Eur. Phys. J. B (2021) 94:176
https://doi.org/10.1140/epjb/s10051-021-00185-w

THE EUROPEAN
PHYSICAL JOURNAL B

Regular Article - Statistical and Nonlinear Physics

Evolutionary dynamics of trust in the N-player trust game with individual reward and punishment

Xing Fang and Xiaojie Chen[a]

School of Mathematical Sciences, University of Electronic Science and Technology of China, Chengdu 611731, China

Received 16 July 2021 / Accepted 13 August 2021 / Published online 31 August 2021

Abstract. Trust plays an important role in human society. However, how does trust evolve is a huge challenge. The trust game is a well-known paradigm to measure the evolution of trust in a population. Reward and punishment as the common types of incentives can be used to improve the trustworthiness. However, it remains unclear how reward and punishment actually influence the evolutionary dynamics of trust. Here, we introduce individual reward and punishment into the N-player trust game model in an infinite well-mixed population, where investors use a part of the returned fund to reward trustworthy trustees and meanwhile punish untrustworthy trustees. We then investigate the evolutionary dynamics of trust by means of replicator equations. We show that the introduction of reward and punishment can lead to the stable coexistence state of investors and trustworthy trustees, which indicates that the evolution of trust can be greatly promoted. We reveal that the attraction domain of the coexistence state becomes larger as investors increase the incentive strength from the returned fund for reward and punishment. In addition, we find that the increase of the reward coefficient can enlarge the attraction domain of the coexistence state, which implies that reward can better promote the evolution of trust than punishment.

1 Introduction

Trust plays a significant role in the development of human society [1–7]. With the help of trust, rational individuals with limited cognitive capacities can manage the complex situations [8]. Hence trust can work as a complexity-reduction mechanism for solving complicated social problems. For example, when people trust each other, transaction costs can be reduced [9,10].

However, how to understand the evolution of trust in a population is a huge puzzle. The trust game (TG) is an effective paradigm to measure the evolution of trust in quantitative manners [11]. Recently, Abbass proposed the N-player Trust Game (NTG) in an infinite well-mixed population and revealed that when the population consists of even the slightest number of untrustworthy individuals initially, the society converges to the state where there are zero trusters and many untrustworthy individuals [12]. To promote the level of trust in the TG, many mechanisms have been considered [13–16]. For example, Chica et al. considered the networked structure of populations allowing interactions and imitation with only neighboring players and found that trust can be promoted when individuals play the game on a social network [14,15]. Lim considered the asymmetric nature of the TG and introduced a two-population model of the TG with asymmetric demographic parameters, and it is shown that stochastic evolutionary dynamics with the asymmetric parameters can lead to the evolution of high trust and high trustworthiness [16]. Charness et al. studied the effect of reputation systems in which investors sometimes have foreknowledge of trustee behavior and found that reputation about trustees can boost the evolution of trust and trustworthiness [13].

Reward and punishment, as the common incentive means, have been extensively used to control the states of systems in the evolutionary game theory [17–40]. And the threat of punishment or the promise of reward can effectively induce self-interested players to prefer actions that promote the level of trust [41–43]. However, there are few works that have considered the reward and punishment mechanism into the TG and it is still not clear how the mechanism of reward and punishment influences the evolution of trust in the game from the theoretical perspective. For this purpose, we introduce individual reward and punishment into the NTG by considering that investors share a part of the returned fund to reward trustworthy trustees and meanwhile punish untrustworthy trustees. By using replicator equations, we find that the introduction of reward and punishment could induce the stable coexistence state of investors and trustworthy trustees. Furthermore, we investigate the impact of the incentive strength and reveal that the attraction domain of the

[a] e-mail: xiaojiechen@uestc.edu.cn (corresponding author)

coexistence state becomes larger with the increase of the incentive strength. We find that the attraction domain of the coexistence state is enlarged as the reward coefficient increases, which implies that reward can better promote the evolution of trust than punishment.

To make a thorough investigation into the mentioned problem, the remaining paper is divided into the following sections. In Sect. 2, we describe our model about the NTG with individual reward and punishment in an infinite well-mixed population. Then we present the theoretical analysis for our model in Sect. 3 and provide numerical calculations to illustrate our theoretical results in Sect. 4. Finally, we make our conclusion in Sect. 5.

2 Model

2.1 N-player trust game

We consider the NTG in an infinite well-mixed population. At each time step, N ($N > 2$) individuals are randomly chosen to form a group and offered the opportunity to participate in the game. Each individual could choose three strategies: being an investor also called trustor (strategy I), being a trustworthy trustee (strategy T), and being an untrustworthy trustee (strategy U). An investor needs to pay t_{v} to trustees, where $t_{\mathrm{v}} > 0$ represents the trusted value. If there are k_{I} investors, the total fund contributed by investors is $k_{\mathrm{I}}t_{\mathrm{v}}$. Then each trustee receives the same amount of the fund, which is $k_{\mathrm{I}}t_{\mathrm{v}}/k_{\mathrm{TU}}$, where k_{TU} is the number of trustees. Here we have $k_{\mathrm{TU}} = k_{\mathrm{T}} + k_{\mathrm{U}}$, where k_{T} is the number of trustworthy trustees and k_{U} is the number of untrustworthy trustees in the group. A trustworthy trustee then returns the received fund multiplied by R_{T} to investors and will have $R_{\mathrm{T}}k_{\mathrm{I}}t_{\mathrm{v}}/k_{\mathrm{TU}}$. At the same time, it can also have the same amount of return as the investor. However, untrustworthy trustees return nothing to investors and instead keep all they have for themselves. Each untrustworthy trustee finally has the received fund $R_{\mathrm{U}}k_{\mathrm{I}}t_{\mathrm{v}}/k_{\mathrm{TU}}$, where $1 < R_{\mathrm{T}} < R_{\mathrm{U}} < 2R_{\mathrm{T}}$. For simplicity, in this work we consider the temptation to defect ratio $r \in (0, 1)$ following previous work [14], which is defined as

$$r = \frac{R_{\mathrm{U}} - R_{\mathrm{T}}}{R_{\mathrm{T}}} \tag{1}$$

2.2 Reward and punishment

We then consider the mechanism of reward and punishment into our model. We assume that each investor shares a part of the returned fund from trustworthy trustees, $pR_{\mathrm{T}}k_{\mathrm{T}}t_{\mathrm{v}}/k_{\mathrm{TU}}$, which is used to reward T individuals and punish U individuals in the group. Here $p \in [0, 1]$ represents the incentive strength. For $p = 1$, investors contribute all their returned fund as the incentive budget. While for $p = 0$, they contribute nothing as the incentive budget. Furthermore, this incentive budget is divided into two parts. The first part is used to reward T players and each T player in the group can get a reward $\alpha pR_{\mathrm{T}}k_{\mathrm{I}}t_{\mathrm{v}}/k_{\mathrm{TU}}$; the second part is used to punish U players and each U player in the group can get a fine $(1-\alpha)pR_{\mathrm{T}}k_{\mathrm{I}}k_{\mathrm{T}}t_{\mathrm{v}}/(k_{\mathrm{TU}}k_{\mathrm{U}})$. Here α ($0 \le \alpha \le 1$) represents the reward coefficient. For $\alpha = 1$, investors use all the incentive budget to reward T individuals. While for $\alpha = 0$, they use all the incentive budget to punish U individuals in the group.

Accordingly, the payoffs of investors, trustworthy trustees, and untrustworthy trustees can be respectively written as

$$\pi_{\mathrm{I}} = \begin{cases} (1-p)\dfrac{R_{\mathrm{T}}N_{\mathrm{T}}}{N - N_{\mathrm{I}} - 1}t_{\mathrm{v}} - t_{\mathrm{v}}, & \text{if } N_{\mathrm{I}} \neq N-1; \\ 0, & \text{otherwise} \end{cases} \tag{2}$$

$$\pi_{\mathrm{T}} = \begin{cases} (1+\alpha p)\dfrac{R_{\mathrm{T}}N_{\mathrm{I}}}{N - N_{\mathrm{I}}}t_{\mathrm{v}}, & \text{if } N_{\mathrm{I}} \neq N; \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

$$\pi_{\mathrm{U}} = \begin{cases} \dfrac{R_{\mathrm{U}}N_{\mathrm{I}}}{N - N_{\mathrm{I}}}t_{\mathrm{v}} - \dfrac{(1-\alpha)pN_{\mathrm{I}}N_{\mathrm{T}}R_{\mathrm{T}}}{(N-N_{\mathrm{I}})(N_{\mathrm{U}}+1)}t_{\mathrm{v}}, & \text{if } N_{\mathrm{I}} \neq N; \\ 0, & \text{otherwise} \end{cases} \tag{4}$$

where N_{I}, N_{T}, and N_{U} represent the number of investors, trustworthy trustees, and untrustworthy trustees among other $N-1$ players in the group, respectively.

2.3 Replicator equations

The evolutionary behavior of a population playing the trust game could be studied by replicator dynamics [44–47]. We let x be the frequency of I players in the population, y be the frequency of T players, and z be the frequency of U players with $x+y+z = 1$. The evolution of trust can be described by the following equations

$$\begin{cases} \dot{x} = x(f_{\mathrm{I}} - \phi), \\ \dot{y} = y(f_{\mathrm{T}} - \phi), \\ \dot{z} = z(f_{\mathrm{U}} - \phi), \end{cases} \tag{5}$$

where f_{i} denotes the expected payoff of strategy i ($i = I$, T, or U) and $\phi = xf_{\mathrm{I}} + yf_{\mathrm{T}} + zf_{\mathrm{U}}$ represents the average payoff of the whole population.

Here the expected payoff of strategy I is given by

$$\begin{aligned} f_{\mathrm{I}} &= \sum_{N_{\mathrm{I}}=0}^{N-1}\sum_{N_{\mathrm{T}}=0}^{N-1-N_{\mathrm{I}}}\binom{N-1}{N_{\mathrm{I}}}\binom{N-1-N_{\mathrm{I}}}{N_{\mathrm{T}}} \\ &\quad \times x^{N_{\mathrm{I}}}y^{N_{\mathrm{T}}}z^{N-1-N_{\mathrm{I}}-N_{\mathrm{T}}}\pi_{\mathrm{I}} \\ &= t_{\mathrm{v}}[\frac{(1-p)R_{\mathrm{T}}(1-x-z)}{1-x} - 1](1-x^{N-1}). \end{aligned} \tag{6}$$

And the expected payoff of strategy T is given by

$$\begin{aligned} f_{\mathrm{T}} &= \sum_{N_{\mathrm{I}}=0}^{N-1} \sum_{N_{\mathrm{T}}=0}^{N-1-N_{\mathrm{I}}} \binom{N-1}{N_{\mathrm{I}}} \binom{N-1-N_{\mathrm{I}}}{N_{\mathrm{T}}} \\ &\quad \times x^{N_{\mathrm{I}}} y^{N_{\mathrm{T}}} z^{N-1-N_{\mathrm{I}}-N_{\mathrm{T}}} \pi_{\mathrm{T}} \\ &= \frac{(1+\alpha p) R_{\mathrm{T}} t_{\mathrm{v}} x}{1-x} (1-x^{N-1}). \end{aligned} \tag{7}$$

Similarly, we can write the expected payoff of strategy U as

$$\begin{aligned} f_{\mathrm{U}} &= \sum_{N_{\mathrm{I}}=0}^{N-1} \sum_{N_{\mathrm{T}}=0}^{N-1-N_{\mathrm{I}}} \binom{N-1}{N_{\mathrm{I}}} \binom{N-1-N_{\mathrm{I}}}{N_{\mathrm{T}}} \\ &\quad \times x^{N_{\mathrm{I}}} y^{N_{\mathrm{T}}} z^{N-1-N_{\mathrm{I}}-N_{\mathrm{T}}} \pi_{\mathrm{U}} \\ &= \frac{(1+r+p-\alpha p) R_{\mathrm{T}} t_{\mathrm{v}} x}{1-x} (1-x^{N-1}) \\ &\quad - \frac{(1-\alpha) p R_{\mathrm{T}} t_{\mathrm{v}} x}{z} [1-(1-z)^{N-1}]. \end{aligned} \tag{8}$$

Below, we study the evolutionary dynamics of trust in the N-player trust game in an infinite well-mixed population by means of theoretical analysis and numerical calculations.

3 Theoretical analysis

Since we have $y = 1 - x - z$, thus the system equation can be rewritten as

$$\begin{cases} \dot{x} = x[(1-x)(f_{\mathrm{I}} - f_{\mathrm{T}}) + z(f_{\mathrm{T}} - f_{\mathrm{U}})], \\ \dot{z} = z[(1-z)(f_{\mathrm{U}} - f_{\mathrm{T}}) + x(f_{\mathrm{T}} - f_{\mathrm{I}})], \end{cases} \tag{9}$$

where

$$f_{\mathrm{I}} - f_{\mathrm{T}} = t_{\mathrm{v}} [\frac{(1-p)(1-z)R_{\mathrm{T}} - (2+\alpha p - p)R_{\mathrm{T}} x}{1-x} - 1] \times (1-x^{N-1})$$

and

$$\begin{aligned} f_{\mathrm{T}} - f_{\mathrm{U}} &= \frac{(2\alpha p - r - p) R_{\mathrm{T}} t_{\mathrm{v}} x}{1-x} (1-x^{N-1}) \\ &\quad + \frac{(1-\alpha) p R_{\mathrm{T}} t_{\mathrm{v}} x}{z} [1-(1-z)^{N-1}]. \end{aligned}$$

Solving $f_{\mathrm{I}} = f_{\mathrm{T}}$ results in

$$z = \frac{(1-p)R_{\mathrm{T}} - 1 - [(2+\alpha p - p)R_{\mathrm{T}} - 1]x}{(1-p)R_{\mathrm{T}}}. \tag{10}$$

Based on the above analysis, we have Theorem 1 about the equilibrium points in the system equation.

Theorem 1 *For $0 < r < 1$, $0 \le p \le 1$, $0 \le \alpha \le 1$, and $(1-p)R_{\mathrm{T}} > 1$, the system (5) or (9) has a continuum of fixed points $(x, y, z) = (0, m, 1-m)$ on the TU edge and a unique fixed point $(x, y, z) = (\frac{(1-p)R_{\mathrm{T}}-1}{(2+\alpha p-p)R_{\mathrm{T}}-1}, \frac{(1+\alpha p)R_{\mathrm{T}}}{(2+\alpha p-p)R_{\mathrm{T}}-1}, 0)$ on the IT edge. Besides, there are three vertex fixed points, namely, $(x, y, z) = (1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$.*

Proof (1) On the edge of TU, we have $x = 0$, resulting in $f_{\mathrm{T}} = f_{\mathrm{U}} = 0$. Thus regardless of the value of z, we have $\dot{x} \equiv 0$ and $\dot{z} \equiv 0$. So there exists a continuum of fixed points $(x, y, z) = (0, m, 1-m)$ on the TU edge.

(2) On the edge of IT, we have $z = 0$. By solving the equation $f_{\mathrm{I}} = f_{\mathrm{T}}$, we get

$$\begin{cases} x = \dfrac{(1-p)R_{\mathrm{T}} - 1}{(2+\alpha p - p)R_{\mathrm{T}} - 1}, \\ y = \dfrac{(1+\alpha p)R_{\mathrm{T}}}{(2+\alpha p - p)R_{\mathrm{T}} - 1}. \end{cases} \tag{11}$$

Thus, we have a unique fixed point $(x, y, z) = (\frac{(1-p)R_{\mathrm{T}}-1}{(2+\alpha p-p)R_{\mathrm{T}}-1}, \frac{(1+\alpha p)R_{\mathrm{T}}}{(2+\alpha p-p)R_{\mathrm{T}}-1}, 0)$ on the IT edge.

(3) On the edge of IU, since $y = 0$ and $x + z = 1$, we have $\dot{z} = z(1-z)(f_{\mathrm{U}} - f_{\mathrm{I}}) > 0$. Thus, the direction of the dynamics goes from I to U and there is no fixed point on the IU edge.

(4) For the vertex points, it is obvious to find that the three points $(1, 0, 0)$, $(0, 1, 0)$, and $(0, 0, 1)$ can always make $\dot{x} = 0$ and $\dot{z} = 0$, which hence are the fixed points. □

Since the mathematical expressions of the payoffs of strategies in the population are high-order and nonlinear, it is extremely difficult to obtain the interior equilibrium points for the system equation. Here we only give a sufficient condition in which the interior equilibrium points do not exist, which is described by Lemma 1.

Lemma 1 *In the system (5) or (9), there is no interior equilibrium point when $\alpha p > r$ is satisfied.*

Proof Since $x \in [0, 1]$, $z \in [0, 1]$, $1 - z \ge x$, and $1 - x^{N-1} = (1-x)\sum_{k=0}^{N-2} x^k$, we have

$$\begin{aligned} f_{\mathrm{T}} - f_{\mathrm{U}} &= \frac{(2\alpha p - r - p) R_{\mathrm{T}} t_{\mathrm{v}} x}{1-x} (1-x^{N-1}) \\ &\quad + \frac{(1-\alpha) p R_{\mathrm{T}} t_{\mathrm{v}} x}{z} [1-(1-z)^{N-1}] \\ &= (2\alpha p - r - p) R_{\mathrm{T}} t_{\mathrm{v}} x \sum_{k=0}^{N-2} x^k \\ &\quad + (1-\alpha) p R_{\mathrm{T}} t_{\mathrm{v}} x \sum_{k=0}^{N-2} (1-z)^k \\ &\ge (\alpha p - r) R_{\mathrm{T}} t_{\mathrm{v}} x \sum_{k=0}^{N-2} x^k \\ &\ge 0 \end{aligned}$$

When $x = 0$, we have $f_T = f_U = 0$. So when $\alpha p > r$, $f_T - f_U > 0$ if $x > 0$. Correspondingly, there is no interior equilibrium point in this condition. □

Furthermore, in order to study the fixed points in the system and do the stability analysis, we define that

$$\begin{cases} f(x,z) = x[(1-x)(f_I - f_T) + z(f_T - f_U)], \\ g(x,z) = z[(1-z)(f_U - f_T) + x(f_T - f_I)]. \end{cases} \quad (12)$$

Then the Jacobian matrix of the system is

$$J = \begin{bmatrix} \frac{\partial f(x,z)}{\partial x} & \frac{\partial f(x,z)}{\partial z} \\ \frac{\partial g(x,z)}{\partial x} & \frac{\partial g(x,z)}{\partial z} \end{bmatrix}, \quad (13)$$

where

$$\begin{cases} \frac{\partial f(x,z)}{\partial x} = [(1-x)(f_I - f_T) + z(f_T - f_U)] \\ \qquad + x[-(f_I - f_T) + (1-x)\frac{\partial}{\partial x}(f_I - f_T) \\ \qquad + z\frac{\partial}{\partial x}(f_T - f_U)], \\ \frac{\partial f(x,z)}{\partial z} = x[f_T - f_U + (1-x)\frac{\partial}{\partial z}(f_I - f_T) \\ \qquad + z\frac{\partial}{\partial z}(f_T - f_U)], \\ \frac{\partial g(x,z)}{\partial x} = z[f_T - f_I + (1-z)\frac{\partial}{\partial x}(f_U - f_T) \\ \qquad + x\frac{\partial}{\partial x}(f_T - f_I)], \\ \frac{\partial g(x,z)}{\partial z} = [(1-z)(f_U - f_T) + x(f_T - f_I)] \\ \qquad + z[-(f_U - f_T) + (1-z)\frac{\partial}{\partial z}(f_U - f_T) \\ \qquad + x\frac{\partial}{\partial z}(f_T - f_I)]. \end{cases}$$

Accordingly, we have some following results about the stability of the fixed points described by Theorem 2.

Theorem 2 *For $0 < r < 1$, $0 \le p \le 1$, $0 \le \alpha \le 1$, and $(1-p)R_T > 1$, the equilibrium points $(1,0,0)$,$(0,1,0)$ are unstable and the equilibrium point $(0,0,1)$ is stable. And when $\alpha p > r$, the boundary equilibrium point $(x,y,z) = (\frac{(1-p)R_T - 1}{(2-p+\alpha p)R_T - 1}, \frac{(1+\alpha p)R_T}{(2-p+\alpha p)R_T - 1}, 0)$ is stable. Besides, when $m > \frac{1}{(1-p)R_T}$, the continuum of equilibrium points on the TU edge is unstable.*

Proof (1) For $(x,y,z) = (1,0,0)$, the Jacobian is

$$J(1,0,0) = \begin{bmatrix} f_T - f_I & f_T - f_U \\ 0 & f_U - f_I \end{bmatrix}, \quad (14)$$

where $f_T - f_I = (1+\alpha p)R_T t_v (N-1) > 0$, thus the fixed point is unstable.

(2) For $(x,y,z) = (0,1,0)$, the Jacobian is

$$J(0,1,0) = \begin{bmatrix} [(1-p)R_T - 1]t_v & 0 \\ 0 & 0 \end{bmatrix}. \quad (15)$$

Because $(1-p)R_T > 1$, thus the fixed point is unstable.

(3) For $(x,y,z) = (0,0,1)$, the Jacobian is

$$J(0,0,1) = \begin{bmatrix} -t_v & 0 \\ t_v & 0 \end{bmatrix}. \quad (16)$$

For fixed point $(0,0,1)$, because $f_T - f_I > 0$ always holds with the condition $z \to 1_-$, which leads to $\dot{z} = z[(1-z)(f_U - f_T) + x(f_T - f_I)] > 0$ in the interior space near the equilibrium point $(0,0,1)$. Thus, z will always evolve into the $z = 1$ state for small perturbations at fixed point $(0,0,1)$ [48,49]. So the fixed point $(0,0,1)$ is stable.

(4) For $(x,y,z) = (\frac{(1-p)R_T - 1}{(2-p+\alpha p)R_T - 1}, \frac{(1+\alpha p)R_T}{(2-p+\alpha p)R_T - 1}, 0)$, referred to as $(x_0, y_0, 0)$, the corresponding Jacobian is

$$J(x_0, y_0, 0) = \begin{bmatrix} a_{11} & a_{12} \\ 0 & a_{22} \end{bmatrix}, \quad (17)$$

where

$$\begin{aligned} a_{11} &= \frac{\partial f}{\partial x}(x_0, z_0) \\ &= [(1-p)R_T - 1][(\frac{(1-p)R_T - 1}{(2+\alpha p - p)R_T - 1})^{N-1} - 1]t_v < 0, \\ a_{12} &= \frac{\partial f}{\partial z}(x_0, z_0) \\ &= \frac{(1-p)R_T[(1-p)R_T - 1](1 - x_0^{N-1})t_v}{(2+\alpha p - p)R_T - 1} \\ &\quad - \frac{[(1-p)R_T - 1]^2 t_v}{[(2+\alpha p - p)R_T - 1]^2}(1-\alpha)(1-N)pR_T \\ &\quad - \frac{[(1-p)R_T - 1]^2(1 - x_0^{N-1})t_v}{(1+\alpha p)[(2+\alpha p - p)R_T - 1]}(p + r - 2\alpha p), \end{aligned}$$

and

$$\begin{aligned} a_{22} &= \frac{\partial g}{\partial z}(x_0, z_0) \\ &= f_U - f_T \\ &= \frac{(p + r - 2\alpha p)R_T t_v x_0}{1 - x_0}(1 - x_0^{N-1}) \\ &\quad - \frac{[(1-p)R_T - 1](1-\alpha)pR_T t_v}{(2+\alpha p - p)R_T - 1}(N-1). \end{aligned}$$

When $\alpha p > r$ is satisfied, we have $a_{22} = f_U - f_T < 0$ according to the Lemma 1. Thus, the fixed point $(x,y,z) = (\frac{(1-p)R_T - 1}{(2-p+\alpha p)R_T - 1}, \frac{(1+\alpha p)R_T}{(2-p+\alpha p)R_T - 1}, 0)$ is stable. In particular, for very large N, the positive or negative sign of a_{22} will be determined by the second term of right side $-\frac{[(1-p)R_T - 1](1-\alpha)pR_T t_v}{(2+\alpha p - p)R_T - 1}(N-1)$. In

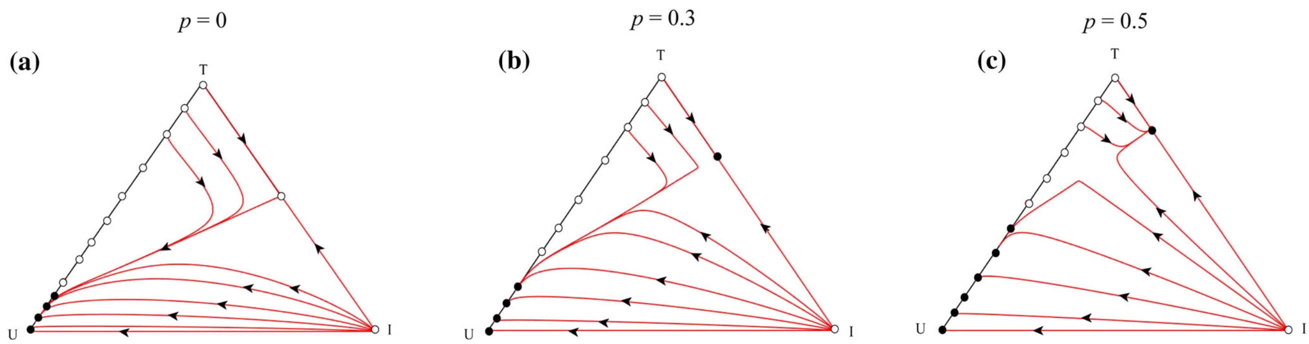

Fig. 1 Evolutionary dynamics of the population with investors (I), trustworthy trustees (T), and untrustworthy trustees (U) in the S_3 for different values of incentive strength p. The triangle represents the state space $S_3 = \{(x, y, z) : x, y, z \geq 0, x + y + z = 1\}$, where x, y, and z are the frequencies of I, T, and U, respectively. Open circles represent unstable equilibrium points and filled circles represent stable equilibrium points. Other parameter values: $N = 20$, $\alpha = 0.5$, $r = \frac{1}{3}$, $R_{\rm T} = 6$, and $t_{\rm v} = 1$

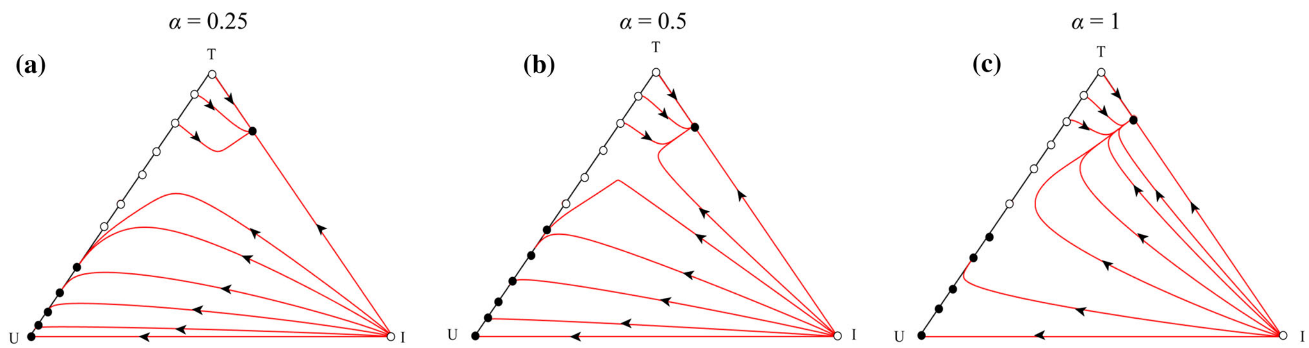

Fig. 2 Evolutionary dynamics of the population with investors (I), trustworthy trustees (T), and untrustworthy trustees (U) in the S_3 for different reward coefficient values α. The triangle represents the state space $S_3 = \{(x, y, z) : x, y, z \geq 0, x + y + z = 1\}$, where x, y, and z are the frequencies of I, T, and U, respectively. Open circles represent unstable equilibrium points and filled circles represent stable equilibrium points. Other parameter values: $N = 20$, $p = 0.5$, $r = \frac{1}{3}$, $R_{\rm T} = 6$, and $t_{\rm v} = 1$

this case, we also have $a_{22} < 0$ and accordingly the fixed point is stable.

(5) For $(x, y, z) = (0, m, 1 - m)$, the Jacobian matrix is

$$J(0, m, 1 - m) = \begin{bmatrix} a_{11} & 0 \\ a_{21} & 0 \end{bmatrix}, \qquad (18)$$

where

$$a_{11} = \frac{\partial f}{\partial x}(x, z) = [(1 - p)R_{\rm T}m - 1]t_{\rm v}$$

and

$$\begin{aligned} a_{21} &= \frac{\partial g}{\partial x}(x, z) \\ &= z[-f_{\rm I} + (1 - z)\frac{\partial}{\partial x}(f_{\rm U} - f_{\rm T})] \\ &= (1 - m)t_{\rm v} - mt_{\rm v}R_{\rm T}(1 + \alpha p - r - p) \\ &\quad -m^2 t_{\rm v} R_{\rm T}[-1 + r + p(\alpha - 1)(2 - m^{N-2})]. \end{aligned}$$

When $(1 - p)R_{\rm T}m > 1$, we have $a_{11} > 0$, in this case the corresponding fixed points are unstable. Based on Refs. [50–52], we can know that the segment consisting of the fixed points is stable. □

4 Numerical calculations

We now present some numerical calculations to confirm our theoretical results shown above. As shown in Figs. 1 and 2, we can see that there are three vertex equilibrium points, that is, all I $(x = 1)$, all T $(y = 1)$, and all U $(z = 1)$, corresponding to three vertices of the simplex S_3, respectively. We can find that only the fixed point corresponding to the U vertex is stable. On the edge IT, there exists a stable boundary fixed point (see Figs. 1b, 1c, and 2), which indicates that the evolution of trust can be greatly promoted. When $p = 0$, the mechanism of reward and punishment does not work and the system converges to the state without investors and with many untrustworthy individuals (see Fig. 1a). Besides, on the edge TU there exists a continuum of unstable

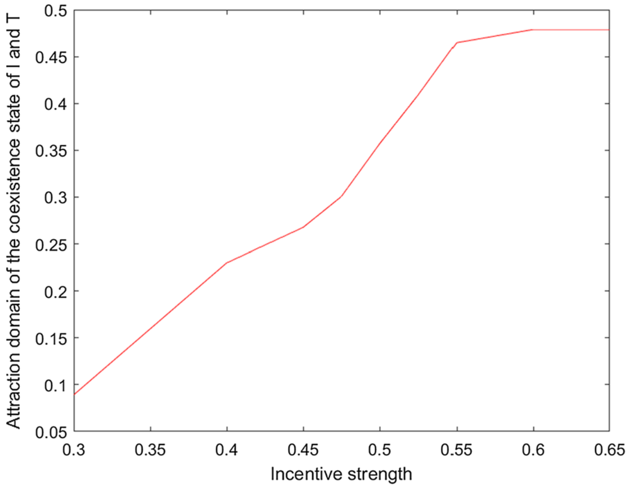

Fig. 3 The attraction domain of the coexistence state of investors and trustworthy trustees as a function of incentive strength p. Other parameter values: $N = 20$, $r = \frac{1}{3}$, $R_{\mathrm{T}} = 6$, $\alpha = 0.5$, and $t_{\mathrm{v}} = 1$

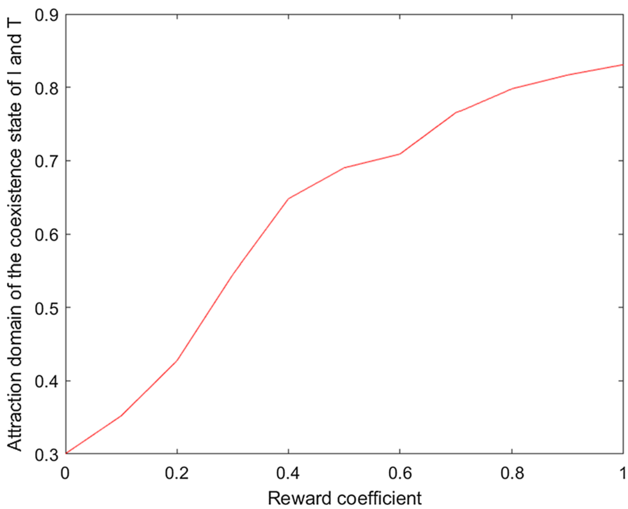

Fig. 4 The attraction domain of the coexistence state of investors and trustworthy trustees as a function of reward coefficient α. Other parameter values: $N = 20$, $r = \frac{1}{5}$, $R_{\mathrm{T}} = 6$, $p = 0.5$, and $t_{\mathrm{v}} = 1$

equilibrium points (see Figs. 1 and 2). These numerical results agree well with the theoretical findings presented in Theorems 1 and 2.

In order to better illustrate the role of the considered reward and punishment in the evolutionary outcomes, we first present the attraction domain of the coexistence state as a function of incentive strength, as shown in Fig. 3. We can see that the attraction domain of the coexistence state becomes larger with the increase of the incentive strength. This reflects that the evolution of trust can be better promoted when investors are willing to give more returns to trustees. Furthermore, we investigate how the reward coefficient influences the attraction domain of the coexistence state. As shown in Fig. 4, we present the attraction domain in dependence of the reward coefficient α. We find that the attraction domain of the coexistence state is enlarged as the reward coefficient increases, which implies that in our model when investors allocate more incentive budget as the reward, the evolution of trust can be better enhanced and indicates that reward can perform better than punishment for the evolution of trust. Indeed this result can be understood intuitively since the promise of reward is more attractive for trustees and reward could lead to a win-win situation for both sides [39].

5 Conclusion

In this study, we have introduced the mechanism of reward and punishment into an infinite well-mixed population of agents who play the N-player trust game, and then investigated the evolutionary dynamics of trust by means of replicator equations. We have shown that the introduction of reward and punishment can lead to the coexistence of investors and trustworthy trustees, which indicates that the evolution of trust can be greatly promoted and accordingly the system can reach a rational and optimal social state. We have also investigated the effect of incentive strength and found that the attraction domain of the coexistence state becomes larger as investors increase the incentive strength from the returned fund for reward and punishment. In addition, we have revealed that compared with punishment reward can perform better for the evolution of trust. Finally, we would like to stress that we focus on well-mixed populations in our present study where individuals perform rand interactions. However, the interactions among individuals are typically not random but rather that they are limited to a set of neighbors in a structured population [27]. Accordingly, our present results which are valid in well-mixed populations may be changed in structured populations, which is worth investigating for future study.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant nos. 61976048, 62036002) and the Fundamental Research Funds of the Central Universities of China.

Author contributions

XC and XF designed the research, XF performed the research, and XF and XC wrote the manuscript.

Data Availability Statement This manuscript has no associated data or the data will not be deposited. [Authors' comment: The data that support the results of this study are available from the corresponding authors upon request.]

References

1. A. Baier, Ethics **96**, 231 (1986)
2. E. Fehr, U. Fischbacher, Nature **425**, 785 (2003)
3. M.R. Delgado, R.H. Frank, E.A. Phelps, Nat. Neurosci. **8**, 1611 (2005)
4. M. Berens, J. Organ. Chang. Manag. **20**, 463 (2007)
5. E. Petraki, H.A. Abbass, in Proceedings of 7th IEEE Symposium on Computational Intelligence for Security and Defense Applications, Hanoi, Vietnam (2014)
6. A.W. Brooks, H. Dai, M.E. Schweitzer, Soc. Psychol. Personal. Sci. **5**, 467 (2014)
7. F.S. Grodzinsky, K.W. Miller, M.J. Wolf, Philos. Technol. **28**, 91 (2015)
8. R. Golembiewski, Am. Polit. Sci. Rev. **75**, 480 (1981)
9. S. Knack, P. Keefer, Q. J. Econ. **112**, 1251 (1977)
10. A. Alesina, E.L. Ferrara, J. Pubilc Econ. **85**, 207 (2002)
11. J. Berg, J. Dickhaut, K. McCabe, Games Econ. Behav. **10**, 122 (1995)
12. H.A. Abbass, G. Greenwood, E. Petraki, IEEE Trans. Evol. Comput. **20**, 470 (2016)
13. G. Charness, N. Du, C. Yang, Games Econ. Behav. **72**, 361 (2011)
14. M. Chica, R. Chiong, M. Kirley, H.A. Ishibuchi, IEEE Trans. Evol. Comput. **22**, 866 (2018)
15. M. Chica, R. Chiong, J. Ramasco, H.A. Abbass, Commun. Nonlinear Sci. Numer. Simul. **79**, 104870 (2019)
16. I.S. Lim, Phys. Rev. E **102**, 062419 (2020)
17. K. Sigmund, Trends Ecol. Evol. **22**, 593 (2007)
18. A. Szolnoki, M. Perc, EPL **92**, 38003 (2010)
19. A. Szolnoki, G. Szabó, M. Perc, Phys. Rev. E **83**, 036101 (2011)
20. A. Szolnoki, M. Perc, New J. Phys. **14**, 093016 (2012)
21. M. Perc, Sci. Rep. **2**, 344 (2012)
22. N.J. Raihani, A. Thornton, R. Bshary, Trends Ecol. Evol. **27**, 288 (2012)
23. A. Szolnoki, M. Perc, Phys. Rev. X **3**, 041021 (2013)
24. X. Chen, A. Szolnoki, M. Perc, New J. Phys. **16**, 083016 (2014)
25. K. Li, R. Cong, T. Wu, L. Wang, Phys. Rev. E **91**, 042810 (2015)
26. C. Wang, L. Wang, J. Wang, S. Sun, C. Xia, Appl. Math. Comput. **293**, 18 (2017)
27. M. Perc, J.J. Jordan, D.G. Rand, Z. Wang, S. Boccaletti, A. Szolnoki, Phys. Rep. **687**, 1 (2017)
28. C. Du, D. Jia, L. Jin, L. Shi, Eur. Phys. J. B **91**, 234 (2018)
29. J. Riehl, P. Ramazi, M. Cao, Annu. Rev. Control **45**, 87 (2018)
30. L. Liu, S. Wang, X. Chen, M. Perc, Chaos **28**, 103105 (2018)
31. T.A. Han, L.T. Thanh, Sci. Rep. **8**, 15997 (2018)
32. T. Chen, X. Hu, Y. Wang, L. Wang, Eur. Phys. J. B **91**, 112 (2018)
33. Y. Zhang, F. Fu, PLOS One **13**, 5 (2018)
34. Y. Zhang, X. Chen, A. Liu, C. Sun, Appl. Math. Comput. **321**, 641 (2018)
35. L. Ratliff, R. Dong, S. Sekar, T. Fiez, Annu. Rev. Control Robot. Auton. Syst. **2**, 305 (2019)
36. X. Li, H. Wang, C. Xia, M. Perc, Front. Phys. **7**, 125 (2019)
37. Y. Zhang, J. Liu, A. Li, Complexity **2019**, 2915020 (2019)
38. Y. Fang, T.P. Benko, M. Perc, H. Xu, Q. Tan, Proc. R. Soc. A. **475**, 20190349 (2019)
39. L. Hu, N. He, Q. Weng, X. Chen, M. Perc, Chaos Soliton. Fract **134**, 109694 (2020)
40. Q. Chang, Y. Zhang, Eur. Phys. J. B **94**, 100 (2021)
41. M. Rigdon, J. Econ. Behav. Organ. **70**, 93 (2009)
42. H. Sun, M. Bigoni, J. Behav. Exp. Econ. **77**, 158 (2018)
43. X. Pan, D. Houser, J. Econ. Behav. Organ. **158**, 236 (2019)
44. J. Hofbauer, K. Sigmund, *Evolutionary Games and Population Dynamics* (Cambridge University Press, Cambridge, 1998)
45. M.A. Nowak, *Evolutionary Dynamics: Exploring the Equations of Life* (Harvard University Press, Harvard, 2006)
46. K. Sigmund, *The Calculus of Selfishness* (Princeton University Press, Princeton, 2010)
47. W.H. Sandholm, *Population Games and Evolutionary Dynamics* (MIT Press, London, 2010)
48. T. Sasaki, Å. Brännström, U. Dieckmann, K. Sigmund, Proc. Natl. Acad. Sci. USA **109**, 1165 (2012)
49. F. Huang, X. Chen, L. Wang, J. Theor. Biol. **449**, 60 (2018)
50. C. Hauert, N. Haiden, K. Sigmund, Discret. Contin. Dyn. Syst. Ser. B. **4**, 575 (2004)
51. C. Hauert, A. Traulsen, H. De Silva née Brandt, M.A. Nowak, K. Sigmund, Biol. Theor. **3**, 114 (2008)
52. T. Sasaki, S. Uchida, Proc. R. Soc. B. **280**, 20122498 (2013)