social norm, we define $p_1^{'} = p_1 - p_4$, $p_2^{'} = p_2 - p_4$, $p_3^{'} = p_3 - p_4$ and $\bar{p} = x_1p_1^{'} + x_2p_2^{'} + x_3p_3^{'}$ with the corresponding p_i $(i = 1, 2, 3)$ in equation 2 and 3 respectively. The dynamical systems of strategies frequency in both norms can be derived with the formula 5.

$$\begin{cases} \dot{x}_1 = x_1(p_1^{'} - \bar{p}) \\ \dot{x}_2 = x_2(p_2^{'} - \bar{p}) \\ \dot{x}_3 = x_3(p_3^{'} - \bar{p}) \end{cases} \tag{5}$$

And the detailed expressions of for both dynamical systems are provided in Appendix C.

3. Analysis

3.1. Equilibrium analysis in three social norms

Some analytically accessible results about the existence and stability of equilibriums of the strategy dynamics in three social norms are collected in the following propositions.

Proposition 1. *In all three social norms, the state that all players take 'DD' strategy ($x_1 = 0$, $x_2 = 0$ for 'GGBG' norm and $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ for 'GGBGBG' and 'GGBBBG' norm) is the evolutionary stable state.*

The economic intuition of proposition 1 is that if all other agents take 'DD' strategy, the best choice for one agent is the take 'DD' strategy in all three social norms. Because whatever his reputation is, he will get defection as recipient for all others take 'DD' strategy. Without a incentive to get a good reputation, his best choice is defection which has no cost at all. So in such a evolutionary stable state, there are totally no cooperation in the society, and this is the so called social dilemma. These states are illustrated by the points DD in figure 4(a), 4(b), and 4(c) respectively.

Proposition 2. *In 'GGBG' norm, the state that all players take 'CD' strategy ($x_1 = 0$, $x_2 = 1$) is evolutionary stable on the condition that* $\frac{c}{(1-2\mu)b} < 1$.

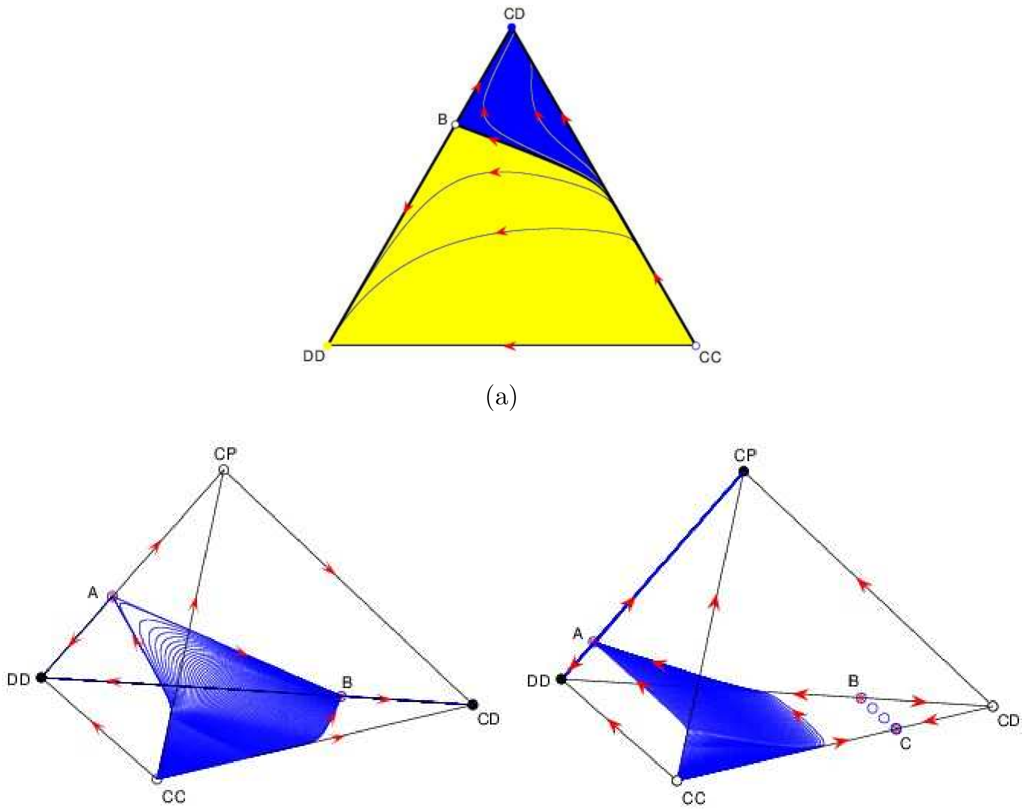
Rearranging the condition we get $\frac{c}{b} < 1 - 2\mu$. This implies that if the social resolution $(1 - 2\mu)$ exceeds the cost-benefit ratio $(\frac{c}{b})$, cooperation can prevail in a society with non-punishment social norm. Because we assume that with probability μ $(0 < \mu < 1/2)$, an incorrect reputation is assigned, $1 - 2\mu$ represents the social resolution to individual reputation. In the state that all other players take 'CD' strategy, the best choice of one agent is the take 'CD' strategy. To be cooperated as a recipient, an agent must have a good reputation as all others take 'CD' strategy. To get a good reputation, one have to cooperate good recipients and cooperate or defect bad recipients, and best choice to bad recipients is to defect because defection has no cost. In such a equilibrium, most interactions are cooperation, so it is a cooperative evolutionary stable state (CESS). This state is illustrated by point CD in figure 4(a).

Proposition 3. *In 'GGBGBG' norm, the state that all players take 'CD' strategy ($x_1 = 0$, $x_2 = 1$, $x_3 = 0$) is an evolutionary stable state on the condition that* $\frac{c}{(1-2\mu)b} < 1$.

Similar to non-punishment norm, the economic intuition is that if all other agents take 'CD' strategy, the best choice for one agent is the take 'CD' strategy. To be cooperated as a recipient, an agent must have a good reputation as all others take 'CD' strategy. The only difference is that an individuals have another choice of punishment, but it is seldom used. This state is also a cooperative evolutionary stable state (CESS) as illustrated by point CD in figure 4(b).

Proposition 4. *In 'GGBBBG' norm, the state that all players take 'CP' strategy ($x_1 = 0$, $x_2 = 0$, $x_3 = 1$) is an evolutionary stable state on the condition that* $c < \alpha$ *and* $\frac{(1-\mu)c+\mu\alpha}{(1-2\mu)(b+\beta)} < 1$.

This state is illustrated by point CP in figure 4(c). The economic intuition is that if all other agents take 'CP' strategy, the best choice for one agent



(a)

(b)

(c)

Figure 4: Phase portrait of three social norms (a. non-punishment, b. punishment-optional and c. punishment-provoking social norm) with $b = 3$, $c = 2$, $\alpha = 1$, $\beta = 4$ and $\mu = 0.02$. Each vertex represents a state with individuals taking the same corresponding strategy, such as points DD in all three norms represent the state that all individuals taking 'DD' strategy. The arrows indicate the evolution direction. In 'GGBG' norm (a), the blue part is the attraction basin of cooperative evolutionary stable state CD and the yellow part is the attraction basin of evolutionary stable state DD. The separatrix line is the stable manifold of saddle point B. State CC is unstable. In 'GGBGBG' norm (b), a separatrix surface which is the stable manifold of saddle point B divides the phase space into to two parts. The part over the surface is the attraction basin of cooperative evolutionary stable state CD and the nether part is the attraction basin of stale state DD. State CC and CP are unstable. In 'GGBBBG' norm (c), there is also a separatrix surface dividing the phase space into to two parts. And the upper and nether regions are the attraction basin of state CD and CP respectively. State CC and CD are unstable. The line connecting B and C consists of Lyapunov stable equilibriums denoted by circles. The attraction basin ratio of cooperative stable state are 15%, 60% and 81% for non-punishment, punishment-optional and punishment-provoking social norm.

is the take 'CP' strategy. To be cooperated as a recipient, an agent must have a good reputation as all others take 'CP' strategy. To get a good reputation, one have to cooperate good recipients and cooperate or punish bad recipients, and best choice to bad recipients is to punish because the cost of cooperation is larger than punishing ($c < \alpha$). Most interactions in such a equilibrium are cooperation, and it is also a cooperative evolutionary stable state (CESS). Comparing the stability condition of cooperative evolutionary stable state in three social norms, we can find that 'CP' state (the state that all players take 'CP' strategy) in punishment-provoking social norm has the broader stability condition than 'CD' state (the state that all players take 'CD' strategy) in non-punishment and punishment-optional social norm whenever the cost of cooperation (c) is less than the cost of punishment (α). If $c < \alpha$, $\frac{(1-\mu)c+\mu\alpha}{(1-2\mu)(b+\beta)} < \frac{c}{(1-2\mu)b}$, because $c > (1-\mu)c + \mu\alpha$ in the numerator and $(1-2\mu)b < (1-2\mu)(b+\beta)$ in the denominator. Moreover, if $\frac{(1-\mu)c+\mu\alpha}{(1-2\mu)(b+\beta)} < 1 < \frac{c}{(1-2\mu)b}$, 'CP' state in punishment-provoking social norm is stable, whereas 'CD' state in non-punishment and punishment-optional social norm is unstable.

Proofs to these propositions can be found in Appendix D.

3.2. Attraction basin ratio of CESS in three social norms

The attraction basins of a cooperative evolutionary stable states (all players taking 'CD' strategy in 'GGBG' and 'GGBGBG' norm or all players taking 'CP' strategy in 'GGBBBG' norm) are the sets of all initial strategy distributions in feasible domain that converge to the CESS. We will investigate the ratio of the extent of the CESS attraction basin to that of the entire feasible domain. For the lack of analytical tools, we rely heavily on the numerical method to calculate the extent of the attraction basin of CESS. Figure 4 gives the phase portraits of three norms with the same typical parameter setting $b = 3$, $c = 2$, $\alpha = 1$, $\beta = 4$ and $\mu = 0.02$).

In the phase portrait of 'GGBG' norm of figure 4(a), the point CD (representing the state that all players takes 'CD' strategy, $x_1 = 0$, $x_2 = 1$) and the point DD (representing the state that all players takes 'DD' strategy, $x_1 = 0$, $x_2 = 0$) are both stable equilibrium as mentioned above. And CD is the cooperative evolutionary state. Point B ($x_1 = 0$, $x_2 = \frac{1}{1-2\mu}\frac{c}{b}$) is a saddle node whose unstable manifold is along the CD-DD line, and the stable manifold constitutes the separatrix line dividing the plane into two regions. The blue region is the basin of attraction of cooperative evolutionary state

CD, points lying in this region converges to CD points. The ratio of CESS attraction basin is the extent of blue area over that of the entire feasible domain, i.e. the area of triangle CC-CD-DD. Under this parameter setting, the ratio of CESS attraction basin is about 0.15.

The phase portrait of 'GGBGBG' norm in a simplex-4 is given in figure 4(b). CD ($x_1 = 0$, $x_2 = 1$, $x_3 = 0$) and DD ($x_1 = 0$, $x_2 = 0$, $x_3 = 0$) are the stable equilibrium and CD is the CESS. Points A ($x_1 = 0$, $x_2 = 0$, $x_3 = \frac{1}{1-2\mu}\frac{(1-\mu)(\alpha+c)}{(1-\mu)(b+\beta)+\alpha}$) is a saddle node with a one-dimensional stable manifold and a two-dimensional unstable manifold. Points B ($x_1 = 0$, $x_2 = \frac{1}{1-2\mu}\frac{c}{b}$, $x_3 = 0$) is a saddle node with a one-dimensional unstable manifold along CD-DD line and a two-dimensional stable manifold which constitutes the separatrix surface dividing the simplex into two parts. The region over the surface is the basin of attraction of cooperative evolutionary state CD which is about 0.60 under this parameter setting. We can easily notice that even punishment just as an optional action, it can widen the attraction basin of the cooperative evolutionary stable state.

The phase portrait of 'GGBBBG' norm in a simplex-4 is given in figure 4(c). CP ($x_1 = 0$, $x_2 = 0$, $x_3 = 1$) and DD ($x_1 = 0$, $x_2 = 0$, $x_3 = 0$) are the asymptotically stable equilibrium and CP is the CESS. In the facet of CC-CD-DD, the line connecting B ($x_1 = 0$, $x_2 = \frac{1}{1-2\mu}\frac{c}{b}$, $x_3 = 0$) and C($x_1 = \frac{1}{1-2\mu}\frac{c}{b}$, $x_2 = \frac{1}{1-2\mu}\frac{c}{b}$, $x_3 = 0$) consists of equilibriums which are Lyapunov stable, but very small disturbance such as mutation will drive the system to deviate from these equilibrium and head to the asymptotically stable state CP. Points A ($x_1 = 0$, $x_2 = 0$, $x_3 = \frac{1}{1-2\mu}\frac{(1-\mu)\alpha+\mu c}{\alpha+\beta+b-c}$) is a saddle node with a one-dimensional unstable manifold along CP-DD line and a two-dimensional stable manifold which constitutes the separatrix surface dividing the simplex into two parts. The region over the surface is the basin of attraction of cooperative evolutionary state CP which is about 0.81 under this parameter setting.

Putting figure 4(b) and 4(c) together, we get figure 5 in which the simplex-4 is divided into three regions. The bottom region under the separatrix surface of 'GGBBBG' norm is in the attraction basin of non-cooperative evolutionary state DD in both norms. The middle region between the separatrix surfaces of 'GGBBBG' and 'GGBGBG' norm is in the attraction basin of non-cooperative evolutionary state DD in 'GGBGBG' norm and attraction basin of cooperative evolutionary state CP in 'GGBBBG' norm. From the same initial points S1, the trajectory in 'GGBGBG' norm converges to DD

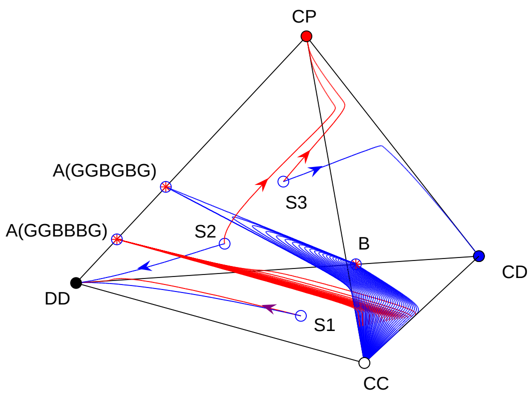


Figure 5: Typical trajectories under punishment-optional (blue lines and blue separatrix surface) and punishment-provoking (red lines) norm.

(yellow), and the trajectory in 'GGBGBG' norm converges to CP (red). The top region is in the attraction basin of cooperative evolutionary state CD in 'GGBGBG' norm and attraction basin of cooperative evolutionary state CP in 'GGBBBG' norm. This shows that the CESS in 'GGBBBG' norm, i.e. the state that all players taking 'CP' strategy, has the largest attraction basin ratio.

To get the overview of effect of parameters (α, β, b, c) to CESS attraction basin, we calculate the ratios of CESS attraction basin under different parameter settings by plenty of numerical computation and present the results in figure 6. The red lines are the ratios of CESS attraction basin in 'GGBBBG' norm, blue for 'GGBGBG' norm and green for 'GGBG' norm. In most normal parameter settings, the ratios of CESS attraction basin in punishment-provoking 'GGBBBG' norm is largest, followed by punishment-optional 'GGBGBG' norm and the non-punishment 'GGBG' norm smallest. Only in some abnormal situation such as that the punishment cost α is very large (right part of figure 6(a)) or the cooperation cost c is much small (left part of figure 6(c)), the ratio of CESS attraction basin in punishment-provoking 'GGBBBG' norm is not the largest.

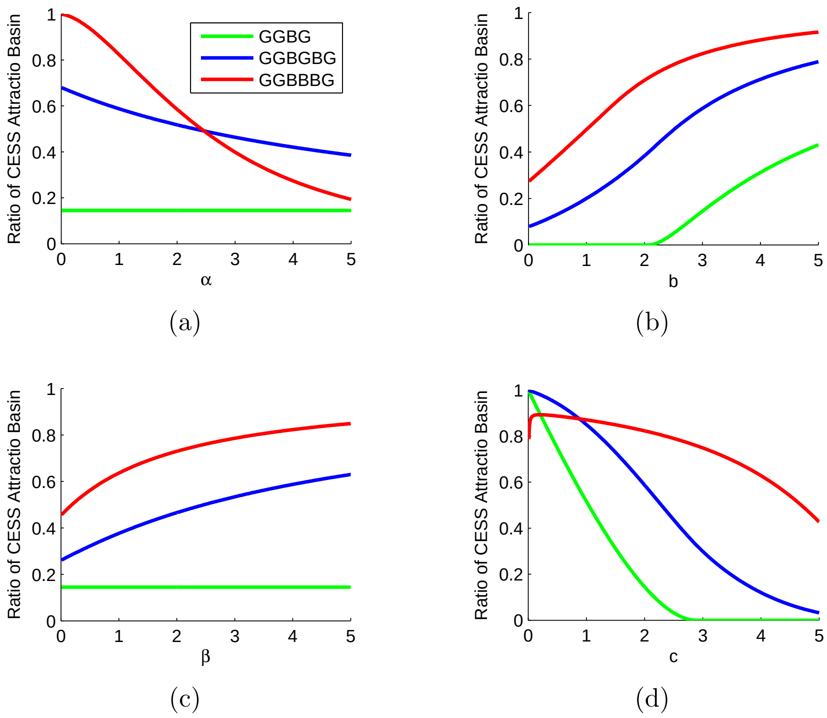


Figure 6: Ratios of CESS attraction basin of different social norms under different parameter settings ($b = 3$, $c = 2$, $\alpha = 1$, $\beta = 4$ if not specified, and $\mu = 0.02$).

This implies that when there are too many defectors in the society, and it is not in the attraction basin of cooperative evolutionary stable state if the society takes a non-punishment social norm, it may struggle out of the social dilemma and converge to the cooperative states by using a punishment-optional or punishment-provoking social norm.

3.3. Converge speed in three social norms

Even starting from such initial strategy frequencies that societies with any of the three social norms will approach to cooperative evolutionary stable states, they may converge to cooperative states with different speed in different social norms.

Figure 7(a) gives the comparison of converge speed to cooperative evolutionary stable states of societies with non-punishment (green) and punishment-optional (blue) social norm starting from the very close initial states, $x_1 = 0.02, x_2 = 0.72$ for 'GGBG' norm and $x_1 = 0.02, x_2 = 0.71, x_3 = 0.01$ for 'GGBGBG' norm. The horizontal axis is time, the vertical axis is the cooperation ratio, defined as the ratio of individuals who taking 'CC', 'CD' or 'CP' strategy. We can easily notice that a society with punishment-optional norm converges to the cooperative states more rapidly than a society with non-punishment norm. Consequently, the average payoff of a society with punishment-optional norm in each time increases more quickly than that of a society with non-punishment norm, except in a very short period of time in the beginning when a punishment-optional society suffers losses from the punishment larger than the gain from the increased cooperation, as in 7(b). After the society reaches the cooperative evolutionary stable state, i.e., all individuals taking CD strategy, the average payoff of a society with punishment-optional norm is exactly equal to that of a society with non-punishment norm. However, the average accumulative payoff from the starting time of a punishment-optional society will be significantly larger than that of a non-punishment society for very long time, as in 7(c).

Figure 7(d) gives the comparison of converge speed of societies with punishment-optional (blue) and punishment-provoking (red) social norm starting from the exactly same initial states, $x_1 = 0.05, x_2 = 0.15, x_3 = 0.3$. It is obvious that a society with punishment-provoking social norm converges to the cooperative states more rapidly than a society with punishment-optional norm. Consequently, the average payoff of a society with punishment-provoking norm increases more quickly than that of a society with punishment-optional norm, except in a very short period of time

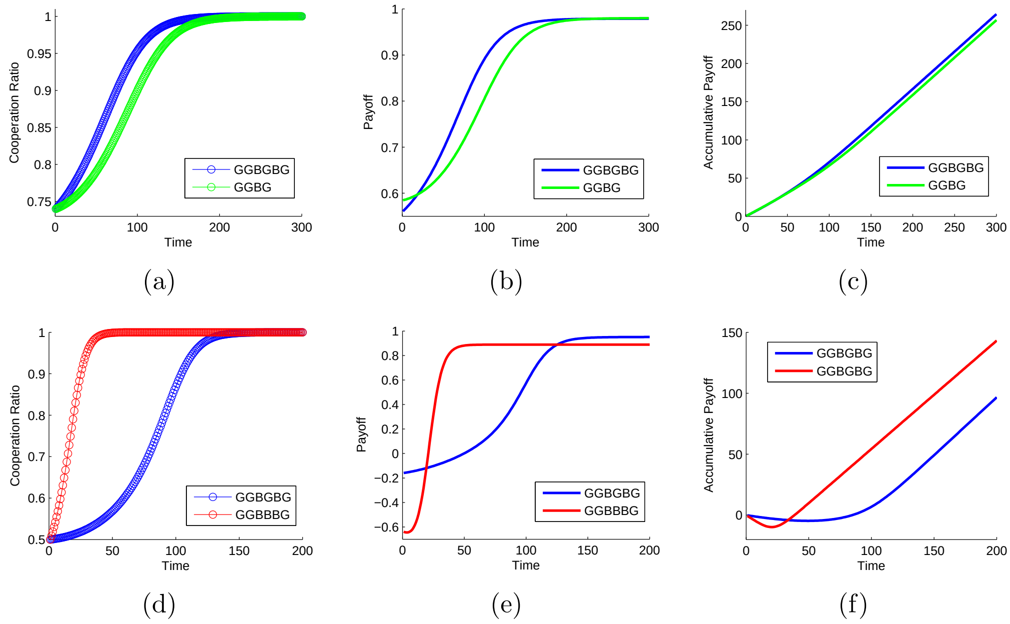


Figure 7: Comparison of converge speed in three social norms

in the beginning. After the society reaches the cooperative evolutionary stable state, i.e., all individuals taking CD (in 'GGBGBG' norm) or CP (in 'GGBBBG' norm) strategy, the average payoff of a society with punishment-provoking norm is slightly smaller than a society with punishment-optional norm for there are errors in reputation assignment and punishment is executed to 'bad' individuals with cost to both punisher and punished. However, the average accumulative payoff from the starting time of a punishment-provoking society will be significantly larger than that of a punishment-optional society for very long time as in 7(f).

This implies that even the initial state is in the attraction basin of cooperative evolutionary stable state for all social norms, if the society cares about the social overall benefit in a considerable long period of time, a punishment-optional or punishment-provoking social norm will be a better choice for it can increase the convergence speed to cooperative evolutionary stable states.

4. Conclusion and Discussion

Costly punishment is widespread in human society although both theoretical analysis and laboratory experiments show that punishment provides little efficiency and can hardly increase the total benefit of a population.

We give a possible explanation to this contradiction.We study the role of costly punishment in the evolution process to cooperation on the social norm level in a two-level evolution framework. Social norm is a global shared rule used to update the reputation of agents according to their actions. It is the collective choice of a population and evolves gradually according to the total benefit of all members in a society for a considerable long period of time. Agents are embedded in a certain social norm and they choose their strategies to maximize their individual benefit according to the current social environment including the strategies of other agents and the given social norm.

We explicitly model the strategies frequency dynamics in social norms with different punishment attitudes, and we find that the attraction basin ratio of cooperative evolutionary stable state (CESS) of non-punishment social norm is very small, that of punishment-optional social norm is larger and that of the punishment-provoking social norm is the largest. From the same initial state, society with punishment-provoking social norm can converge to cooperative evolutionary state more rapidly than society with punishment-optional social norm and society with punishment-optional social norm can converge to cooperative evolutionary state more rapidly than society with non-punishment social norm.

So costly punishment is inevitable for the evolution to cooperation in two situations. The first situation is that there are too many defectors in the society, i.e. it is not in the attraction basin of cooperative evolutionary stable state if the society takes a non-punishment social norm, so it can only struggle out of defection by using a punishment-optional or punishment-provoking social norm. The second is that even the initial state is in the attraction basin of cooperative evolutionary stable state for all social norms, if the society is not patient enough and wishes to reach the cooperative states more quickly, a punishment-optional or punishment-provoking social norm can increase the convergence speed.

We emphasize the role of punishment in the route to cooperation from initial state far away from the equilibrium. In the equilibrium when most individuals cooperate, the best choice for a society is the non-punishment social norm and the best choice for individuals under such norm is to withhold help rather than punish them as Ohtsuki et al. (2009) indicate. However the society is more likely to start from a state out of equilibrium for at least two possible reasons. The first possible reason is that human being evolves from the primitives most of whom are inherently self-regarding, and competition between societies favors those with punishment-provoking social norm. Another possible reason is that our society has the birth and death mechanism that old people die and new babies are born continuously. AS the babies are born self-interested, the society is always out of the stable state with monomorphic strategy. Because the punishment-provoking social norm has the larger attraction basin of cooperative evolutionary stable state (CESS) and can increase the converge speed to CESS, societies may evolve to take the punishment-provoking social norm for overall benefit of the society in the long history, and people will adapt to cooperate to good guys and punish bad ones. The punishment-provoking social norm can pass from generation to generation as the new born self-interested individuals will learn and acclimatize themselves to the social norm and take the appropriate strategies gradually as they growing up.

The punishment-provoking social norm predominate by the long history of co-evolution of social norm and individual strategy, and the individuals in such a social norm have evolved to acclimatize themselves to the social norm and tend to punish the social norm violator. Such an acclimatization can be embedded in individuals as a culture or even as a physiological reaction, such as the neural studies of de Quervain et al. (2004) and Singer et al. (2006) indicate that people receive pleasure from punishing norm violators. Once the subjects with such culture or physiological reaction participate experiments of Fehr and Gächter (2002), Dreber et al. (2008), Henrich et al. (2006) and etc., they will naturally punish the non-cooperators although it can not bring any material benefit to them, as the experiment results show. We argue that the evidence of so-called altruistic punishment can only indicate that the subjects are from a society with punishment-provoking or punishment-encouraging social norm but not the other-regarding reference of subjects. So we use the term 'costly punishment' instead of 'altruistic punishment', because the subjects are actually not altruistic, they think they will be rewarded in the future subconsciously or they will get a physiological pleasure.

There are much room for further research. We only investigate four out of nine strategies which are normal and reasonable in this paper. What will happen if individuals take the 'abnormal' strategies? Such as Herrmann et al. (2008) find that antisocial punishment, the sanctioning of people who behave cooperatively, exist across the societies. In current paper, individuals are paired randomly to interact. While in reality, people prefer to interact with those familiar and trusted. Is there more cooperation in a society if individuals have the opportunity select partners in the model? Only three typical social norms are analyzed in current paper, and the cooperation dynamics in more other norms need further exploration. We model the strategies frequency dynamics under fixed social norms and compare the fitness of different social norms in the evolution route of societies. Although this method can provide insight of the driving force of the social norm evolution, we can not capture the evolution process of social norms and many other interesting topics such as the effect of relative evolution speed of social norm and individual strategy. To address these issues, more powerful tool such as agent-based modeling (Tesfatsion, 2001) is needed. Agent-based simulation can not only provide individual level mechanism of the population level evolutionary dynamics, but also enable us to investigate more properties cannot be captured by the analytical analysis (Dawid, 2007).

Acknowledgements

We are grateful for useful comments and suggestions by Prof. Herbert Dawid from Bielefeld University, Prof. Beom Jun Kim from Sung Kyun Kwan University , Prof. Zengru Di from Beijing Normal Univerity and participants of Econophysics Colloquium 2010 Taipei. This work was supported by the National Science Foundation of China under Grant No.70871013 and the Program for New Century Excellent Talents in University under Grant No. NCET-07-0089.

Appendix A. Stable reputation distribution of given strategies frequency

In the main text, we only give a simple introduction of how to get the stable reputation distribution of given strategies frequency with the example of 'CC' strategy under non-punishment social norm. The detailed process to get the stable reputation frequency of all strategies for all three social norms is provided here.

For non-punishment (GGBG) social norm, there are three strategies 'CC', 'CD' and 'DD', with corresponding frequencies denoted by x_1, x_2 and x_3, and $x_1 + x_2 + x_3 = 1$. The ratios of players with good reputation in 'CC' , 'CD' and 'DD' players are denoted by g_1, g_2 and g_3 respectively. Thus the ratio of good players in entire population is $g = x_1g_1 + x_2g_2 + x_3g_3$. The reputation dynamics of three kinds players is illustrated by figure 2 in main text.

A 'CC' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action no matter what reputation the recipient has, and this tends to make him a good reputation. Due to the assignment error, he gets a good reputation with a probability $1 - \mu$ and bad reputation with probability μ. The 'CC' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as good with probability g_1. So the new ratio of good reputation in 'CC' players is $g_1^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_1$.

A 'CD' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action to good recipients and defection action to bad recipients, and both should bring him a good reputation with norm 'GGBG'. Due to the assignment error, he gets a good reputation with a probability $1 - \mu$ and bad reputation with probability μ. The 'CD' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as good with probability g_s. So the new ratio of good reputation in 'CD' players is $g_2^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_2$.

A 'DD' player has $\frac{1}{2}$ chance to be a donor, and takes defection action no matter what reputation the recipient has. He has a chance of $1 - g$ to meet a bad recipient. According to social norm 'GGBG', he should get a good reputation. Considering the reputation assignment error, he will get a good reputation with probability $(1-\mu)(1-g)$ and a bad reputation with probability $\mu(1-g)$. He also has a chance of g to meet a good recipient and get a bad reputation with probability $(1-\mu)g$ and a good reputation with probability μg. The 'DD' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as good with probability g_3. So the new ratio of good reputation in 'DD' players is $g_3^{'} = \frac{1}{2}(1-g)(1-\mu) + \frac{1}{2}g\mu + \frac{1}{2}g_3$.

$$\begin{cases} g_1^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_1 \\ g_2^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_2 \\ g_3^{'} = \frac{1}{2}(1-g)(1-\mu) + \frac{1}{2}g\mu + \frac{1}{2}g_3 \end{cases} \tag{A.1}$$

Since $g = x_1g_1 + x_2g_2 + x_3g_3$, we can solve the linear recursion and get the stable reputation distribution of each strategy $g_1^* = g_2^* = 1 - \mu$, $g_3^* = (1-\mu)[1 - \frac{1-2\mu}{1+(1-2\mu)x_3}]$. And the good reputation ratio in entire population is $g^* = \frac{1-\mu}{1+(1-2\mu)x_3}$.

For punishment-optional (GGBGBG) social norm, there are four strategies 'CC', 'CD', 'CP' and 'DD', with corresponding frequencies denoted by x_1, x_2, x_3 and x_4, and $x_1 + x_2 + x_3 + x_4 = 1$. The ratios of players with

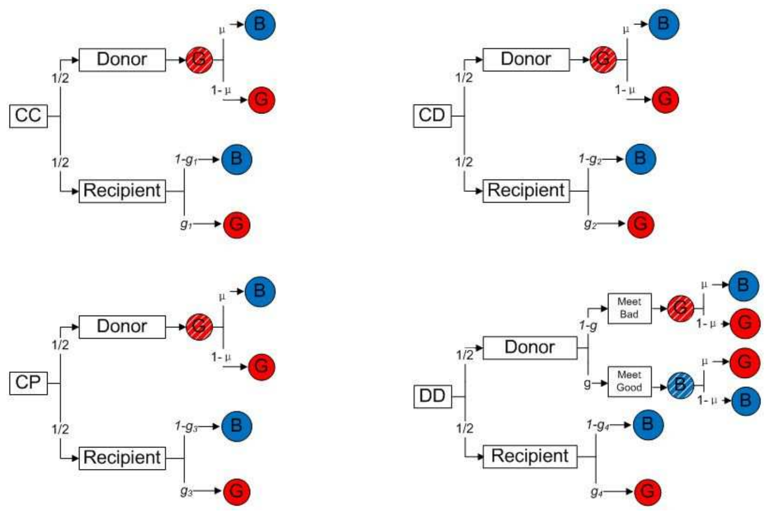


Figure A.1: Reputation dynamics of individuals taking different strategies in punishment-optional (GGBGBG) social norm.

good reputation in 'CC' , 'CD', 'CP' and 'DD' players are denoted by g_1, g_2, g_3 and g_4 respectively. Thus the ratio of good players in entire population is $g = x_1g_1 + x_2g_2 + x_3g_3 + x_4g_4$. The reputation dynamics of three kinds players is illustrated by figure A.1.

A 'CC' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action no matter what reputation the recipient has, and this tends to bring him a good reputation according to social norm 'GGBGBG'. Due to the assignment error, he gets a good reputation with a probability $1 - \mu$ and bad reputation with probability μ. The 'CC' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as good with probability g_1. So the new ratio of good reputation in 'CC' players is $g_1^{'} = \frac{1}{2}(1 - \mu) + \frac{1}{2}g_1$.

A 'CD' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action to good recipients and defection action to bad recipients, and both should bring him a good reputation with norm 'GGBGBG'. Due to the assignment error, he gets a good reputation with a probability $1 - \mu$ and bad reputation with probability μ. The 'CD' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as as good with probability g_2. So the new ratio of good reputation in 'CD' players is $g_2^{'} = \frac{1}{2}(1 - \mu) + \frac{1}{2}g_2$.

A 'CP' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action to good recipients and punishment action to bad recipients, and both should bring him a good reputation with norm 'GGBGBG'. Due to the assignment error, he gets a good reputation with a probability $1-\mu$ and bad reputation with probability μ. The 'CP' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as as good with probability g_3. So the new ratio of good reputation in 'CP' players is $g_3^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_3$.

A 'DD' player has $\frac{1}{2}$ chance to be a donor, and takes defection action no matter what reputation the recipient has. He has a chance of $1-g$ to meet a bad recipient. According to social norm 'GGBGBG', he should get a good reputation. Considering the reputation assignment error, he will get a good reputation with probability $(1-\mu)(1-g)$ and a bad reputation with probability $\mu(1-g)$. He also has a chance of g to meet a good recipient and get a bad reputation with probability $(1-\mu)g$ and a good reputation with probability μg. The 'DD' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as good with probability g_4. So the new ratio of good reputation in 'DD' players is $g_4^{'} = \frac{1}{2}(1-g)(1-\mu)+\frac{1}{2}g\mu+\frac{1}{2}g_4$.

$$\begin{cases} g_1^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_1 \\ g_2^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_2 \\ g_3^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_3 \\ g_4^{'} = \frac{1}{2}(1-g)(1-\mu) + \frac{1}{2}g\mu + \frac{1}{2}g_4 \end{cases} \tag{A.2}$$

Since $g = x_1g_1 + x_2g_2 + x_3g_3 + x_4g_4$, we can solve the linear recursion and get the stable reputation distribution. The ratios of good reputation in players taking 'CC','CD','CP' and 'DD' strategy are $g_1^* = g_2^* = g_3^* = 1-\mu$, $g_4^* = (1-\mu)[1 - \frac{1-2\mu}{1+(1-2\mu)x_4}]$. And the good reputation ratio in entire society is $g^* = \frac{1-\mu}{1+(1-2\mu)x_4}$.

For punishment-provoking (GGBBBG) social norm, there are four strategies 'CC', 'CD', 'CP' and 'DD', with corresponding frequencies denoted by x_1, x_2, x_3 and x_4, and $x_1 + x_2 + x_3 + x_4 = 1$. The ratios of players with good reputation in 'CC' , 'CD', 'CP' and 'DD' players are denoted by g_1, g_2, g_3 and g_4 respectively. Thus the ratio of good players in entire population is $g = x_1g_1 + x_2g_2 + x_3g_3 + x_4g_4$. The reputation dynamics of three kinds players is illustrated by figure A.2.

A 'CC' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action no matter what reputation the recipient has, and this tends to bring him a good reputation according to social norm 'GGBGBG'. Due to the assignment error, he gets a good reputation with a probability $1-\mu$ and bad reputation with probability μ. The 'CC' player also has $\frac{1}{2}$ chance to be a recipient; his

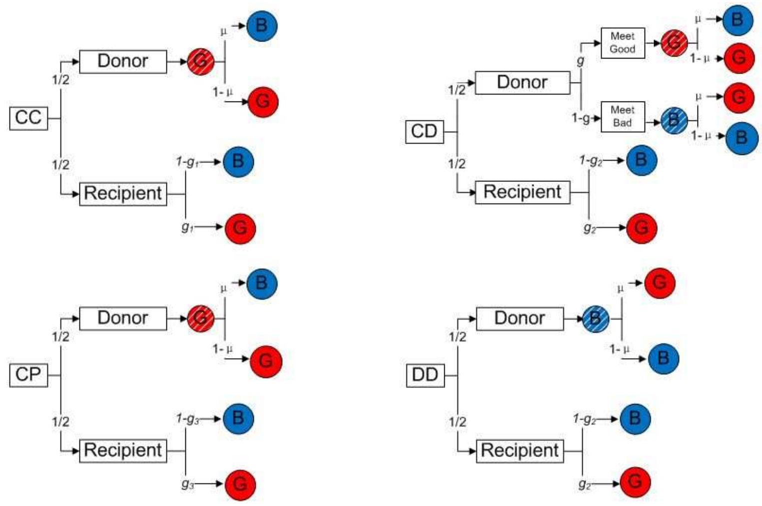


Figure A.2: Reputation dynamics of individuals taking different strategies in punishment-provoking (GGBBBG) social norm.

reputation does not change and remains as good with probability g_1. So the new ratio of good reputation in 'CC' players is $g_1^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_1$.

A 'CD' player has $\frac{1}{2}$ chance to be a donor and takes cooperation action to good recipients and defection action to bad recipients. With probability g he will meet a good recipient, his cooperation action to good recipient should bring him a good reputation. Due to the assignment error, he gets a good reputation with a probability $1-\mu$ and bad reputation with probability μ. With probability $1-g$ he will meet a bad recipient, his defection action to bad recipient should bring him a bad reputation according to 'GGBBBG' norm. Due to the assignment error, he gets a bad reputation with a probability $1-\mu$ and good reputation with probability μ. The 'CD' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as as good with probability g_2. So the new ratio of good reputation in 'CD' players is $g_2^{'} = \frac{1}{2}[(1-\mu)g + \mu(1-g)] + \frac{1}{2}g_2$.

A 'CP' player has $\frac{1}{2}$ chance to be a donor, and takes cooperation action to good recipients and punishment action to bad recipients, and both should bring him a good reputation with norm 'GGBGBG'. Due to the assignment error, he gets a good reputation with a probability $1-\mu$ and bad reputation with probability μ. The 'CP' player also has $\frac{1}{2}$ chance to be a recipient; his reputation does not change and remains as as good with probability g_3. So the new ratio of good reputation in 'CP' players is $g_3^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_3$.

A 'DD' player has $\frac{1}{2}$ chance to be a donor, and takes defection action no matter what reputation the recipient has. Both should bring him a bad reputation according to social norm 'GGBBBG'. Considering the reputation assignment error, he will get a bad reputation with probability $1-\mu$ and a good reputation with probability μ. So the new ratio of good reputation in 'DD' players is $g_4^{'} = \frac{1}{2}\mu + \frac{1}{2}g_4$.

$$\begin{cases} g_1^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_1 \\ g_2^{'} = \frac{1}{2}[(1-\mu)g + \mu(1-g)] + \frac{1}{2}g_2 \\ g_3^{'} = \frac{1}{2}(1-\mu) + \frac{1}{2}g_3 \\ g_4^{'} = \frac{1}{2}\mu + \frac{1}{2}g_4 \end{cases} \tag{A.3}$$

Since $g = x_1g_1 + x_2g_2 + x_3g_3 + x_4g_4$, we can solve the linear recursion and get the stable reputation distribution. The ratios of good reputation in 'CC' and 'CP' players are $g_1^* = g_3^* = 1-\mu$, and that in 'CD' players is $g_2^* = \mu + (1-2\mu)g^*$ and 'DD' players $g_4^* = \mu$. And the good reputation ratio in entire society is $g^* = \frac{(1-\mu)(x_1+x_3)+\mu(x_2+x_4)}{1-(1-2\mu)x_2}$.

Appendix B. Expected revenue of strategies in given strategies frequency

In the main text, we only give a simple introduction of how to calculate the expected revenue of strategies in given strategies frequency with the example of 'CC' strategy under non-punishment social norm. The detailed process to calculate the expected revenue of all strategies in given strategies frequency for all three social norms is provided here.

For non-punishment (GGBG) social norm, the calculation expected payoff of 'CC', 'CD' and 'DD' strategy is illustrated in figure 3 in main text.

For a 'CC' player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c regardless of the reputation of the recipient. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a 'CC' player to have a good reputation is $1-\mu$ in stable reputation distribution, he is expected to get revenue $(1-\mu)b$ when meeting a 'CD' donor. He will also meet a 'DD' donor with probability x_3 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CC' is $p_1 = \frac{1}{2}(-c) + \frac{1}{2}[bx_1 + bx_2(1-\mu)]$.

For a 'CD' player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c to good recipients with probability g and defects with no cost to bad recipients with probability $1-g$. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a 'CD' player to have a good reputation is $1-\mu$ in stable reputation distribution, he is expected to get revenue $(1-\mu)b$ when meeting a 'CD' donor. He will also meet a 'DD' donor with probability x_3 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CD' is $p_2 = \frac{1}{2}g(-c) + \frac{1}{2}[bx_1 + bx_2(1-\mu)]$.

For a 'DD' player, he has $\frac{1}{2}$ chance to be a donor and defects with no cost regardless of the reputation of the recipient. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the ratio of a 'DD' player to have a good reputation is g_3 in stable reputation distribution as defined above, he is expected to get revenue $g_3 b$ when meeting a 'CD' donor. He will also meet a 'DD' donor with probability x_3 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CC' is $p_3 = \frac{1}{2}(0) + \frac{1}{2}[bx_1 + bx_2 g_3)]$.

Summarily, the expected revenues of all three strategies in non-punishment (GGBG) social norm are

$$\begin{cases} p_1 = \frac{1}{2}(-c) + \frac{1}{2}[bx_1 + bx_2(1-\mu)] \\ p_2 = \frac{1}{2}g(-c) + \frac{1}{2}[bx_1 + bx_2(1-\mu)] \\ p_3 = \frac{1}{2}(0) + \frac{1}{2}[bx_1 + bx_2 g_3)] \end{cases} \tag{B.1}$$

.

For punishment-optional (GGBGBG) social norm, the calculation expected payoff of 'CC', 'CD', 'CP' and 'DD' strategy is illustrated in figure B.1.

For a 'CC' player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c regardless of the reputation of the recipient. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a 'CC' player to have a good reputation is $1-\mu$ in stable reputation distribution, he is expected to get revenue $(1-\mu)b$ when meeting a 'CD' donor. He will also meet a 'CP' donor with

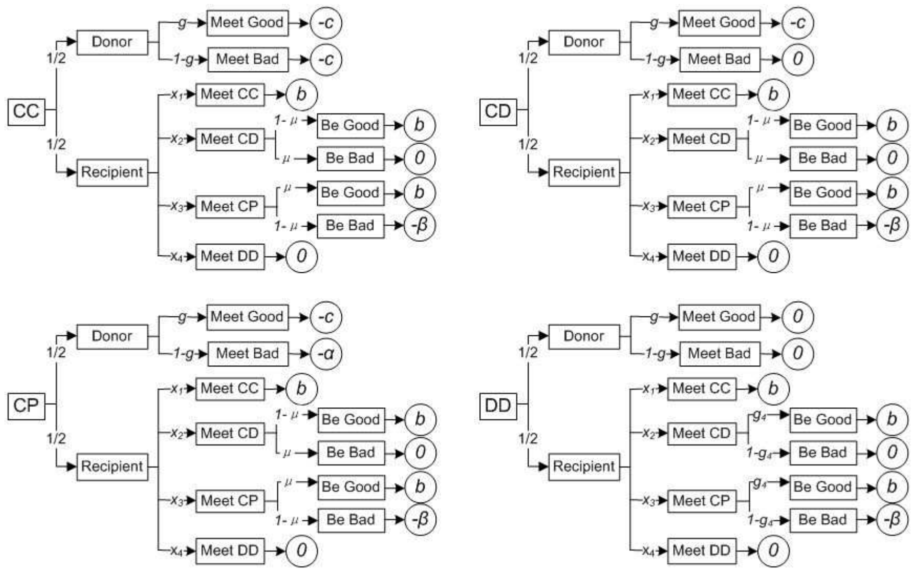


Figure B.1: Calculate the expected revenue of strategies in punishment-optional (GGBG-BG) social norm.

probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'CC' player to have a good (or bad) reputation is $1-\mu$ (or μ), he is expected to get revenue $(1-\mu)b-\mu\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CC' is $p_1 = \frac{1}{2}(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1-\mu)]$.

For a 'CD' player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c to good recipients with probability g and defects with no cost to bad recipients with probability $1-g$. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a 'CD' player to have a good reputation is $1-\mu$ in stable reputation distribution, he is expected to get revenue $(1-\mu)b$ when meeting a 'CD' donor. He will also meet a 'CP' donor with probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'CD' player to have a good (or bad) reputation is $1-\mu$ (or μ), he is expected to get revenue $(1-\mu)b-\mu\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CD' is $p_2 = \frac{1}{2}g(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1 - \mu)]$.

For a 'CP' player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c to good recipients with probability g and punishes with cost α bad recipients with probability $1 - g$. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a 'CP' player to have a good reputation is $1 - \mu$ in stable reputation distribution, he is expected to get revenue $(1-\mu)b$ when meeting a 'CD' donor. He will also meet a 'CP' donor with probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'CP' player to have a good (or bad) reputation is $1-\mu$ (or μ), he is expected to get revenue $(1-\mu)b - \mu\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CP' is $p_3 = \frac{1}{2}g(-c) + \frac{1}{2}(1 - g)(-\alpha) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1 - \mu)]$.

For a 'DD' player, he has $\frac{1}{2}$ chance to be a donor and defects with no cost regardless of the reputation of the recipient. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the ratio of a 'DD' player to have a good reputation is g_4 in stable reputation distribution as defined above, he is expected to get revenue g_4b when meeting a 'CD' donor. He will meet a 'CP' donor with probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'DD' player to have a good (or bad) reputation is g_4 (or $1 - g_4$), he is expected to get revenue $g_4b - (1 - g_4)\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'DD' is $p_4 = \frac{1}{2}[bx_1 + b(x_2 + x_3)g_4] + \frac{1}{2}x_3(1 - g_4)(-\beta)$.

Summarily, the expected revenues of all four strategies in punishment-optional (GGBGBG) social norm are

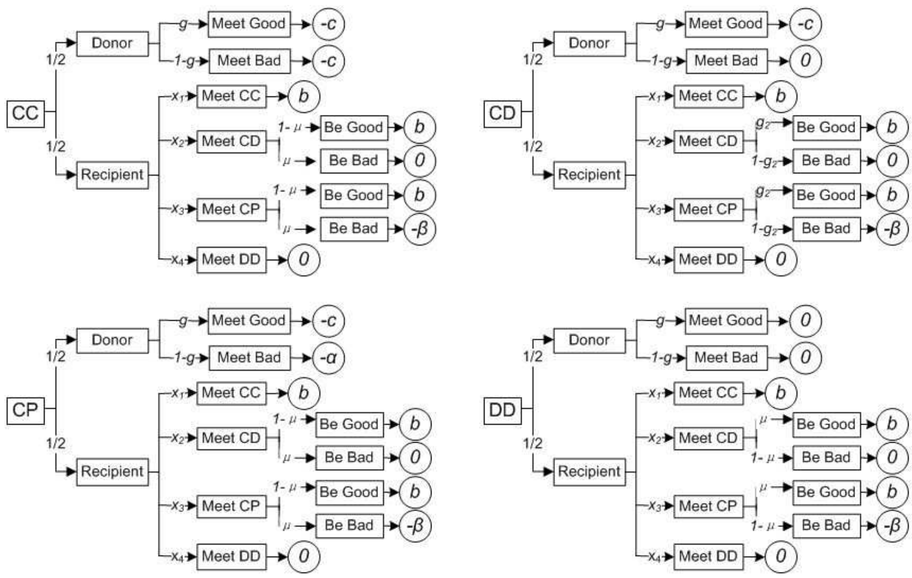


Figure B.2: Calculate the expected revenue of strategies in punishment-provoking (GG-BBBG) social norm.

$$
\begin{cases}
p_1 = \frac{1}{2}(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2+x_3)(1-\mu)] \\
p_2 = \frac{1}{2}g(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2+x_3)(1-\mu)] \\
p_3 = \frac{1}{2}g(-c) + \frac{1}{2}(1-g)(-\alpha) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2+x_3)(1-\mu)] \\
p_4 = \frac{1}{2}[bx_1 + b(x_2+x_3)g_4] + \frac{1}{2}x_3(1-g_4)(-\beta)
\end{cases}
\tag{B.2}
$$

.

For punishment-provoking (GGBBBG) social norm, the calculation expected payoff of ‘CC’, ‘CD’, ‘CP’ and ‘DD’ strategy is illustrated in figure B.2.

For a ‘CC’ player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c regardless of the reputation of the recipient. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a ‘CC’ player with probability x_1 and be expected to get b revenue. He will meet a ‘CD’ player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a ‘CC’ player to have a good reputation is $1-\mu$ in stable reputation distribution, he is expected to get revenue $(1-\mu)b$ when meeting a ‘CD’ donor. He will meet a ‘CP’ donor with probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'CC' player to have a good (or bad) reputation is $1-\mu$ (or μ), he is expected to get revenue $(1-\mu)b-\mu\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CC' is $p_1 = \frac{1}{2}(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2 + x_3)(1-\mu)]$.

For a 'CD' player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c to good recipients with probability g and defects with no cost to bad recipients with probability $1-g$. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a 'CD' player to have a good reputation is g_2 in stable reputation distribution, he is expected to get revenue $g_2 b$ when meeting a 'CD' donor. He will also meet a 'CP' donor with probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'CD' player to have a good (or bad) reputation is g_2 (or $1-g_2$), he is expected to get revenue $g_2 b - (1-g_2)\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CD' is $p_2 = \frac{1}{2}g(-c) + \frac{1}{2}x_3(1-g_2)(-\beta) + \frac{1}{2}[bx_1 + bg_2(x_2+x_3)]$.

For a 'CP' player, he has $\frac{1}{2}$ chance to be a donor and cooperates with a cost c to good recipients with probability g and punishes with cost α bad recipients with probability $1-g$. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the probability of a 'CP' player to have a good reputation is $1-\mu$ in stable reputation distribution, he is expected to get revenue $(1-\mu)b$ when meeting a 'CD' donor. He will also meet a 'CP' donor with probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'CP' player to have a good (or bad) reputation is $1-\mu$ (or μ), he is expected to get revenue $(1-\mu)b-\mu\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'CP' is $p_3 = \frac{1}{2}g(-c) + \frac{1}{2}(1-g)(-\alpha) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2+x_3)(1-\mu)]$.

For a 'DD' player, he has $\frac{1}{2}$ chance to be a donor and defects with no cost regardless of the reputation of the recipient. With another $\frac{1}{2}$ chance, he will be a recipient. As a recipient, he will meet a 'CC' player with probability x_1 and be expected to get b revenue. He will meet a 'CD' player with probability x_2 and be expected to get b if his reputation is good and 0 if his reputation is bad. Because the ratio of a 'DD' player to have a good reputation is μ in stable reputation distribution as defined above, he is expected to get revenue μb when meeting a 'CD' donor. He will meet a 'CP' donor with probability x_3 and be expected to get b if his reputation is good and $-\beta$ if his reputation is bad. Because the probability of a 'DD' player to have a good (or bad) reputation is μ (or $1-\mu$), he is expected to get revenue $\mu b - (1-\mu)\beta$ when meeting a 'CP' donor. He will also meet a 'DD' donor with probability x_4 and be expected to get nothing. Totally, the expected revenue of a player with strategy 'DD' is $p_4 = \frac{1}{2}[bx_1 + b(x_2+x_3)\mu] + \frac{1}{2}x_3(1-\mu)(-\beta)$.

Summarily, the expected revenues of all four strategies in punishment-optional (GGBGBG) social norm are

$$\begin{cases} p_1 = \frac{1}{2}(-c) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2+x_3)(1-\mu)] \\ p_2 = \frac{1}{2}g(-c) + \frac{1}{2}x_3(1-g_2)(-\beta) + \frac{1}{2}[bx_1 + bg_2(x_2+x_3)] \\ p_3 = \frac{1}{2}g(-c) + \frac{1}{2}(1-g)(-\alpha) + \frac{1}{2}x_3\mu(-\beta) + \frac{1}{2}[bx_1 + b(x_2+x_3)(1-\mu)] \\ p_4 = \frac{1}{2}[bx_1 + b(x_2+x_3)\mu] + \frac{1}{2}x_3(1-\mu)(-\beta) \end{cases} \tag{B.3}$$

.

Appendix C. Detailed expression of strategies frequency dynamics for punishment-optional and punishment-provoking norms

The expressions of strategies frequency dynamics for punishment-optional and punishment-provoking norms are too complicated to be included in the main text. And we only provide the finished form of dynamical system of strategies frequency dynamics for non-punishment social norm as an demonstration. Here the detailed expression for punishment-optional and punishment-provoking norms are articulated.

The dynamical system of strategies frequency in punishment-optional (GGBGBG) social norm is

$$
\left\{
\begin{aligned}
\dot{x}_1 = &- cx_1 + cx_1^2 + \alpha x_1 x_3 \\
&+ \left[\begin{array}{l}
[(1-2\mu)b + c]x_1x_2 \\
+[(1-2\mu)(b+\beta) + (c-\alpha)]x_1x_3 \\
-(1-2\mu)b[x_1+x_2]x_1x_2 \\
+(1-2\mu)(b+\beta)[x_1+x_3]x_1x_3 \\
+(1-2\mu)(2b+\beta)x_1x_2x_3
\end{array}\right] \Bigg/ [2 - \frac{1-2\mu}{1-\mu}\sum_{i=1}^{3} x_i] \\
\dot{x}_2 = &- cx_1x_2 + \alpha x_2 x_3 \\
&+ \left[\begin{array}{l}
-cx_2 + [(1-2\mu)b + c]x_2^2 \\
+[(1-2\mu)(b+\beta) + (c-\alpha)]x_2x_3 \\
-(1-2\mu)b[x_1+x_2]x_2^2 \\
-(1-2\mu)(b+\beta)[x_1+x_3]x_2x_3 \\
-(1-2\mu)(2b+\beta)x_2^2x_3
\end{array}\right] \Bigg/ [2 - \frac{1-2\mu}{1-\mu}\sum_{i=1}^{3} x_i] \\
\dot{x}_3 = &- \alpha x_3 + cx_1x_3 + \alpha x_3^2 \\
&+ \left[\begin{array}{l}
-(c-\alpha)x_3 + [(1-2\mu)b + c]x_2x_3 \\
+[(1-2\mu)(b+\beta) + (c-\alpha)]x_3^2 \\
-(1-2\mu)b[x_1x_2 + x_2^2 + x_2x_3]x_3 \\
-(1-2\mu)(b+\beta)[x_1x_3 + x_3^2]x_3 \\
-(1-2\mu)(2b+\beta)x_2x_3^2
\end{array}\right] \Bigg/ [2 - \frac{1-2\mu}{1-\mu}\sum_{i=1}^{3} x_i]
\end{aligned}
\right.
\tag{C.1}
$$

.

The dynamical system of strategies frequency in punishment-provoking (GGBBBG) social norm is

$$
\left\{
\begin{aligned}
\dot{x}_1 =& -cx_1 + cx_1^2 + [(1-2\mu)(b+\beta)+\alpha]x_1x_3 \\
&+ (1-2\mu)b[1-x_1-x_3]x_1x_2 - (1-2\mu)(b+\beta)[x_1+x_3]x_1x_3 \\
&- \left[\begin{array}{l}
-\frac{\mu c}{1-2\mu}x_1x_2 + \frac{\mu(\alpha-c)}{1-2\mu}x_1x_3 - cx_1^2x_2 \\
+\mu b x_1x_2^2 + [\mu(b+\beta)-c]x_1x_2x_3 \\
+(\alpha-c)[x_1+x_3]x_1x_3 \\
+(1-2\mu)b[x_1+x_3]x_1x_2^2 \\
+(1-2\mu)(b+\beta)[x_1+x_3]x_1x_2x_3^2
\end{array}\right] \Big/ [\frac{1}{1-2\mu} - x_2] \\
\dot{x}_2 =& cx_1^2 + \alpha x_2x_3 - (1-2\mu)bx_1x_2^2 \\
&- (1-2\mu)(b+\beta)[x_1+x_3]x_2x_3 - (1-2\mu)bx_2^2x_3 \\
&+ \left[\begin{array}{l}
-\frac{\mu c}{1-2\mu}x_2 - cx_1x_2 + \mu[b+\frac{c}{1-2\mu}]x_2^2 \\
+[(\mu b+\mu\beta-c)+\frac{\mu(c-\alpha)}{1-2\mu}]x_2x_3 \\
+[(1-2\mu)b+c]x_1x_2^2 - \mu b x_2^3 \\
+[(1-2\mu)(b+\beta)+(c-\alpha)]x_1x_2x_3 \\
+[(1-2\mu)b-\mu(b+\beta)+c]x_2^2x_3 \\
+[(1-2\mu)b+(c-\alpha)]x_2x_3^2 \\
-(1-2\mu)b(x_1+x_3)x_2^3 \\
-(1-2\mu)(b+\beta)(x_1+x_3)x_2^2x_3
\end{array}\right] \Big/ [\frac{1}{1-2\mu} - x_2] \\
\dot{x}_3 =& -\alpha x_3 + cx_1x_3 + [(1-2\mu)(b+\beta)+\alpha)]x_3^2 \\
&+ (1-2\mu)b[1-x_1-x_3]x_2x_3 - (1-2\mu)(b+\beta)[x_1+x_3]x_3^2 \\
&- \left[\begin{array}{l}
\frac{\mu(c-\alpha)}{1-2\mu}x_3 - \frac{\mu c}{1-2\mu}x_2x_3 + \mu b x_2^2x_3 \\
-(c-\alpha)x_1x_3^2 + [\mu(b+\beta)-c]x_2x_3^2 \\
-(c-\alpha)x_3^3 + (c-\alpha)(x_1+x_3)x_3 \\
+(1-2\mu)b(x_1+x_3)x_2^2x_3 - cx_1x_2x_3 \\
+(1-2\mu)(b+\beta)(x_1+x_3)x_2x_3^2
\end{array}\right] \Big/ [\frac{1}{1-2\mu} - x_2]
\end{aligned}
\right.
\tag{C.2}
$$

.

Appendix D. Proofs to the propositions in equilibrium analysis

Proof 1 (Proof of Proposition 1). *For non-punishment (GGBG) norm, inserting $x_1 = 0$, $x_2 = 0$ into dynamical system 4, we can get dotx$_1$ = dotx$_2$ = 0, so $x_1 = 0$, $x_2 = 0$ is a fixed point. To prove the stability of $x_1 = 0$, $x_2 = 0$, we calculate the Jacobian matrix (J) of dynamical system 4, and evaluate it at some equilibrium $(0, 0)$. We get,*

$$J\Big|_{\substack{x_1=0\\x_2=0}} = \begin{bmatrix} -c & 0 \\ 0 & c/2 \end{bmatrix} \tag{D.1}$$

It follows that the eigenvalues at the equilibrium $(0,0)$ *are* $\lambda_1 = -c$, $\lambda_2 = -c/2$. *We have assumed that* $c > 0$, *so* $\lambda_1 < 0$, $\lambda_2 < 0$ *and equilibrium* $(0,0)$ *is always stable.*

For punishment-optional (GGBGBG) norm, inserting $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ *into dynamical system C.1, we can get* $dotx_1 = dotx_2 = dotx_3 = 0$, *so* $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ *is a fixed point. To prove the stability, we calculate the Jacobian matrix (J) of dynamical system C.1, and evaluate it at some equilibrium* $(0,0,0)$. *We get,*

$$J\Big|_{\substack{x_1=0\\x_2=0\\x_3=0}} = \begin{bmatrix} -c & 0 & 0 \\ 0 & -c/2 & 0 \\ 0 & 0 & -(c+\alpha)/2 \end{bmatrix} \tag{D.2}$$

Then the eigenvalues at the equilibrium $(0,0,0)$ *are* $\lambda_1 = -c$, $\lambda_2 = -c/2$, $\lambda_3 = -(c+\alpha)/2$. *We have assumed that* $c > 0$, $\alpha > 0$, *so* $\lambda_1 < 0$, $\lambda_2 < 0$, $\lambda_{33} < 0$ *and equilibrium* $(0,0,0)$ *is always stable.*

For punishment-provoking (GGBBBG) norm, inserting $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ *into dynamical system C.2, we can get* $dotx_1 = dotx_2 = dotx_3 = 0$, *so* $x_1 = 0$, $x_2 = 0$, $x_3 = 0$ *is a fixed point. To prove the stability, we calculate the Jacobian matrix (J) of dynamical system C.2, and evaluate it at some equilibrium* $(0,0,0)$. *We get,*

$$J\Big|_{\substack{x_1=0\\x_2=0\\x_3=0}} = \begin{bmatrix} -c & 0 & 0 \\ 0 & -\mu c & 0 \\ 0 & 0 & -\mu c-(1-\mu)\alpha) \end{bmatrix} \tag{D.3}$$

Then the eigenvalues at the equilibrium $(0,0,0)$ *are* $\lambda_1 = -c$, $\lambda_2 = -\mu c$, $\lambda_3 = -\mu c - (1-\mu)\alpha)$. *We have assumed that* $c > 0$, $\mu > 0$, *so* $\lambda_1 < 0$, $\lambda_2 < 0$, $\lambda_{33} < 0$ *and equilibrium* $(0,0,0)$ *is always stable.* □

Proof 2 (Proof of Proposition 2). *Inserting* $x_1 = 0$, $x_2 = 1$ *into dynamical system 4, we can get* $dotx_1 = dotx_2 = 0$, *so* $x_1 = 0$, $x_2 = 1$ *is a fixed point. To prove the stability of* $x_1 = 0$, $x_2 = 1$, *we calculate the Jacobian matrix (J) of dynamical system 4, and evaluate it at some equilibrium* $(0,1)$. *We get,*

$$J\Big|_{\substack{x_1=0\\x_2=1}} = \begin{bmatrix} -\mu c & 0 \\ c-(1-\mu)(1-2\mu)b & (1-\mu)[c-(1-2\mu)b] \end{bmatrix} \tag{D.4}$$

It follows that the eigenvalues at the equilibrium $(0,1)$ *are* $\lambda_1 = -\mu c$, $\lambda_2 = (1-\mu)[c-(1-2\mu)b]$. *We have assumed that* $c > 0, \mu > 0$, *so* $\lambda_1 < 0$. *The stability condition for equilibrium* $(0,1)$ *is* $\lambda_2 < 0$, *that is,* $\frac{1}{1-2\mu}\frac{c}{b} < 1$. □

Proof 3 (Proof of Proposition 3). *Inserting* $x_1 = 0$, $x_2 = 1$, $x_3 = 0$ *into dynamical system C.1, we can get* $dotx_1 = dotx_2 = dotx_3 = 0$, *so* $x_1 = 0$, $x_2 = 1$, $x_3 = 0$ *is a fixed point. To prove the stability, we calculate the Jacobian matrix (J) of dynamical system C.1, and evaluate it at some equilibrium* $(0,1,0)$. *We get,*

$$J\Big|_{\substack{x_1=0\\x_2=1\\x_3=0}} = \begin{bmatrix} -\mu c & 0 & 0 \\ c-(1-\mu)(1-2\mu)b & (1-\mu)[c-(1-2\mu)b] & \mu\alpha+(1-\mu)[c-(1-2\mu)b] \\ 0 & 0 & -\mu\alpha \end{bmatrix} \tag{D.5}$$

Then the eigenvalues at the equilibrium $(0,1,0)$ *are* $\lambda_1 = -\mu c$, $\lambda_2 = (1-\mu)[c-(1-2\mu)b]$, $\lambda_3 = -\mu\alpha$. *We have assumed that* $c > 0$, $b > 0$, $\alpha > 0$, $\mu > 0$, *so* $\lambda_1 < 0$, $\lambda_3 < 0$. *The stability condition for equilibrium* $(0,1,0)$ *is* $\lambda_2 < 0$, *that is,* $\frac{1}{1-2\mu}\frac{c}{b} < 1$. □

Proof 4 (Proof of Proposition 4). *For punishment-provoking (GGBBBG) norm, inserting* $x_1 = 0$, $x_2 = 0$, $x_3 = 1$ *into dynamical system C.2, we can get* $dotx_1 = dotx_2 = dotx_3 = 0$, *so* $x_1 = 0$, $x_2 = 0$, $x_3 = 1$ *is a fixed point. To prove the stability, we calculate the Jacobian matrix (J) of dynamical system C.2, and evaluate it at some equilibrium* $(0,0,1)$. *We get,*

$$J\Big|_{\substack{x_1=0\\x_2=0\\x_3=1}} = \begin{bmatrix} -\mu(c-\alpha) & 0 & 0 \\ 0 & \mu\alpha-\mu(1-2\mu)(b+\beta) & 0 \\ c-\mu(1-2\mu)(b+\beta) & (1-\mu)[c-\mu(1-2\mu)(b+\beta)] & (1-\mu)c+\mu\alpha-(1-2\mu)(b+\beta)) \end{bmatrix} \tag{D.6}$$

Then the eigenvalues at the equilibrium $(0,0,1)$ *are* $\lambda_1 = -\mu(c-\alpha)$, $\lambda_2 = \mu\alpha - \mu(1-2\mu)(b+\beta)$, $\lambda_3 = (1-\mu)c + \mu\alpha - (1-2\mu)(b+\beta))$. *To guarantee stability of equilibrium* $(0,0,1)$, *it is required that* $\lambda_1 < 0$, $\lambda_2 < 0$, $\lambda_3 < 0$. *And* $\lambda_1 < 0$ *leads to condition* $c < \alpha$. *We can validate that* $\lambda_2 < 0$ *if*$c < \alpha$. *And* $\lambda_3 < 0$ *leads to condition* $c < \alpha$. *Therefore, stability condition for equilibrium* $(0,0,1)$ *is* $c < \alpha$ *and* $\frac{1}{1-2\mu}\frac{(1-\mu)\alpha+\mu c}{\alpha+\beta+b-c} < 1$. □

References

Andreoni, J., Harbaugh, W., Vesterlund, L., 2003. The carrot or the stick: Rewards, punishments, and cooperation. The American Economic Review 93 (3), 893–902.

Axelrod, R., 1984. The Evolution of Cooperation. Basic Books, New York.

Bergstrom, T. C., 2002. Evolution of social behavior: Individual and group selection. The Journal of Economic Perspectives 16 (2), 67–88.

Bochet, O., Page, T., Putterman, L., 2006. Communication and punishment in voluntary contribution experiments. Journal of Economic Behavior and Organization 60 (1), 11–26.

Boehm, C., 1993. Egalitarian behavior and reverse dominance hierarchy. Current Anthropology 34 (3), 227.

Bowles, S., Gintis, H., 2004. The evolution of strong reciprocity: cooperation in heterogeneous populations. Theoretical Population Biology 65 (1), 17–28.

Boyd, R., Gintis, H., Bowles, S., Richerson, P. J., 2003. The evolution of altruistic punishment. Proceedings of the National Academy of Sciences of the United States of America 100 (6), 3531–3535.

Dawid, H., 2007. Evolutionary game dynamics and the analysis of agent-based imitation models: The long run, the medium run and the importance of global analysis. Journal of Economic Dynamics and Control 31 (6), 2108–2133.

de Quervain, D. J.-F., Fischbacher, U., Treyer, V., Schellhammer, M., Schnyder, U., Buck, A., Fehr, E., 2004. The neural basis of altruistic punishment. Science 305 (5688), 1254–1258.

Dreber, A., Rand, D. G., Fudenberg, D., Nowak, M. A., 2008. Winners don/'t punish. Nature 452 (7185), 348–351.

Egas, M., Riedl, A., 2008. The economics of altruistic punishment and the maintenance of cooperation. Proceedings of the Royal Society B: Biological Sciences 275 (1637), 871–878.

Falk, A., Fehr, E., Fischbacher, U., 2005. Driving forces behind informal sanctions. Econometrica 73 (6), 2017–2030.

Fehr, E., Fischbacher, U., 2003. The nature of human altruism. Nature 425 (6960), 785–791.

Fehr, E., Fischbacher, U., 2004. Social norms and human cooperation. Trends in Cognitive Sciences 8 (4), 185–190.

Fehr, E., Fischbacher, U., Gächter, S., 2002. Strong reciprocity, human cooperation, and the enforcement of social norms. Human Nature 13 (1), 1–25.

Fehr, E., Gächter, S., 2002. Altruistic punishment in humans. Nature 415 (6868), 137–140.

Fehr, E., Simon, G., 2000. Cooperation and punishment in public goods experiments. The American Economic Review 90 (4), 980–994.

Fowler, J. H., 2005. Altruistic punishment and the origin of cooperation. Proceedings of the National Academy of Sciences of the United States of America 102 (19), 7047–7049.

Gächter, S., Herrmann, B., 2009. Reciprocity, culture and human cooperation: previous insights and a new cross-cultural experiment. Philosophical Transactions of the Royal Society B: Biological Sciences 364 (1518), 791–806.

Gintis, H., 2000. Strong reciprocity and human sociality. Journal of Theoretical Biology 206 (2), 169–179.

Gürerk, O., Irlenbusch, B., Rockenbach, B., 2006. The competitive advantage of sanctioning institutions. Science 312 (5770), 108–111.

Henrich, J., 2004. Cultural group selection, coevolutionary processes and large-scale cooperation. Journal of Economic Behavior and Organization 53 (1), 3–35.

Henrich, J., 2006. Cooperation, punishment, and the evolution of human institutions. Science 312 (5770), 60–61.

Henrich, J., Boyd, R., 2001. Why people punish defectors: Weak conformist transmission can stabilize costly enforcement of norms in cooperative dilemmas. Journal of Theoretical Biology 208 (1), 79–89, doi: DOI: 10.1006/jtbi.2000.2202.

Henrich, J., Boyd, R., Bowles, S., Camerer, C., Fehr, E., Gintis, H., McElreath, R., Alvard, M., Barr, A., Ensminger, J., Henrich, N. S., Hill, K., Gil-White, F., Gurven, M., Marlowe, F. W., Patton, J. Q., Tracer, D., 2005. "Economic man" in cross-cultural perspective: Behavioral experiments in 15 small-scale societies. Behavioral and Brain Sciences 28 (06), 795–815.

Henrich, J., Ensminger, J., McElreath, R., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., Ziker, J., 2010. Markets, religion, community size, and the evolution of fairness and punishment. Science 327 (5972), 1480–1484.

Henrich, J., McElreath, R., Barr, A., Ensminger, J., Barrett, C., Bolyanatz, A., Cardenas, J. C., Gurven, M., Gwako, E., Henrich, N., Lesorogol, C., Marlowe, F., Tracer, D., Ziker, J., 2006. Costly punishment across human societies. Science 312 (5781), 1767–1770.

Herrmann, B., Thoeni, C., Gachter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.

Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge.

Knauft, B. M., Abler, T. S., Betzig, L., Boehm, C., Dentan, R. K., Kiefer, T. M., Otterbein, K. F., Paddock, J., Rodseth, L., 1991. Violence and sociality in human evolution [and comments and replies]. Current Anthropology 32 (4), 391–428.

Marlowe, F. W., Berbesque, J. C., Barr, A., Barrett, C., Bolyanatz, A., Cardenas, J. C., Ensminger, J., Gurven, M., Gwako, E., Henrich, J., Henrich, N., Lesorogol, C., McElreath, R., Tracer, D., 2008. More "ltruistic" punishment in larger societies. Proceedings of the Royal Society B: Biological Sciences 275 (1634), 587–592.

Milinski, M., Rockenbach, B., 2008. Punisher pays. Nature 452 (7185), 297–298.

Nowak, M. A., 2006. Five rules for the evolution of cooperation. Science 314 (5805), 1560–1563.

Nowak, M. A., Sigmund, K., 2005. Evolution of indirect reciprocity. Nature 437 (7063), 1291–1298.

Ohtsuki, H., Iwasa, Y., 2007. Global analyses of evolutionary dynamics and exhaustive search for social norms that maintain cooperation by reputation. Journal of Theoretical Biology 244 (3), 518–531.

Ohtsuki, H., Iwasa, Y., Nowak, M. A., 2009. Indirect reciprocity provides only a narrow margin of efficiency for costly punishment. Nature 457 (7225), 79–82.

Olson, M., 1965. The Logic of Collective Action: Public Goods and the Theory of Groups. Harvard University Press, Cambridge, MA.

Oosterbeek, H., Sloof, R., van de Kuilen, G., 2004. Cultural differences in ultimatum game experiments: Evidence from a meta-analysis. Experimental Economics 7 (2), 171–188.

Ostrom, E., 2000. Collective action and the evolution of social norms. The Journal of Economic Perspectives 14 (3), 137–158.

Ostrom, E., Walker, J., Gardner, R., 1992. Covenants with and without a sword: Self-governance is possible. The American Political Science Review 86 (2), 404–417.

Rockenbach, B., Milinski, M., 2006. The efficient interaction of indirect reciprocity and costly punishment. Nature 444 (7120), 718–723.

Singer, T., Seymour, B., O'Doherty, J. P., Stephan, K. E., Dolan, R. J., Frith, C. D., 2006. Empathic neural responses are modulated by the perceived fairness of others. Nature 439 (7075), 466–469.

Taylor, C., Nowak, M. A., 2007. Transforming the dilemma. Evolution 61 (10), 2281–2292.

Tesfatsion, L., 2001. Introduction to the special issue on agent-based computational economics. Journal of Economic Dynamics and Control 25 (3-4), 281–293.

Wu, J.-J., Zhang, B.-Y., Zhou, Z.-X., He, Q.-Q., Zheng, X.-D., Cressman, R., Tao, Y., 2009. Costly punishment does not always increase cooperation. Proceedings of the National Academy of Sciences 106 (41), 17448–17451.

Young, H. P., 2008. Social Norms. Palgrave Macmillan, Basingstoke.