



Game Theory and the Evolution of Cooperation

Bo-Yu Zhang¹ · Shan Pei¹

Received: 14 October 2020 / Revised: 9 April 2021 / Accepted: 30 April 2021 / Published online: 12 June 2021
© Operations Research Society of China, Periodicals Agency of Shanghai University, Science Press, and Springer-Verlag GmbH Germany, part of Springer Nature 2021

Abstract

Evolution is based on the competition between individuals and therefore rewards only selfish behavior. How cooperation or altruism behavior could prevail in social dilemma then becomes a problematic issue. Game theory offers a powerful mathematical approach for studying social behavior. It has been widely used to explain the evolution of cooperation. In this paper, we first introduce related static and dynamic game methods. Then we review two types of mechanisms that can promote cooperation in groups of genetically unrelated individuals, (i) direct reciprocity in repeated games, and (ii) incentive mechanisms such as reward and punishment.

Keywords Cooperation · Game theory · Direct reciprocity · Reward · Punishment

Mathematics Subject Classification 91A80 · Applications of game theory

1 Introduction

“How did cooperative behavior evolve” is a fundamental question in biology and the social sciences [1]. A great deal of game-theoretic research has been devoted to explain the prevalence of cooperation in biological systems as well as in human society. One reason for the vast literature from members of the game theory community on this topic is that their methods do not work for the underlying game, the Prisoner’s Dilemma (PD) game, which pits cooperative behavior against its nemesis of defection. In particular,

This work was supported by the National Natural Science Foundation of China (Nos. 71771026 and 71922004).

✉ Bo-Yu Zhang
zhangby@bnu.edu.cn
Shan Pei
peishan113@163.com

¹ Laboratory of Mathematics and Complex Systems, Ministry of Education; School of Mathematical Sciences, Beijing Normal University, Beijing 100875, China

defection strictly dominates cooperation since defectors always have higher payoffs by enjoying the benefits from cooperators without paying. That is, the only rational option in the PD game is to defect although it is to everyone's advantage for all players to cooperate.

In the last decades, several mechanisms have been proposed to promote cooperation. Cooperation can be rational when the payoffs of the PD game are modified by assuming some relatedness between the players, by them playing the game an uncertain number of times, by extending the model to a game based on reputation, or introducing incentive mechanisms such as reward and punishment. These predictions are often based on applying either static (e.g., Nash equilibrium analysis and evolutionarily stable strategy) or dynamic methods (e.g., replicator equation) from game theory that assumes a large population of agents paired at random to play the game. Population interactions that are structured either spatially or socially also enhance the evolution of cooperation as do the stochastic effects of finite populations.

In this paper, we review recent studies on the evolution of cooperation. In particular, we focus on game mechanisms that can promote cooperation in groups of genetically unrelated individuals. These mechanisms include direct reciprocity in repeated games and incentive mechanisms such as reward and punishment. The rest of this paper is organized as follows. Section 2 introduces some basic games, equilibrium notations, and dynamic methods in game theory. Section 3 focuses on the repeated game and discusses strategies based on direct reciprocity. Section 4 considers the effects of reward and punishment in one-shot game. Section 5 summarizes and discusses the effects of these two types of mechanisms.

2 Game theory and Evolutionary Dynamics

2.1 Basic Games

2.1.1 The Prisoner's Dilemma Game

A paradigm for modelling human cooperation is the Prisoner's Dilemma (PD) game [2–4]. The standard PD game is a two-player game, where each player has two strategies, cooperation (C) and defection (D). The payoff matrix for the PD game is shown in Eq. (1). Two players are offered a certain payoff, R (i.e., reward), for mutual cooperation, and a lower payoff, P (i.e., punishment), for mutual defection. If one player cooperates while the other defects, then the cooperator gets the lowest payoff, S (i.e., sucker's payoff), and the defector gains the highest payoff, T (i.e., the temptation of defection). Thus, the payoffs satisfy $T > R > P > S$. A further common assumption is $2R > S + T > 2P$, which means that mutual cooperation is the best outcome and mutual defection is the worst outcome.

$$\begin{array}{c} \text{C D} \\ \text{C} \begin{pmatrix} R & S \\ T & P \end{pmatrix} \\ \text{D} \end{array} \quad (1)$$

In the one-shot PD game, D strictly dominates C because a player is better off defecting than cooperating no matter what the opponent does. So mutual defection is the only rational outcome.

We further introduce a simplified version of the PD game, the Donation game [3, 4]. This game has been widely used to study the evolution of cooperation. In this game, C means that a donor pays a cost, c , for a recipient to get a benefit, b , where $b > c$, D means that the donor pays no cost and distributes no benefit. The payoff matrix is then given by

$$\begin{array}{cc} & \begin{array}{c} C \quad D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} b - c & -c \\ b & 0 \end{pmatrix} \end{array} \tag{2}$$

Similarly as the standard PD game, D is a dominant strategy for both players and mutual cooperation leads to a higher average payoff.

2.1.2 The Public Goods Game

The public goods game (PGG) is a generalization of the PD game [4]. In a PGG, each player in a group of size n is given a fixed endowment, c , and chooses how much of that endowment to put into a common pool. The total amount in the pool is multiplied by a factor r with $1 < r < n$ and then redistributed evenly to each player in the group. Denote the contribution of player i by x_i , with $0 \leq x_i \leq c$. The payoff of player i is then the remainder of his/her endowment $c - x_i$ plus what he/she receives from the public pool, which has the form

$$u_i(\mathbf{x}) = c - x_i + \frac{r \sum_{j=1}^n x_j}{n} = c - \left(1 - \frac{r}{n}\right)x_i + \frac{r \sum_{j \neq i} x_j}{n} \tag{3}$$

Furthermore, the average payoff of the group is

$$\bar{u}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n u_i(\mathbf{x}) = c + \frac{(r - 1)}{n} \sum_{i=1}^n x_i \tag{4}$$

Thus, it is to the group’s advantage if all players contribute their total endowment because $r > 1$, but each player, given the contributions of the others, does best by contributing nothing because $r < n$.

To show the relation between PGG and PD game, we consider a two-player PGG with discrete choices. We define C as contributing all the endowment and D as contributing nothing. Thus, the payoff matrix for this two-player PGG can be written as

$$\begin{array}{cc} & \begin{array}{c} C \quad D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} cr & \frac{cr}{2} \\ c(1 + \frac{r}{2}) & c \end{pmatrix} \end{array} \tag{5}$$

This simplified PGG is a special case of the PD game, where D is the only dominant strategy.

2.2 Equilibrium Analysis

2.2.1 Nash Equilibrium

The most commonly used equilibrium notation in game theory is the Nash equilibrium (NE). In a game, if each player has chosen a strategy and no player can benefit by changing his/her strategy while the other players keep theirs unchanged, then the set of strategy choices is called a Nash equilibrium [5].

Consider a 2-person n -strategy symmetric game, where for each player $S = \{s_1, s_2, \dots, s_n\}$ is the strategy set. A strategy pair (s_j, s_i) is a NE if $u(s_j, s_i) \leq u(s_i, s_i)$ for any $j \neq i$, where $u(s_j, s_i)$ is the payoff to the player using strategy j when the other player adopts strategy i .

It has been shown that every game has at least one Nash equilibrium [5], but in general there are many. Trying to select the “best” equilibrium for each game is a difficult problem. Methods to do this have been suggested by Harsanyi and Selten [6], inventing the risk dominant equilibrium, and by many other researchers.

2.2.2 Evolutionarily Stable Strategy

One well-known refinement for NE is the evolutionarily stable strategy (ESS). More than 40 years have passed since the concept of ESS was introduced by Maynard Smith, evolutionary game theory has become a powerful tool to analyze evolutionary processes and human behaviors [7, 8].

A strategy is called evolutionarily stable if all the members of the population adopted it, then no other strategy could invade the population. In a 2-person symmetric game, a strategy s_i is an ESS if for any $j \neq i$, (i) $u(s_j, s_i) \leq u(s_i, s_i)$, and (ii) $u(s_j, s_j) \leq u(s_i, s_j)$ when $u(s_j, s_i) = u(s_i, s_i)$. Condition (i) is the NE condition and condition (ii) guarantees that the strategy is evolutionarily stable that can prevent the invasion of other strategies.

We now provide a biological explanation for ESS. Consider that mutations occur in a population of strategy s_i , where the frequency of mutate strategy s_j is x . Thus, the expected payoffs for the resident strategy s_i and the mutate strategy s_j are $u(s_i) = (1 - x)u(s_i, s_i) + xu(s_i, s_j)$ and $u(s_j) = (1 - x)u(s_j, s_i) + xu(s_j, s_j)$, respectively. If s_i is an ESS, then we always have $u(s_i) > u(s_j)$ in the limit of $x \rightarrow 0$. This implies that any mutate strategy cannot invade the population under the influence of natural selection.

2.2.3 Quantal Response Equilibrium

The definition of NE relies on perfect rationality, where at a NE all players are using the best response strategies given the strategies of the others. However, people in reality are bounded rational and may make mistakes in decision making. In the context of bounded rationality, one important equilibrium notation is the quantal response equilibrium (QRE) introduced by McKelvey and Palfrey [9]. In a QRE, players do not always choose best responses. Instead, they make decisions based on a probabilistic choice model and assume other players do so as well. A general interpretation of this

model is that players observe randomly perturbed payoffs of strategies and choose optimally according to those noisy observations [9–13]. For a given error structure, QRE is defined as a fixed point of this process.

For instance, in a 2-person n -strategy symmetric game, the set of QRE at noise level λ is defined as

$$\pi_\lambda = \{\mathbf{x} = (x_1, \dots, x_n) | x_i = \sigma(\lambda \mathbf{u}(\mathbf{x}))\}, \quad (6)$$

where $\mathbf{u}(\mathbf{x}) = (u_1(\mathbf{x}), \dots, u_n(\mathbf{x}))$ is the vector for the expected payoffs of the n strategies. The quantal response function has one free parameter λ , whose inverse $1/\lambda$ has been interpreted as the temperature, or the intensity of noise. At $\lambda = 0$, players have no information about the game and each strategy is chosen with equal probability. As λ approaches infinity, players achieve full information about the game and choose best responses. In particular, if σ satisfies continuity and cumulativity, then the limit set $\lim_{\lambda \rightarrow \infty} \pi_\lambda$ consists of NE only [14].

McKelvey and Palfrey [9] then defined an equilibrium selection from the set of Nash equilibria by tracing the branch of the QRE correspondence starting at the centroid of the strategy simplex (the only QRE when $\lambda = 0$) and following this branch to its terminus at $\lambda = +\infty$. If σ satisfies C^2 continuity, monotonicity and cumulativity, then the graph of QRE correspondence generically includes a unique branch that starts at the centroid of the strategy simplex and converges to a unique Nash equilibrium as noises vanish [15, 16]. This equilibrium is called the limiting QRE of the game.

2.3 Evolutionary Dynamics

In addition to static equilibrium analysis, dynamic approaches have been widely used to study the evolution of strategies or behaviors. Dynamic approaches in game theory could be roughly classified into two categories. Evolutionary game theory considers the behavior of large populations, where individuals choose which actions to play genetically or using simple myopic rules (e.g., imitation, best response). In contrast, learning models focus on the behavior of small groups in repeated games. Individuals make decisions according to explicit learning rules, which could be simple myopic rules (called heuristic learning or adaptive learning) or more complicated Bayesian rules (called coordinated Bayesian learning or rational learning). The heuristic learning is close to the spirit of evolutionary approach. In contrast, in the Bayesian learning, individuals play the best response to their beliefs about other individuals' strategies and update the beliefs over rounds. For instance, a representative class of Bayesian learning methods is homotopy approaches, such as the tracing procedure of Harsanyi and Selten [6] or the QRE of McKelvey and Palfrey [9], see also McKelvey and Palfrey [11], Turocy [13], and Zhang [15].

In this paper, we mainly focus on the evolutionary game method. One well-known evolutionary dynamics is the replicator dynamics [7, 17, 18]. Replicator dynamics was first motivated biologically in the context of evolution and is still applied in this field. Later, economists related replicator dynamics to learning [19–21] and defined several equilibrium notions [22, 23].

Replicator dynamics describes the evolution of the frequencies of strategies. It assumes that the per capita growth rate of each strategy is proportional to its payoff and the evolutionary success of a strategy is measured quantitatively by the difference between its payoff and the average payoff of the population. When a strategy has payoff above the average, then this strategy is expected to spread.

Consider a symmetric game and denote the frequency of strategy s_i in the population by x_i . The replicator dynamics are then written as

$$\frac{dx_i}{dt} = x_i(u_i(\mathbf{x}) - \bar{u}(\mathbf{x})), \quad (7)$$

where $u_i(\mathbf{x}) = \sum_{j=1}^n x_j u(s_i, s_j)$ is the expected payoff for strategy i and $\bar{u}(\mathbf{x}) = \sum_{i=1}^n x_i u_i(\mathbf{x})$ is the average payoff of the population.

A state $\mathbf{x}^* = (x_1^*, \dots, x_n^*)$ is called evolutionarily stable if it is an asymptotically stable fixed point of the replicator dynamics. The stability analyses of replicator dynamics were given in the framework of population games [17, 18, 21, 24–26], asymmetric games [7, 27–31], and genetically structured population [7, 32–34]. In a symmetric game, a symmetric NE must be a fixed point of the replicator dynamics, and any stable fixed point must be a NE. In particular, in the case of two strategies, a stable fixed point corresponds to an ESS [17].

Other common evolutionary game dynamics include replicator-mutator dynamics, adaptive dynamics, birth-death/death-birth process, exploration dynamics, Fermi update, best response dynamics, fictitious play, perturbed best response dynamics, logit dynamics, stochastic dynamics, and learning-mutation dynamics. Table 1 summarizes the biological or economic context of these dynamics and related equilibrium notations. More details can be found in Hofbauer and Sigmund [17] and Sandholm [12].

3 Repeated Game

The two basic games introduced in Sect. 2.1, the PD game and PGG, are one-shot, where the only NE/ESS is defection (D). However, most of interactions in our daily life (e.g., with colleagues or friends) are not one-shot. We now consider that the game is repeatedly played between the two players. Furthermore, we assume that the two players do not know how many rounds their game will last. They only know that after every round, a further round can occur with a probability $w < 1$, i.e., the expected number of rounds is $t = 1/(1 - w)$. If the number of rounds is known to the both players, then backward induction predicts that rational players ought to play D in each round.

We note that strategies in the repeated game can be very complicated because the choice in a round could depend on what happens in previous rounds. In fact, in the repeated PD game, there is no strategy that can strictly dominate all other strategies. For instance, if the coplayer uses an unconditional strategy such as always defects (AllD) or always cooperates (AllC), then the best response is to play D. However, if the coplayer uses a conditional strategy that punishes defection, then it is better to

Table 1 Evolutionary dynamics. This table provides biological and economic contexts and related equilibrium notations for different evolutionary dynamics. Details for dynamics 1,2,3,6 can be found in Hofbauer and Sigmund [17]. Details for dynamics 1, 4, 5, 6, 7, 8 can be found in Sandholm [12]

	Evolutionary dynamics	Biological and economic context	Equilibrium notation
1	Replicator dynamics	Infinite population + imitation/selection	NE, ESS
2	Replicator-mutator dynamics	Infinite population + imitation/selection + mutation	NE, ESS
3	Adaptive dynamics	Continuous strategy + imitation/selection + mutation	ESS
4	Birth–death/death-birth process	Finite population + selection + mutation	ESS
5	Exploration dynamics/Fermi update	Finite population + imitation + mutation	ESS
6	Best response dynamics/fictitious play	Best response	NE
7	Perturbed best response/logit dynamics	Best response + decision error	QRE
8	Stochastic/learning mutation dynamics	Finite population + best response + mutation	Long-run equilibrium

play C in every round. In this section, we will introduce some typical strategies in the repeated social dilemma games and discuss their evolutionary stabilities against defection.

3.1 Evolution of Cooperation in the Repeated PD Game

One of the most famous strategy in the repeated PD game is tit-for-tat (TFT). A player using this strategy chooses C in the first round and then does whatever the opponent did in the previous round. TFT was first introduced by Rapoport, and it is both the simplest and the most successful strategy in Axelrod's computer tournaments [2, 35]. The success of TFT may attribute to four reasons. First, it is never the first to defect. Thus, cooperation can be sustained when both players are using TFT. Second, it is provokable into retaliation by a defection of the other. As a result, TFT will be not by exploited by defective strategies. Third, it is forgiving after just one act of retaliation. Thus, if the coplayer switches back to C, then the TFT player resumes to C forthwith. Finally, it never tries to get more than its coplayer. In fact, in the repeated PD game, the only way to do better than the coplayer is to choose D. However, if both players want to get a higher payoff, then they will fall into mutual defection. Subsequent empirical studies found that the success of TFT is not limited to human society but that it also extends to animal populations [36].

However, TFT has an Achilles' heel, it cannot correct mistakes. When players may make mistakes in decision making, the expected payoff for two TFT players in an infinitely repeated PD is same as random players (see Fig. 1a). This implies that TFT is evolutionarily unstable in an uncertain world. In the next few years, evolutionary

Fig. 1 TFT, GTFT, and WSLS in repeated PD games with errors. (a) TFT cannot correct mistakes. (b) GTFT can correct mistakes in few rounds. (c) WSLS can exploit ALLC. (d) WSLS is a deterministic corrector

(a) **TFT** : C D* C D C D D D C* D C ...
TFT : C C D C D D* D D D C D ...

(b) **GTFT** : C D* C D C C D C D C C C ...
GTFT : C C D C C D* C D C C C C C ...

(c) **WSLS** : C C D* D D D D C C C C C ...
ALLC : C C C C C C C C D* C C C C C ...

(d) **WSLS** : C C D* D C C C D C C C C ...
WSLS : C C C D C C D* D C C C C ...

game method has been applied to find the “best” strategy in the repeated PD game. The evolution of strategies has been analyzed by adaptive dynamics and replicatory dynamics in the set of reactive strategies and stochastic strategies [37–43].

A reactive strategy is written as a 3D vector (x, p_C, p_D) , where x is the probability of cooperating in the first round and p_C , and p_D are the conditional probabilities of playing C after the coplayer choosing C and D, respectively. For instance, ALLC, ALLD, and TFT can be denoted by $(1, 1, 1)$, $(0, 0, 0)$, and $(1, 1, 0)$, respectively. Adaptive dynamics show that in the set of reactive strategies, there is no evolutionary tendency towards TFT. Once a cooperative population has been established, TFT will be replaced by a more generous strategy, $(1, 1, p)$. This strategy is called generous tit-for-tat (GTFT). A GTFT player starts with C, choose C when the coplayer has cooperated and chooses C with probability p when the coplayer has defected. Comparing with TFT, GTFT is more forgiving when faced with D. Thus, mistakes can be corrected in few rounds. Specifically, when the error probability is small, the expected payoff for two GTFT players per round is very close to the full reward for mutual cooperation (see Fig. 1b).

A memory-one strategy is written as a 5D vector $(x, p_{CC}, p_{CD}, p_{DC}, p_{DD})$, where x is the probability of cooperating in the first round and p_{CC} , p_{CD} , p_{DC} and p_{DD} are the conditional probabilities of playing C after CC, CD, DC and DD interactions, respectively (see Table 2). Thus, reactive strategies form a 3D subset of the memory-one strategies. Evolutionary simulation based on selection and mutation shows that in the set of memory-one strategies, a GTFT population will be replaced by ALLC through neutral drift, and the ALLC population will be undermined by either ALLD or a new strategy, $(1, 1, 0, 0, 1)$ [41]. This strategy is called Win-Stay Lose-Shift (WSLS). In the repeated donation game, a WSLS player starts with C, repeats the previous action if the payoff > 0 and switches if the payoff ≤ 0 . Comparing with GTFT, WSLS has three advantages. First, it can exploit ALLC (see Fig. 1c). Second, it can against the invasion of ALLD if $b/2 > c$ (in contrast, GTFT cannot defeat ALLD when p is large). Finally, it is a deterministic corrector, where a mistake can be correct in 2 rounds (see Fig. 1d).

We now let natural selection to design a strategy in the set of memory-one strategies. As shown in Fig. 2, a TFT population can be established if the initial cooperation rate is higher, after which TFT will be replaced by more generous strategies such as GTFT and ALLC; finally, the population will be undermined by WSLS or ALLD. However, there are two remaining questions for the evolution of cooperation in repeated games. First, ALLD and TFT are bistable under deterministic evolutionary dynamics (e.g., replicator

ers' payoffs [37, 38, 42, 43, 137]. Furthermore, a subset of ZD strategies allows for implicit forms of extortion, where a player using extortionate strategies can enforce a unilateral claim to an unfair share of rewards such that the partner's best response is to be fully cooperative. Thus, extortioners cannot be outperformed by any opponent in a pairwise encounter. Experimental studies showed that computers using such strategies can defeat each of their human opponents in repeated PD game, but extortion resulted in lower payoffs than generosity [49]. This is because human subjects showed a strong concern for fairness. They often punished extortion by refusing to fully cooperate, thereby reducing their own, and even more so, the extortioner's gains. However, when the repeated game is sufficiently long, human subjects' cooperation rates increased after they realized that punishment cannot change the behavioral pattern of their (computer) partners [50]. As a result, the extortionate strategy earns more than the generous strategy.

We note that strategies such as TFT, GTFT, WSLS, and ZD can be generalized to games with continuous strategies and games with more than 2 players. Some researchers have examined models of continuous PD game (or equivalently, 2-person PGG) based on the linear reactive strategy method. They found that cooperative strategies such as TFT and GTFT are more difficult to invade a non-cooperative equilibrium than in the discrete PD game [51–53]. TFT and WSLS have also been generalized to n -person PD game (or equivalently, PGG with discrete choices) [46, 54–56]. In an n -person PD, a TFT_k ($0 < k < n$) strategist cooperates if at least k individuals cooperated in the previous round [46, 54, 55], and a WSLS strategist cooperates if all them group members cooperated or defected in the previous round [56]. Both TFT_{m-1} and WSLS can sustain cooperation in sizable group, and sometimes large group size can facilitate the evolution of cooperation [46, 56]. Finally, Hilbe et al. [49] showed that multiplayer social dilemma games also include ZD like strategies. They distinguished several subclasses of ZD strategies, e.g., fair strategies ensure that the own payoff matches the average payoff of the group, extortionate strategies allow a player to perform above average, and generous strategies let a player perform below average.

Finally, we summarize recent empirical results for strategies in the multi-player repeated PD games and PGG. On the one hand, a strategy called “moody conditional cooperation” was observed in experiments based on the multiplayer spatial PD game [57–61]. The definition of this strategy comprises two main ingredients. The first is conditional cooperation, i.e. people cooperate more when more of their neighbors cooperated in the previous round. The second is that the probability that they display condition cooperation depends on whether they cooperated or defected in the previous round. If defected, then this player is less likely to cooperate in this round. On the other hand, conditional cooperative strategies have been widely observed in repeated PGG experiments, where a conditional cooperator changes his/her contribution in the next round in the direction of the group average contribution of the current round [62–66]. Experimental studies have revealed that about half of the individuals in the repeated PGG can be classified as conditional cooperator [49, 62, 63]. Recent experiments based on PGG with institutionalized incentives show that this proportion seems to be independent of the incentive modes. The proportion of individuals displaying conditional cooperative behavior is stabilized at around 50% in all nine treatments [66].

3.3 Separable PD Game

For the repeated PD game, one of the key assumptions is that the interaction between a pair of individuals will be repeated for several rounds [2, 4, 67, 68]. In most of real-world interactions, people are able to stop the interaction with their coplayers. Some researchers then considered a variant of the repeated PD game, the separable PD game. In this game, players can choose to keep or stop the interaction after they know the choices of their coplayers. At the end of each round, all single subjects form new interaction pairs through random meeting in the next round. Based on individual self-interest in the PD game, both cooperators and defectors prefer a coplayer who cooperates. Thus, if players are able to unilaterally terminate the interactions with their coplayers, then a simple rule will be followed by all individuals: I would like to keep my opponent if he/she is a cooperator; and if my opponent is a defector, I will stop the interaction with him/her and seek a new partner instead (see Fig. 3). This class of conditional dissociation strategies is called “out-for-tat” (OFT) [56, 69–75, 138]. Since OFT means that an individual displaying cooperation (C) will respond to defection (D) by merely leaving, OFT will not tolerate defection but, unlike TFT, it does not seek revenge.

Theoretical analysis based on evolutionary game method shows that OFT leads to the stable coexistence of C and D. Furthermore, when all individuals use OFT, non-OFT strategies (e.g., ALLD, TFT) cannot successfully invade this population [14, 56, 70, 73, 76]. Empirical studies confirmed that most of subjects in separable PD game

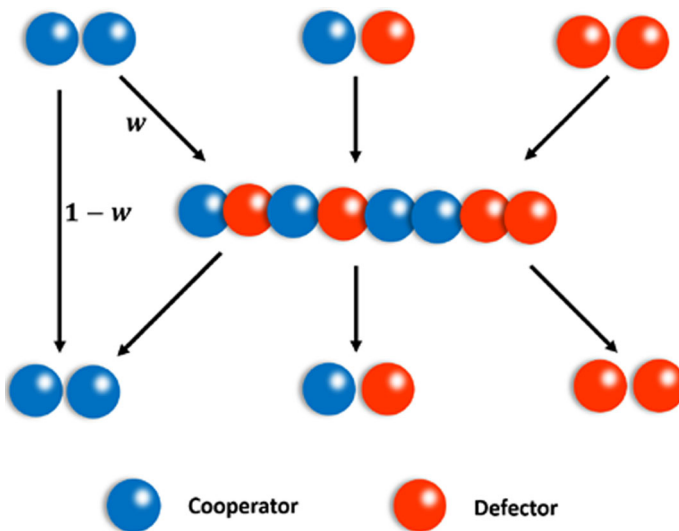


Fig. 3 Separable PD game with OFT players. OFT-cooperators and OFT-defectors are marked by blue and red balls, respectively. At the end of a round, C-D pairs and D-D pairs will be broken since all individuals immediately stop the interaction with a defector. In a C-C pair, both individuals are willing to continue. Thus, the C-C pair will be terminated with probability w (as in the repeated PD game). All single individuals will be paired with a new partner through random meeting in the next round

adopted OFT-like strategies, and the cooperation rate stabilized at about 60% in the repeated play.

In addition to stop the interaction with the current coplayer, another class of studies allows abstaining from a game [71, 74, 77, 78], with players choosing between an outside option (i.e., to be a loner) and the PD game. In such case, cooperators, defectors and loners can coexist if the payoff of the outside option is higher than the payoff of mutual defection. However, replicator dynamics show that voluntary participation usually does not lead to a stable equilibrium, but to an unending limit cycle [77, 78].

4 Reward and Punishment

Although direct reciprocity provides satisfactory explanations of cooperation in social dilemma games with repeated interactions, the sustaining of cooperation in one-shot interaction (or game between strangers) remains a problem [79, 80]. Examples such as the PD game and PGG (see Sect. 2.1) show that self-interested players should prefer a free-rider strategy, whereas it is to everyone's advantage for all players to contribute [4]. In real-life situations that include this sort of dilemma between individual rationality and collective advantage, incentives are often used to promote cooperation, where contributors may be rewarded and free-riders may be punished. In fact, positive and negative incentives exist not only in human societies but also in animal behavior [81]. Understanding the consequences of such 'carrots and sticks' in boosting cooperation is a core topic in the economics and management sciences (see [82–84] for three recent review papers).

4.1 Peer Incentive

Several types of incentives have been proposed to promote cooperation in social dilemma games. Most of these investigations addressed so-called peer (or decentralized) incentives. In this scenario, players are allowed to reward and/or punish others at a cost to themselves. A typical social dilemma game with peer incentives is a two-stage game, where in the first stage players play the social dilemma game and in the second stage they can reward and/or punish their group members based on the result of the first stage (see Fig. 4a).

The social dilemma game with peer incentives can be simplified as a two-stage asymmetric game [4, 85, 86]. To show this, we consider a two-player donation game with peer incentives, where one player is donor and the other is recipient. In the first stage, the donor has two possible strategies, C and D, and in the second stage, the recipient has three strategies, no action (No), R (reward), and P (punishment). We assume that players are pro-social in the sense that they only reward cooperators and punish defectors. Denote the cost of incentive by C_I and the amounts of reward and

(a) Peer incentive



(b) Institutional incentive

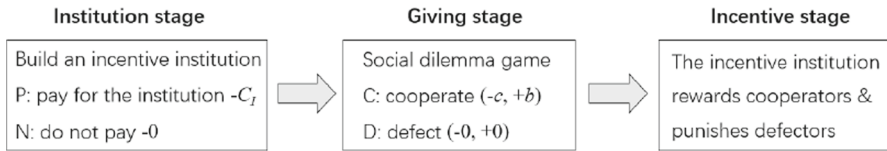


Fig. 4 Peer incentives and institutional incentives. (a) A social dilemma game with peer incentives can be formalized as a two-stage game. (b) A social dilemma game with institutional incentives can be formalized as a three-stage game

punishment by R and P, respectively. The payoff matrix of this two-stage game can be written as

$$\begin{matrix} & \text{No} & \text{R} & \text{P} \\ \text{C} & (-c, b & -c + R, b - C_I & -c, b \\ \text{D} & (0, 0 & 0, 0 & -P, -C_I) \end{matrix}, \tag{8}$$

where the rows are the strategies of the donor and columns are the strategies of the recipient.

It is clear that the defective state (D, No) is a NE, and the cooperative state (C, P) becomes a NE if $P > c$. The subgame perfect NE of this game can be derived by backward induction. Since the incentive is costly, rational players will not choose R and P in the second stage. Thus, the subgame perfect NE in this game is same as the game without peer incentives, i.e., (D, No). Evolutionary game theory analysis based on replicator dynamics also shows that the cooperative state (C, P) is evolutionarily unstable in one-shot game [85–94]. However, when subjects may make mistakes in decision making, intermediate errors can increase cooperation in for games with peer punishment [16, 95]. Furthermore, if the game is played repeatedly, incentive may help to enhance the effects of direct reciprocity and indirect reciprocity [4, 85, 86]. In particular, punishment promotes cooperation better than reward in the case of indirect reciprocity (i.e., players may know the historical behaviors of their coplayers), where the cooperative state can be evolutionarily stable.

Laboratory experiments confirmed that peer punishment can curb free-riding in human populations [96–112]. However, peer punishment suffers from several drawbacks. First, the use of incentives is costly, which then raises an issue of second-order free-riding. In fact, the incentive system itself is a common good that can be exploited, and the use of punishment is individually disadvantageous [99, 112]. Second, punishment decreases the total welfare. As a result, the average payoff for group with punishment often less than group without punishment [103, 104]. Finally, punish-

ment often some players abuse sanctioning opportunities by engaging in antisocial punishment, which harms cooperators [97–99, 106, 107].

In contrast, the effect of peer reward is controversial. Some experimental studies showed that reward can promote cooperation if the reward-to-cost ratio (i.e., R/C_I) is greater than one. However, given the budget-balanced condition (i.e., $R/C_I = 1$), peer reward alone is relatively inefficient in promote cooperation [110, 113].

4.2 Institutional Incentive

In the institutional incentive scenario, it is not individuals who reward or punish. Rather, an institution rewards and punishes individuals based on their contributions. The use of institutional incentives is a common feature in many parts of human society such as government institutions and businesses. It can overcome the problem of second-order free-riding and avoid antisocial punishment. However, this approach is more wasteful than peer incentives because subjects have to pay a fee to maintain the institution even if no one is being rewarded or punished (which can be viewed as paying for the upkeep of a police force).

A social dilemma game with institutional incentives can be formalized as a three-stage game. In the first stage, players can decide to pay for the incentive institution. In the second stage, they play the social dilemma game. In the third stage, the institution will reward or punish each player based on his/her behaviors in the two-stage game, where the amount of incentive depends on the total payment in the first stage. In particular, players do not need to make decisions in the third stage.

We consider an n -person PGG with institutional incentives [114–122]. For simplicity, we first assume that the institutional incentive is compulsory, where each player is forced to pay C_I in the first stage. Thus, the total amount of incentive is $C_I n$. In the second stage, they play PGG with discrete choices. Specifically, each player can decide whether to contribute a fixed amount c_0 knowing that this amount will be multiplied by $r > 1$ and divided equally among all n players in the group. If n_C is the number of those players who contribute (i.e. cooperators) and n_D the number of those who do not (i.e. defectors), then the payoffs of a cooperator and a defector are $(rcn_C)/n - c$ and rcn_C/n , respectively.

In the case of institutional reward (IR), the incentive is shared among the n_C cooperators. Thus, each cooperator obtains a reward of $C_I n/n_C$. In the case of institutional punishment (IP), each defector analogously receives a punishment of $C_I n/n_D$. Payoff for cooperators and defectors in IR are

$$\begin{aligned} u_C &= \frac{rcn_C}{n} - c - C_I + \frac{C_I n}{n_C}, \\ u_D &= \frac{rcn_C}{n} - C_I, \end{aligned} \quad (9)$$

and in IP are

$$u_C = \frac{rcn_C}{n} - c - C_I,$$

$$u_D = \frac{rcn_C}{n} - C_I - \frac{C_I n}{n_D}. \quad (10)$$

Theoretical studies based on evolutionary game theory have revealed that the effect of institutional incentives on cooperation can be understood in terms of the incentive size [114, 115, 118, 119, 123, 124]. If $C_I < \frac{c}{n} - \frac{rc}{n^2}$, then both reward and punishment have no effect on promoting cooperation, and selfish players maintain a free-riding strategy. If $C_I > c - \frac{rc}{n}$, then both reward and punishment compel all players to cooperate. If $\frac{c}{n} - \frac{rc}{n^2} < C_I < c - \frac{rc}{n}$, then full contribution and free-riding are bistable under punishment, and reward can cause only the stable coexistence of free-riders and cooperators. Experimental results are consistent with the theoretical prediction, where an intermediate size of punishment is sufficient to support high cooperation level, but reward can maintain stable cooperation only for a large incentive size [66, 113, 125–129].

4.2.1 Compulsory Institution vs Voluntary Institution

In the above model, every player has to pay for the incentive institution before contributing to the PGG. In fact, this payment can be seen as an entry fee for the PGG. If the payment is voluntary (e.g., players may choose not to pay for the incentive institution in the first stage), then the issue of second-order free-riding is raised because the incentive institution itself is a common good that can be exploited. Previous studies show that IP with voluntary payment is functional if punishment is also imposed on second-order free-riders, i.e. subjects who cooperate but do not pay for the institution should also be punished [121, 127, 129]. By contrast, if IR is budget-balanced, then rational players will not pay for the institution [94, 123].

Since the entry fee in our model is compulsory, a subsequent question is whether people will choose such an incentive institution. One experimental study showed that comparing with standard PGG or PGG with peer punishment, people are more likely to choose institutional punishment if the institution punishes both first-order and second-order free-riders [129]. Another experimental study indicated that approximately half of the subjects would like to pay 20% of their wealth for a reward mechanism before playing the PGG [130]. These findings suggest that the incentive institution could be sustainable.

4.2.2 Absolute Incentive vs Relative Incentive

In PD game or PGG with discrete choices, a reasonable incentive institution should reward cooperators and punish defectors. However, in PGG with continuous choices, it is difficult to determine who should be rewarded or punished.

Incentive institutions considered in the previous studies could be roughly classified into two types. One type is characterized as an absolute incentive institution, where the institution punishes (or rewards) all individuals whose contribution is less (or higher) than a predefined threshold [114, 118, 121, 123, 127, 129]. A similar institution is that the reward (or punishment) amount increases (decreases) with the absolute contribution, see e.g., Galbiati and Vertova [140] and Putterman et al. [126]. Both

theoretical and experimental studies indicated that absolute punishment can eliminate extremely selfish behaviors in a cooperative population [121, 127, 129]. In contrast, absolute reward is relatively ineffective in moving the equilibrium from the selfish one to the cooperative one [114, 118].

Another type is characterized as a relative incentive institution, where individuals who contribute an amount lower than the group average are more likely to be punished, and those who contribute higher than the group average are more likely to be rewarded [66, 115, 125, 128, 131–134]. A similar institution would be that the reward (or punishment) amount increases (decreases) with the relative contribution, see e.g., Falkinger et al. [135]. For relative punishment, a full contribution becomes a Nash equilibrium if the institution punishes the lowest contributor such that his or her payoff is slightly lower than that of the second lowest contributor [131]. In contrast, relative reward can promote cooperation only if lower contributors also have the chance to win the reward [136].

5 Summary

In this paper, we provide a review of sustaining cooperation in groups of unrelated individuals from the perspective of game theory. We mainly focus on two mechanisms, direct reciprocity and incentives.

Direct reciprocity relies on the repeated interaction. In the framework of direct reciprocity, a vast amount of strategies can help to promote cooperation. In the repeated PD game, strategies that can sustain full cooperation including GRIM, TFT, and some strategies in the set of ZD strategies. However, when subjects may make mistakes in decision making, the above strategies can no longer sustain cooperation because they cannot correct mistakes. Evolutionary game analysis shows that successful strategies in repeated PD with errors are GTFT and WSLs. Finally, when subjects are able to stop the interaction with their coplayers, OTF performs better than TFT, where an OFT player responds to defection by breaking the interaction rather than seeking revenge. However, OFT cannot sustain full cooperation, it leads to the stable coexistence of C and D. This implies that promoting cooperation in separable PD is a more difficult issue.

Comparing to direct reciprocity, incentive mechanism can promote cooperation even in one-shot interaction. Incentive mechanisms can be classified to two classes. One class is peer (or decentralized) incentives, where players can impose fines or bonuses on others at a cost to themselves. Another class is institutional (or centralized) incentives. In this scenario, it is not individuals who reward or punish but rather an institution that rewards and punishes individuals based on their contributions. Thus, Institutional incentives can overcome the problem of second-order free-riding, whereas peer incentives cannot.

Finally, when devising incentive systems, it is important to recognize that the effect of reward is not equivalent to punishment. Punishment can eliminate extremely selfish behaviors in a cooperative population but it does not work for the selfish population. In contrast, rewards alone are relatively ineffective in moving the equilibrium from the selfish one to the cooperative one. Finally, combining rewards and punishments

has a very strong effect. Rewards can act as a catalyzer if the population consists of a majority of defectors.

References

- [1] Pennisi, E.: How did cooperative behavior evolve? *Science* **309**(5731), 93–93 (2005)
- [2] Axelrod, R.: *The Evolution of Cooperation*. Basic Books, New York (1984)
- [3] Nowak, M.A.: *Evolutionary Dynamics*. Belknap Press, Cambridge (2006)
- [4] Sigmund, K.: *The Calculus of Selfishness*. Princeton University Press, Princeton (2010)
- [5] Nash, J.: Equilibrium points in n -person games. *Proc. Natl. Acad. Sci. USA* **36**, 48–49 (1950)
- [6] Harsanyi, J.C., Selten, R.: *A general theory of equilibrium selection in games*. MIT Press, Cambridge (1988)
- [7] Maynard Smith, J.: *Evolution and the Theory of Games*. Cambridge University Press, Cambridge (1982)
- [8] Maynard Smith, J.M., Price, G.R.: The logic of animal conflict. *Nature* **246**(5427), 15–18 (1973)
- [9] McKelvey, R.D., Palfrey, T.A.: Quantal response equilibria for normal form games. *Games Econ. Behav.* **10**, 6–38 (1995)
- [10] Goeree, J.K., Holt, C.A., Palfrey, T.A.: Regular quantal response equilibrium. *Exp. Econ.* **8**, 347–367 (2005)
- [11] McKelvey, R.D., Palfrey, T.A.: Quantal response equilibria for extensive form games. *Exp. Econ.* **1**, 9–41 (1998)
- [12] Sandholm, W.H.: *Population Games and Evolutionary Dynamics*. MIT Press, Cambridge (2010)
- [13] Turocy, T.L.: A dynamic homotopy interpretation of the logistic quantal response equilibrium correspondence. *Games Econ. Behav.* **51**, 243–263 (2005)
- [14] Zhang, B., Fan, S., Li, C., Zheng, X., Bao, J., Cressman, R., Tao, Y.: Opting out against defection leads to stable coexistence with cooperation. *Sci. Rep.* **6**, 35902 (2016). <https://doi.org/10.1038/srep35902>
- [15] Zhang, B.: Quantal response methods for equilibrium selection in normal form game. *J. Math. Econ.* **64**, 113–123 (2016)
- [16] Zhang, B., Hofbauer, J.: Quantal response methods for equilibrium selection in 2×2 coordination games. *Games Econ. Behav.* **97**, 19–31 (2016)
- [17] Hofbauer, J., Sigmund, K.: *Evolutionary games and population dynamics*. Cambridge University Press, Cambridge (1998)
- [18] Taylor, P., Jonker, L.: Evolutionary stable strategies and game dynamics. *Math. Biosci.* **50**, 145–156 (1978)
- [19] Börgers, T., Sarin, R.: Learning through reinforcement and replicator dynamics[J]. *J. Econ. Theory* **77**(1), 1–14 (1997)
- [20] Schlag, K.H.: Why imitate, and if so, how? A boundedly rational approach to multi-armed bandits. *J. Econ. Theory* **78**(1), 130–156 (1998)
- [21] Weibull, J.W.: *Evolutionary Game Theory*. MIT press, Cambridge (1995)
- [22] Samuelson, L.: *Evolutionary Games and Equilibrium Selection*. MIT press, Boston (1997)
- [23] Binmore, K., Samuelson, L.: Muddling through: noisy equilibrium selection. *J. Econ. Theory* **74**(2), 235–265 (1997)
- [24] Cressman, R.: *The Stability Concept of Evolutionary Game Theory: A Dynamic Approach*. Springer, Berlin (1992)
- [25] Hofbauer, J., Schuster, P., Sigmund, K.: A note on evolutionary stable strategies and game dynamics. *J. Theor. Biol.* **81**(3), 609–612 (1979)
- [26] Zeeman, E.C.: *Population Dynamics from Game Theory/Global Theory of Dynamical Systems*, pp. 471–497. Springer, Berlin (1980)
- [27] Cressman, R., Ansell, C., Binmore, K.: *Evolutionary Dynamics and Extensive form Games*, vol. 5. MIT Press, Cambridge (2003)
- [28] Maynard Smith, J., Hofbauer, J.: The “battle of the sexes”: a genetic model with limit cycle behavior. *Theor. Popul. Biol.* **32**(1), 1–14 (1987)
- [29] Schuster, P., Sigmund, K.: Coyne, philandering and stable strategies. *Anim. Behav.* **29**(1), 186–192 (1981)

- [30] Selten, R.: A note on evolutionarily stable strategies in asymmetric animal conflicts. *J. Theor. Biol.* **84**, 93–101 (1980)
- [31] Zhang, B., Hofbauer, J.: Equilibrium selection via replicator dynamics. *Int. J. Game Theory* **44**, 433–448 (2015)
- [32] Eshel, I.: Evolutionarily stable strategies and viability selection in Mendelian populations. *Theor. Popul. Biol.* **22**(2), 204–217 (1982)
- [33] Lessard, S.: Evolutionary dynamics in frequency-dependent two-phenotype models. *Theor. Popul. Biol.* **25**(2), 210–234 (1984)
- [34] Maynard Smith, J.: Will a sexual population evolve to an ESS? [*J.*] *Am. Nat.* **117**(6), 1015–1018 (1981)
- [35] Axelrod, R., Hamilton, W.D.: The evolution of cooperation. *Science* **211**, 1390–1396 (1981)
- [36] Milinski, M.: Tit for tat in sticklebacks and the evolution of cooperation. *Nature* **325**, 433–435 (1987)
- [37] Adami, C., Hintze, A.: Evolutionary instability of zero-determinant strategies demonstrates that winning is not everything. *Nat. Commun.* **4**, 2193 (2013)
- [38] Hilbe, C., Nowak, M.A., Sigmund, K.: The evolution of extortion in iterated prisoner's dilemma games. *Proc. Natl. Acad. Sci. USA* **110**, 6913–6918 (2013)
- [39] Nowak, M.A., Sigmund, K.: The evolution of stochastic strategies in the Prisoner's Dilemma. *Acta Appl. Math.* **20**, 247–265 (1990)
- [40] Nowak, M.A., Sigmund, K.: Tit for tat in heterogeneous populations. *Nature* **355**, 250–253 (1992)
- [41] Nowak, M.A., Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the Prisoner's Dilemma game. *Nature* **364**, 56–58 (1993)
- [42] Press, W.H., Dyson, F.D.: Iterated prisoner's dilemma contains strategies that dominate any evolutionary opponent. *Proc. Natl. Acad. Sci. USA* **109**, 10409–10413 (2012)
- [43] Stewart, A.J., Plotkin, J.B.: From extortion to generosity, evolution in the iterated prisoner's dilemma. *Proc. Natl. Acad. Sci. USA* **110**, 15348–15353 (2013)
- [44] Fudenberg, D., Harris, C.: Evolutionary dynamics with aggregate shocks. *J. Econ. Theor.* **57**, 420–441 (1992)
- [45] Imhof, L.A., Fudenberg, D., Nowak, M.A.: Evolutionary cycles of cooperation and defection. *Proc. Natl. Acad. Sci. USA* **102**, 10797–10800 (2005)
- [46] Kurokawa, S., Ihara, Y.: Emergence of cooperation in public goods games. *Proc. R. Soc. B.* **276**, 1379–1384 (2009)
- [47] Nowak, M.A., Sasaki, A., Taylor, C., Fudenberg, D.: Emergence of cooperation and evolutionary stability in finite populations. *Nature* **428**, 646–650 (2004)
- [48] Dong, Y., Li, C., Tao, Y., Zhang, B.: Evolution of conformity in social dilemmas. *PLoS ONE* **10**, e0137435 (2015). <https://doi.org/10.1371/journal.pone.0137435>
- [49] Hilbe, C., Röhl, T., Milinski, M.: Extortion subdues human players but is finally punished in the prisoner's dilemma. *Nat. Commun.* **5**, 3976 (2014)
- [50] Wang, Z., Zhou, Y., Lien, J.W., Zheng, J., Xu, B.: Extortion can outperform generosity in the iterated prisoners' dilemma. *Nat. Commun.* **7**, 11125 (2016)
- [51] Le, S., Boyd, R.: Evolutionary dynamics of the continuous iterated Prisoner's Dilemma. *J. Theor. Biol.* **245**, 258–267 (2007)
- [52] Wahl, L., Nowak, M.A.: The continuous Prisoner's Dilemma: I. linear reactive strategies. *J. Theor. Biol.* **200**, 307–321 (1999)
- [53] Wahl, L., Nowak, M.A.: The continuous Prisoner's Dilemma: II. Linear reactive strategies with noise. *J. Theor. Biol.* **200**, 323–338 (1999)
- [54] Boyd, R., Richerson, P.J.: The evolution of reciprocity in sizable groups. *J. Theor. Biol.* **132**, 337–356 (1988)
- [55] Hauert, C., Schuster, H.: Effects of increasing the number of players and memory size in the iterated Prisoner's Dilemma: a numerical approach. *Proc. R. Soc. B.* **264**, 513–519 (1997)
- [56] Pinheiro, F.L., Vasconcelos, V.V., Santos, F.C., Pacheco, J.M.: Evolution of all-or-none strategies in repeated public goods dilemmas. *PLoS Comput. Biol.* **10**, e1003945 (2014)
- [57] Gracia-Lázaro, C., Ferrer, A., Ruiz, G., Tarancón, A., Cuesta, J.A., Sánchez, A., et al.: Heterogeneous networks do not promote cooperation when humans play a Prisoner's Dilemma. *Proc. Natl. Acad. Sci. USA* **109**, 12922–12926 (2012)
- [58] Grujić, J., Fosco, C., Araujo, L., Cuesta, J.A., Sánchez, A.: Social experiments in the mesoscale: humans playing a spatial Prisoner's Dilemma. *PLoS ONE* **5**, e13749 (2010). <https://doi.org/10.1371/journal.pone.0013749>

- [59] Grujić, J., Gracia-Lázaro, C., Milinski, M., Semmann, D., Traulsen, A., Cuesta, J.A., et al.: A comparative analysis of spatial Prisoners Dilemma experiments: conditional cooperation and payoff irrelevance. *Sci. Rep.* **4**, 4615 (2014)
- [60] Grujić, J., Röhl, T., Semmann, D., Milinski, M., Traulsen, A.: Consistent strategy updating in spatial and nonspatial behavioral experiments does not promote cooperation in social networks. *PLoS ONE* **7**, e47718 (2012). <https://doi.org/10.1371/journal.pone.0047718>
- [61] Traulsen, A., Semmann, D., Sommerfeld, R.D., Krambeck, H.J., Milinski, M.: Human strategy updating in evolutionary games. *Proc. Natl. Acad. Sci. USA* **107**, 2962–2966 (2010)
- [62] Fischbacher, U., Gächter, S.: Social preferences, beliefs, and the dynamics of free riding in public goods experiments. *Am. Econ. Rev.* **100**, 541–556 (2010)
- [63] Fischbacher, U., Gächter, S., Fehr, E.: Are people conditionally cooperative? Evidence from a public goods experiment. *Econ. Lett.* **71**, 397–404 (2001)
- [64] Keser, C., van Winden, F.: Conditional cooperation and voluntary contributions to public goods. *Scand. J. Econ.* **102**, 23–39 (2000)
- [65] Kurzban, R., Houser, D.: Experiments investigating cooperative type in humans: a complement to evolutionary theory and simulations. *Proc. Natl. Acad. Sci. USA* **102**, 1803–1807 (2005)
- [66] Wu, J.J., Li, C., Zhang, B., Cressman, R., Tao, Y.: The role of institutional incentives and the exemplar in promoting cooperation. *Sci. Rep.* **4**, 6421 (2014)
- [67] Axelrod, R., Dion, D.: The further evolution of cooperation. *Science* **242**, 1385–1390 (1988)
- [68] Nowak, M.A.: Five rules for the evolution of cooperation. *Science* **314**(5805), 1560–1563 (2006)
- [69] Aktipis, C.A.: Know when to walk away: contingent movement and the evolution of cooperation. *J. Theor. Biol.* **231**, 249–260 (2004)
- [70] Fujiwara-Greve, T., Okuno-Fujiwara, M.: Voluntarily separable repeated prisoner's dilemma. *Rev. Econ. Stud.* **76**, 993–1021 (2009)
- [71] Hauk, E.: Multiple prisoner's dilemma games with (out) an outside option: an experimental study. *Theory Decis.* **54**(3), 207–229 (2003)
- [72] Hayashi, N.: From TIT-for-TAT to OUT-for-TAT. *Soc. Theory Methods* **8**, 19–32 (1993)
- [73] Izquierdo, S.S., Izquierdo, L.R., Vega-Redondo, F.: The option to leave: conditional dissociation in the evolution of cooperation. *J. Theor. Biol.* **267**, 76–84 (2010)
- [74] Orbell, J.M., Dawes, R.M.: Social welfare, cooperators' advantage, and the option of not playing the game. *Am. Sociol. Rev.* **58**, 787–800 (1993)
- [75] Schuessler, R.: Exit threats and cooperation under anonymity. *J. Conflict Resolut.* **33**, 728–749 (1989)
- [76] Zheng, X.D., Li, C., Yu, J.R., Wang, S.C., Fan, S.J., Zhang, B.Y., Tao, Y.: A simple rule of direct reciprocity leads to the stable coexistence of cooperation and defection in the prisoner's dilemma game. *J. Theor. Biol.* **420**, 12–17 (2017)
- [77] Hauert, C., De Monte, S., Hofauer, J., Sigmund, K.: Volunteering as red queen mechanism for cooperation in public goods games. *Science* **296**, 1129–1132 (2002)
- [78] Semmann, D., Krambeck, H.J., Milinski, M.: Volunteering leads to rock–paper–scissors dynamics in a public goods game. *Nature* **425**, 390–393 (2003)
- [79] Hamilton, W.D.: The genetical evolution of social behaviour I. *J. Theor. Biol.* **7**, 1–52 (1964)
- [80] Trivers, R.L.: The evolution of reciprocal altruism. *Q. Rev. Biol.* **46**, 35–57 (1971)
- [81] Clutton-Brock, T.H., Parker, G.A.: Punishment in animal societies. *Nature* **373**, 209–216 (1995)
- [82] Balliet, D., Mulder, L.B., VanLange, P.A.M.: Reward, punishment, and cooperation: a meta-analysis. *Psychol. Bull.* **137**, 594–615 (2011)
- [83] Guala, F.: Reciprocity: weak or strong? What punishment experiments do (and do not) demonstrate. *Behav. Brain Sci.* **35**, 1–15 (2012)
- [84] Perc, M., Jordan, J.J., Rand, D.G., Wang, Z., Boccaletti, S., Szolnoki, A.: Statistical physics of human cooperation. *Phys. Rep.* **687**, 1–51 (2017)
- [85] Hilbe, C., Sigmund, K.: Incentives and opportunism: from the carrot to the stick. *Proc. R. Soc. B* **277**, 2427–2433 (2010)
- [86] Sigmund, K., Hauert, C., Nowak, M.A.: Reward and punishment. *Proc. Natl. Acad. Sci. USA* **98**(10), 757–10762 (2001)
- [87] Boyd, R., Gintis, H., Bowles, S., Richerson, P.J.: The evolution of altruistic punishment. *Proc. Natl. Acad. Sci. USA* **100**, 3531–3535 (2003)
- [88] dos Santos, M.: The evolution of anti-social rewarding and its countermeasures in public goods games. *Proc. R. Soc. B* **282** (2015). <https://doi.org/10.1098/rspb.2014.1994>

- [89] Fowler, J.H.: Altruistic punishment and the origin of cooperation. *Proc. Natl Acad. Sci. USA* **102**, 7047–7049 (2005)
- [90] Gao, L., Wang, Z., Pansini, R., Li, Y.T., Wang, R.W.: Collective punishment is more effective than collective reward for promoting cooperation. *Sci. Rep.* **5**, 17752 (2015)
- [91] Hauert, C., Traulsen, A., Brandt, H., Nowak, M.A., Sigmund, K.: Via freedom to coercion: the emergence of costly punishment. *Sci.* **316**, 1905–1907 (2007)
- [92] Hilbe, C., Traulsen, A.: Emergence of responsible sanctions without second order free riders, anti-social punishment or spite. *Sci. Rep.* **2**, 458 (2012)
- [93] Okada, I., Yamamoto, H., Toriumi, F., Sasaki, T.: The effect of incentives and meta-incentives on the evolution of cooperation. *PLoS Comput. Biol.* **11**, e1004232 (2015). <https://doi.org/10.1371/journal.pcbi.1004232>
- [94] Szolnoki, A., Perc, M.: Antisocial pool rewarding does not deter public cooperation. *Proc. R. Soc. B.* **282** (2015). <https://doi.org/10.1098/rspb.2015.1975>
- [95] Traulsen, A., Hauert, C., De Silva, H., Nowak, M.A., Sigmund, K.: Exploration dynamics in evolutionary games. *Proc. Natl Acad. Sci. USA* **106**, 709–712 (2009)
- [96] Andreoni, J., Harbaugh, W.T., Vesterlund, L.: The carrot or the stick: rewards, punishments and cooperation. *Am. Econ. Rev.* **93**, 893–902 (2003)
- [97] Cinyabuguma, M., Page, T., Putterman, L.: Can second-order punishment deter perverse punishment. *Exp. Econ.* **9**, 265–279 (2006)
- [98] Denant-Boemont, L., Masclet, D., Noussair, C.: Punishment, counterpunishment and sanction enforcement in a social dilemma experiment. *Econ. Theory* **33**, 145–167 (2007)
- [99] Dreber, A., Rand, D.G., Fudenberg, D., Nowak, M.A.: Winners don't punish. *Nature* **452**, 348–351 (2008)
- [100] Egas, M., Riedl, A.: The economics of altruistic punishment and the maintenance of cooperation. *Proc. R. Soc. B* **275**, 871–878 (2008)
- [101] Fehr, E., Fischbacher, U.: The nature of human altruism. *Nature* **425**, 785–791 (2003)
- [102] Fehr, E., Gächter, S.: Cooperation and punishment in public goods experiments. *Am. Econ. Rev.* **90**, 980–994 (2000)
- [103] Fehr, E., Rockenbach, B.: Detrimental effects of sanctions on human altruism. *Nature* **422**, 137–140 (2003)
- [104] Gächter, S., Renner, E., Sefton, M.: The long-run benefits of punishment. *Science* **322**, 1510–1510 (2008)
- [105] Gurek, O., Irlenbusch, B., Rockenbach, B.: The competitive advantage of sanctioning institutions. *Sci.* **312**, 108–111 (2006)
- [106] Herrmann, B., Thoni, C., Gächter, S.: Antisocial punishment across societies. *Science* **319**, 1362–1367 (2008)
- [107] Nikiforakis, N.: Punishment and counter punishment in public good games: can we really govern ourselves? *J. Public. Econ.* **92**, 91–112 (2008)
- [108] Rand, D.G., Dreber, A., Ellingsen, T., Fudenberg, D., Nowak, M.A.: Positive interactions promote public cooperation. *Science* **325**, 1272–1275 (2009)
- [109] Rand, D.G., Nowak, M.A.: The evolution of antisocial punishment in optional public goods games. *Nat. Commun.* **2**, 434 (2011)
- [110] Sefton, M., Schupp, R., Walker, M.: The effect of rewards and sanctions in provision of public goods. *Econ. Inq.* **45**, 671–690 (2007)
- [111] Walker, J.M., Halloran, M.: Rewards and sanctions and the provision of public goods in one-shot settings. *Exp. Econ.* **7**, 235–247 (2004)
- [112] Wu, J.J., Zhang, B., Zhou, Z.X., He, Q.Q., Zheng, X.D., Cressman, R., Tao, Y.: Costly punishment does not always increase cooperation. *Proc. Natl Acad. Sci. USA* **106**, 17448–17451 (2009)
- [113] Sutter, M., Haigner, S., Kocher, M.G.: Choosing the carrot or the stick? Endogenous institutional choice in social dilemma situations. *Rev. Econ. Stud.* **77**, 1540–1566 (2010)
- [114] Chen, X., Sasaki, T., Brännström, A., Dieckmann, U.: First carrot, then stick: how the adaptive hybridization of incentives promotes cooperation. *J. R. Soc. Interface* **12** (2015). <https://doi.org/10.1098/rsif.2014.0935>
- [115] Cressman, R., Song, J.W., Zhang, B., Tao, Y.: Cooperation and evolutionary dynamics in the public goods game with institutional incentives. *J. Theor. Biol.* **299**, 144–151 (2012)
- [116] Oya, G., Ohstuki, H.: Stable polymorphism of cooperators and punishers in a public goods game. *J. Theor. Biol.* **419**, 243–253 (2017)

- [117] Perc, M.: Sustainable institutionalized punishment requires elimination of second-order free-riders. *Sci. Rep.* **2**, 344 (2012)
- [118] Sasaki, T., Brännström, A., Dieckmann, U., Sigmund, K.: The take-it-or-leave-it option allows small penalties to overcome social dilemmas. *Proc. Natl Acad. Sci. USA* **109**, 1165–1169 (2012)
- [119] Sasaki, T., Uchida, S., Chen, X.: Voluntary rewards mediate the evolution of pool punishment for maintaining public goods in large populations. *Sci. Rep.* **5**, 8917 (2015)
- [120] Schoenmakers, S., Hilbe, C., Blasius, B., Traulsen, A.: Sanctions as honest signals—the evolution of pool punishment by public sanctioning institutions. *J. Theor. Biol.* **356**, 36–46 (2014)
- [121] Sigmund, K., De-Silva, H., Traulsen, A., Hauert, C.: Social learning promotes institutions for governing the commons. *Nature* **466**, 861–863 (2010)
- [122] Szolnoki, A., Perc, M.: Second-order free-riding on antisocial punishment restores the effectiveness of prosocial punishment. *Phys. Rev. X* **7**, 041027 (2017). <https://doi.org/10.1103/PhysRevX.7.041027>
- [123] Dong, Y., Sasaki, T., Zhang, B.: The competitive advantage of institutional reward. *Proc. R. Soc. B* **286**, 20190001 (2019). <https://doi.org/10.1098/rspb.2019.0001>
- [124] Zhang, B., An, X., Dong, Y.: Conditional cooperator enhances institutional punishment in public goods game. *Appl. Math. Comput.* **390**, 125600 (2020). <https://doi.org/10.1016/j.amc.2020.125600>
- [125] Cressman, R., Wu, J.J., Li, C., Tao, Y.: Game experiments on cooperation through reward and punishment. *Biol. Theory* **8**, 158–166 (2013)
- [126] Putterman, L., Tyran, J.L., Kamei, K.: Public goods and voting on formal sanction schemes. *J. Public Econ.* **95**, 1213–1222 (2011)
- [127] Traulsen, A., Röhl, T., Milinski, M.: An economic experiment reveals that humans prefer pool punishment to maintain the commons. *Proc. R. Soc. B* **279**, 3716–3721 (2012)
- [128] Yamagishi, T.: The provision of a sanctioning system as a public good. *J. Pers. Soc. Psychol.* **51**, 110–116 (1986)
- [129] Zhang, B., Li, C., De Silva, H., Bednarik, P., Sigmund, K.: The evolution of sanctioning institutions: an experimental approach to the social contract. *Exp. Econ.* **17**, 285–303 (2014)
- [130] Yang, C.L., Zhang, B., Charness, G., Li, C., Lien, J.: Endogenous rewards promote cooperation. *Proc. Natl. Acad. Sci. USA* **115**, 9968–9973 (2018)
- [131] Andreoni, J., Gee, L.L.: Gun for hire: delegated enforcement and peer punishment in public goods provision. *J. Public Econ.* **96**, 1036–1046 (2012)
- [132] Dong, Y., Zhang, B., Tao, Y.: The dynamics of human behavior in the public goods game with institutional incentives. *Sci. Rep.* **6**(1), 1–7 (2016)
- [133] Kamijo, Y., Nihonsugi, T., Takeuchi, A., Funaki, Y.: Sustaining cooperation in social dilemmas: comparison of centralized punishment institutions. *Games Econ. Behav.* **84**, 180–195 (2014)
- [134] Qin, X., Wang, S.: Using an exogenous mechanism to examine efficient probabilistic punishment. *J. Econ. Psychol.* **39**, 1–10 (2013)
- [135] Falkinger, J., Fehr, E., Gächter, S., Winter-Ember, R.: A simple mechanism for the efficient provision of public goods: experimental evidence. *Am. Econ. Rev.* **90**(1), 247–264 (2000)
- [136] Huang, S., Dong, Y., Zhang, B.: The optimal incentive in promoting cooperation: punish the worst and do not only reward the best. *Singap. Econ. Rev.* (2020). <https://doi.org/10.1142/S0217590820500071>
- [137] Hilbe, C., Nowak, M.A., Traulsen, A.: Adaptive dynamics of extortion and compliance. *PLoS ONE* **8**, e77886 (2013). <https://doi.org/10.1371/journal.pone.0077886>
- [138] Izquierdo, S.S., Izquierdo, L.R., Vega-Redondo, F.: Leave and let leave: a sufficient condition to explain the evolutionary emergence of cooperation. *J. Econ. Dyn. Control* **46**, 91–113 (2014)
- [139] Zhang, B., Li, C., Tao, Y.: Evolutionary stability and the evolution of cooperation in heterogeneous graphs. *Dyn. Games Appl.* **6**, 567–579 (2016)
- [140] Galbiati, R., Vertova, P.: Obligations and cooperative behaviour in public good games. *Games Econ. Behav.* **64**(1), 146–170 (2008)