Journal of Theoretical Biology 299 (2012) 144–151

Contents lists available at ScienceDirect

Journal of Theoretical Biology

journal homepage: www.elsevier.com/locate/yjtbi

Cooperation and evolutionary dynamics in the public goods game with institutional incentives

Ross Cressman [a,*], Jie-Wen Song [b,c], Bo-Yu Zhang [d], Yi Tao [c,*]

[a] Department of Mathematics, Wilfrid Laurier University, Waterloo, Ontario, Canada
[b] School of Mathematical Sciences, Beijing Normal University, Beijing, China
[c] Key Lab of Animal Ecology and Conservational Biology, Centre for Computational and Evolutionary Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing, China
[d] Department of Mathematics, University of Vienna, Austria

ARTICLE INFO

Available online 11 August 2011

Keywords:
Cooperation
Public goods game
Reward and punishment
Replicator equation
Adaptive dynamics

ABSTRACT

The one-shot public goods game is extended to include institutional incentives (i.e. reward and/or punishment) that are meant to promote cooperation. It is shown that the Nash equilibrium (NE) outcomes predict either partial or fully cooperative behavior in these extended multi-player games with a continuous strategy space. Furthermore, for some incentive schemes, multiple NE outcomes are shown to emerge. Stability of all these equilibria under standard evolutionary dynamics (i.e. the replicator equation and the canonical equation of adaptive dynamics) is characterized.

1. Introduction

The public goods game (PGG) is often used to investigate how cooperation can emerge and be maintained in theoretical models of multi-player group interactions. In the standard one-shot PGG (Sigmund, 2010), the only rational individual behavior (i.e. the only NE outcome) is for no player to contribute to the public good (i.e. to defect) whereas it is to everyone's advantage for all players to contribute their total endowment (i.e. to cooperate). Experimental evidence (Fehr and Gächter, 2000; Herrmann et al., 2008; Rand et al., 2009) shows that actual contributions in PGG are somewhere between these two extremes with typical contribution levels from 20% to 70% of the endowment.

In real-life situations that include this sort of dilemma between individual rationality and collective advantage, incentives are often used to promote cooperation. For example, individuals who are high contributors may be rewarded (e.g. businesses often give bonuses to their best performing employees) and those who contribute little may be punished (e.g. people who cheat on taxes are subject to fines by the state). Theoretical models (Sigmund et al., 2001; Hauert et al., 2004) of PGG with incentives are often based on group members rewarding and/or punishing each other (called peer incentives) after everyone's contribution is reported. In this paper, the effects of institutional incentives are modeled instead. Here, each group member knows the incentive amounts and the probabilities that he will be rewarded and/or punished given his contribution and those of the other group members. Since these reward/punishment schemes, based on institutional rather than peer incentives, are more closely related to such examples as those mentioned above, it is important to study their effect on promoting cooperation in theoretical models.

Our reward scheme is also related to the use of lotteries to finance public goods (Morgan, 2000; Corazzini et al., 2010). For example, the institution (e.g. government, charitable organization) holds a lottery to raise funds for a public project. The more an individual contributes (and the less contributed by other people), the better his chance to win the prize provided by the lottery (i.e. the reward).

Our main goal is then to investigate how institutional incentives affect individual rational behavior. This is initially done by characterizing the NE outcome for our PGG with incentives. As we will see, defection is not always rational in these games. In fact, for certain incentive schemes, there is more than one possible rational behavior. For such game-theoretic situations, evolutionary dynamics can be used to predict the expected outcome (Hofbauer and Sigmund, 1998). We therefore analyze stability of the NE outcome under such standard evolutionary dynamics as the replicator equation and the canonical equation of adaptive dynamics.

The paper is organized as follows. Section 2 briefly summarizes the standard one-shot PGG and evolutionary dynamics for this multi-player game. In particular, it is shown that these dynamics agree with the NE prediction of defection by all group members. Section 3 extends the analysis of the PGG model to include institutional reward (Section 3.1); punishment (Section 3.2); reward and punishment (Section 3.3). The theory is illustrated in these sections by considering in detail four specific connected examples where the effectiveness of these three schemes in promoting cooperation can be compared.

* Corresponding authors.
E-mail addresses: rcressma@wlu.ca (R. Cressman), yitao@ioz.ac.cn (Y. Tao).

0022-5193/$ - see front matter © 2011 Elsevier Ltd. All rights reserved.
doi:10.1016/j.jtbi.2011.07.030

2. The public goods game and evolutionary dynamics

Suppose there are $n \geq 2$ players who are each given an initial endowment $E > 0$. Each player decides how much x of this endowment to contribute to a common pool (i.e. $x \in [0,E]$). All contributions to the common pool are multiplied by a factor $r > 1$ and then evenly distributed among all n players. A player's payoff is then the remainder of his endowment $(E-x)$ plus what he receives from the public pool. If he contributes x and the other $n-1$ players contribute $x_2, x_3, \ldots, x_n$ respectively, this payoff is given as

$$\pi(x; x_2, x_3, \ldots, x_n) = E - x + \frac{r}{n}\sum_{i=1}^{n} x_i = E + \left(\frac{r}{n} - 1\right)x + \frac{r}{n}\sum_{i=2}^{n} x_i \quad (1)$$

We assume that the player receives only a fraction of his own contribution to the common pool (i.e. $r/n < 1$). That is, the player's return on his contribution is less than 100%. Thus, $1 < r < n$.

It is well known that the only Nash equilibrium (NE) of this game is for each player to contribute 0 (i.e. to free-ride). To see this, recall that a NE $(x_1, x_2, x_3, \ldots, x_n)$ of this n-player game must satisfy $\pi(y; x_2, x_3, \ldots, x_n) \leq \pi(x_1; x_2, x_3, \ldots, x_n)$ for all $y \in [0,E]$. That is, $(r/n-1)y \leq (r/n-1)x_1$ and so $x_1 \leq y$ for all $y \in [0,E]$. Thus, player 1 must free-ride.[1] Intuitively, since a player's payoff decreases as the amount he contributes increases given that the other players' contributions remain the same, each player has an incentive to reduce his contribution.

The public goods game is an example of an n-player game that has the form of a population game since a player's payoff $\pi(x; x_2, x_3, \ldots, x_n)$ depends only on his strategy x and the average strategy $x_{-1} \equiv (1/(n-1))\sum_{i=2}^{n} x_i$ of the rest of his group. Moreover, it is a symmetric game in the sense that payoffs do not depend on designating who is player 1, who is player 2, etc. Evolutionary dynamics for symmetric population games have been studied extensively, especially when there is a finite set of pure strategies (Hofbauer and Sigmund, 1998; Sandholm, 2010). When there is a continuum of pure strategies (as in our case where the strategy set is the one-dimensional interval $[0,E]$), the standard evolutionary dynamics are the canonical equation of adaptive dynamics (Dieckmann and Law, 1996) and the replicator dynamics for a continuous strategy space (Cressman and Hofbauer, 2005).

2.1. Replicator dynamics

The replicator equation assumes that the population state is described by a Borel probability measure P over [0,E]. That is, if B is a Borel subset of $[0,E]$, then $P(B)$ equals the proportion of the population using strategies in B. In particular, the population is assumed to be sufficiently large that finite population effects can be ignored and so the expected payoff of an individual playing y in a group whose other $n-1$ players are chosen at random is

$$\pi(y; P) \equiv \int_{[0,E]^{n-1}} \pi(y; x_2, x_3, \ldots, x_n) P(dx_2)\ldots P(dx_n) = E + \left(\frac{r}{n} - 1\right)y + \frac{r}{n}\sum_{i=2}^{n}\int_{[0,E]^{n-1}} x_i P(dx_2)\ldots P(dx_n) = E + \left(\frac{r}{n} - 1\right)y + \frac{r}{n}\bar{x}(n-1)$$

where $\bar{x} \equiv \int_{[0,E]} yP(dy)$ is the average contribution of an individual in the population.

This measure-theoretic formulation of the game generalizes the standard approach for symmetric population games with a finite set of pure strategies. For instance, if all players use one of N strategies $X_1, X_2, \ldots, X_N$ in $[0,E]$ (i.e. if P has finite support), then the average strategy of the population is $\bar{X} = \sum_{i=1}^{N} p_i X_i$ where p_i is the proportion of the population using strategy X_i. That is, $P = \sum_{i=1}^{N} p_i \delta_{X_i}$ where δ_X is the Dirac delta distribution that places all its weight at X (i.e. $\delta_X(\{X\}) = 1$). It is well-known (Hofbauer and Sigmund, 1998) that the support of P is invariant under the replicator equation

$$\dot{p}_i = p_i(\pi(X_i; P) - \pi(P; P)) \quad (2)$$

where $\pi(P; P) \equiv \sum_{i=1}^{N} p_i \pi(X_i; P) = E + (r-1)\bar{X}$ is the average payoff of an individual in the population. Since $\pi(X_i; P) = E + (r/n-1)X_i + (r/n)\bar{X}(n-1)$, $\dot{p}_i = p_i(1-r/n)(\bar{X} - X_i)$. In particular, $\dot{p}_1 > 0$ if X_1 is the strategy corresponding to the smallest contribution of all players initially present. Thus, the population evolves to everyone using this strategy (i.e. $p_1 \to 1$). These results generalize to the case of infinitely many strategies as follows.

The replicator equation, when the support of P (i.e. the smallest closed subset of $[0,E]$ whose complement has measure 0) is infinite, is given by

$$\frac{dP(B)}{dt} = \int_B (\pi(y; P) - \pi(P; P))P(dy) \quad (3)$$

Here $\pi(P; P)$ is $\int_{[0,E]} \pi(y; P)P(dy) = E + (r-1)\bar{x}$. Since our strategy space is compact and the payoff functions are continuous, this is a well-defined dynamics with a unique solution P_t for $t \geq 0$ in the set $\Delta([0,E])$ of Borel measures for every initial P_0 (Bomze, 1991; Oechssler and Riedel, 2001). Consider the evolution of $\bar{x}$ under (3). This is

$$\begin{aligned}\frac{d\bar{x}}{dt} &= \int_{[0,E]} y\frac{dP}{dt}(dy) = \int_{[0,E]} y(\pi(y; P) - \pi(P; P))P(dy)\\ &= \int_{[0,E]} y\left[E + \left(\frac{r}{n} - 1\right)y + \frac{r}{n}\bar{x}(n-1) - (E + (r-1)\bar{x})\right]P(dy)\\ &= \int_{[0,E]} \left[\left(\frac{r}{n} - 1\right)y^2 + y\bar{x}\left(1 - \frac{r}{n}\right)\right]P(dy)\\ &= \left(\frac{r}{n} - 1\right)\int_{[0,E]} (y - \bar{x})^2 P(dy)\\ &\leq 0\end{aligned}$$

with equality if and only if $P = \delta_x$ for some $x \in [0,E]$. Since P_t has the same support as P_0 for all $t \geq 0$, $\bar{x}$ evolves to the smallest element x^* in the support of P_0.[2] From this it follows that P_t evolves to δ_{x^*} in the weak topology of $\Delta([0,E])$.

In particular, if there are some free-riders in the original population distribution, the proportion of players who contribute more than ε will eventually be less than ε no matter how small $\varepsilon > 0$ is taken (i.e. P_t evolves to δ_0 in the weak topology). If rare mutations are also allowed, we can expect that, at some point, free-riders will appear and from then on the average contribution in the population will evolve to 0.

2.2. Adaptive dynamics

The canonical equation of adaptive dynamics (Dieckmann and Law, 1996) assumes the population is monomorphic (i.e. at some δ_x) and that this monomorphism evolves through trait substitution in the direction of nearby strategies that have higher payoff than the resident strategy when played against the resident population. For x in the interior of $[0,E]$ (i.e. for $0 < x < E$), the

[1] The same argument shows that each of the players must free-ride at a pure-strategy NE. The extension of this argument shows that no player can use a mixed strategy at a NE.

[2] This conclusion also follows from the fact that x^* strictly dominates y (i.e. $\pi(y,P) < \pi(x^*,P)$ for all $P \in \Delta([0,E])$) for any y other than x^* in the support of P_t (see Cressman and Hofbauer, 2005).

equation is given by

$$\frac{dx}{dt} = k(x)\frac{\partial \pi(y;x)}{\partial y}\bigg|_{y=x} \tag{4}$$

where $k(x)$ is a positive function of x that is related to how fast new traits appear.[3] Here, $\pi(y;x)$ is used in place of $\pi(y;x_2, x_3,\ldots,x_n)$ since $x_2 = x_3 = \cdots = x_n = x$ (i.e. $\pi(y;x) = E+(r/n-1)y+(r/n)x(n-1)$). Thus $dx/dt = r/n-1 < 0$ and so x evolves to 0.

In summary, for the PGG (without incentives), the evolutionary outcome under either the replicator (with occasional mutations) or the canonical equation is that the population eventually adopts the unique NE of free-riding.

3. Institutional incentives

Institutions provide incentives to their members in order to promote higher contributions to the common pool (i.e. to increase cooperative behavior). Incentive schemes vary greatly from one institution to another but a common feature is that individual members who are considered high contributors are given rewards while lower contributing individuals are sanctioned (i.e. punished).

We model institutional incentives as follows. After the standard PGG of Section 2 is played between n players, a second stage is added where one institutional member is chosen to be rewarded (institutional reward, IR) or one institutional member is chosen to be punished (institutional punishment, IP) or both (institutional reward and punishment, IRP). The probability of being chosen for a reward and/or punishment depends on the member's contribution and those of the rest of the group. The incentive amount (i.e. the additional payoff given to the member in IR or taken away in IP) is fixed at $A > 0$. We examine how NE behavior and evolutionary outcomes depend on this amount and the probabilities of being chosen for IR (Section 3.1), IP (Section 3.2) and IRP (Section 3.3).

3.1. Institutional reward

An institution rewards those individuals who make higher contributions in order to encourage all its members to be more cooperative. This objective is reflected in our model by assuming that the probability, $F(x;x_2,x_3,\ldots,x_n)$, an individual who contributes x in a group whose other members contribute $x_2,x_3,\ldots,x_n$ receives the reward is a continuously differentiable function that satisfies

(i) exactly one individual receives the reward;
(ii) $F = 1/n$ if all members contribute the same;
(iii) F is an increasing function of x;
(iv) F depends only on contribution x and the average contribution x_{-1} of the rest of the group.

The institution may prefer to only reward its highest contributor (i.e. $F(x;x_2,x_3,\ldots,x_n) = 1$ if $x > \max\{x_2,x_3,\ldots,x_n\}$) but this is not a continuous function. Moreover, this incentive scheme is inconsistent with (iv) since the highest contribution may not be x when $x > x_{-1}$. We examine consequences of rewarding only the highest contributor at the end of this section.

One interpretation of our conditions on F is that the institution uses an imperfect ranking system of its members (based on not knowing the exact contribution of each individual) and rewards the person with the highest rank. Alternatively, an institution that knows the exact contributions may still feel a reward scheme satisfying our conditions will be more effective at promoting cooperation by providing an incentive for members who are not the highest contributor to incrementally increase their contributions.

By assumption (iv), IR is a symmetric population game and so the simplified notation $F(x;x_{-1})$ can be used in place of $F(x;x_2, x_3,\ldots,x_n)$ for our purposes (and $\pi(x;x_{-1})$ in place of $\pi(x;x_2, x_3,\ldots,x_n)$). The reward changes the payoff function (1) to

$$\pi(x;x_{-1}) = E+\left(\frac{r}{n}-1\right)x+\frac{r}{n}x_{-1}(n-1)+AF(x;x_{-1}) \tag{5}$$

If the reward A is high enough, free-riding is no longer a NE. In general, x^* is a symmetric NE (i.e. $\pi(x;x^*) \le \pi(x^*;x^*)$ for all $x \in [0,E]$) if and only if

$$A \le \left(1-\frac{r}{n}\right)\frac{x-x^*}{F(x;x^*)-F(x^*;x^*)} \tag{6}$$

for all $x \ne x^*$.

That is, 0 is a NE if and only if

$$A \le \left(1-\frac{r}{n}\right)\frac{x}{F(x;0)-F(0;0)} \tag{7}$$

for all $x \in (0,E]$.[4] Let us examine how this compares to the stability for evolutionary dynamics. First, adaptive dynamics is

$$\frac{dx}{dt} = \frac{\partial \pi(y;x)}{\partial y}\bigg|_{y=x} = \left(\frac{r}{n}-1\right)+A\frac{\partial F(y;x)}{\partial y}\bigg|_{y=x}$$

for $0 < x < E$. If $A < (1-r/n)((\partial F(y;0)/\partial y)|_{y=0})^{-1}$, then $x=0$ is convergence stable, since $dx/dt < 0$ for all positive x close to 0.[5] The inequalities for NE (7) and for convergence stability are closely related since $((\partial F(y;0)/\partial y)|_{y=0})^{-1} = \lim_{x\to 0^+} x/(F(x;0)-F(0;0))$. Specifically, except in the threshold case where $A = (1-r/n)\times((\partial F(y;0)/\partial y)|_{y=0})^{-1}$, if 0 is a NE, then 0 is convergence stable. The converse is not true (i.e. convergence stability of 0 does not imply 0 is a NE) since convergence stability of $x^* \in [0,E]$ relies only on properties of F near x^* whereas the NE conditions on F require comparisons for all $x \in [0,E]$. However, if $F(x;x^*)$ is a concave function of x, then the converse is true since $x/(F(x;0)-F(0;0))$ is then an increasing function of x.

Analogous results for convergence stability of the other endpoint are straightforward to obtain. In summary, E is convergence stable if $A > (1-r/n)((\partial F(y;E)/\partial y)|_{y=E})^{-1}$ and, from (6), a NE if and only if $A \ge (1-r/n)(E-x)/(F(E;E)-F(x;E))$ for all $x \in [0,E)$. Furthermore if $F(x;x^*)$ is a concave function of x, then E is convergence stable if and only if it is a NE except in the threshold case $A((\partial F(y;E)/\partial y)|_{y=E}) = (1-r/n)$.

Finally, the NE and convergence stability criteria are well-known for an $x^* \in (0,E)$ (Dieckmann and Law, 1996). An interior NE x^* is a rest point of the canonical equation but may not be convergence stable. Conversely, a (convergence stable) rest point may not be a NE. In general, there may be several rest points, some of which are convergence stable or NE and others that are not as in Section 3.3 below (see also Cressman, 2009). In our IR model, the classification of rest points is based on properties of the reward function F. An interior x^* is a rest point if and only if $A((\partial F(y;x^*)/y)|_{y=x^*}) = (1-r/n)$ and a rest point is convergence stable if and only if

$$\frac{d}{dx}\left(\frac{dx}{dt}\right) = \frac{d}{dx}\left(\frac{\partial F(y;x)}{\partial y}\bigg|_{y=x}\right)\bigg|_{x=x^*} \equiv F_{11}+F_{12} < 0 \tag{8}$$

[3] A rest point of the canonical equation that is asymptotically stable under all such choices of $k(x)$ is called convergence stable (Christiansen, 1991). Here, we can ignore this factor (i.e. take $k(x) = 1$) since our strategy space is one-dimensional and so $k(x)$ only effects the speed of evolution and not the eventual outcome.

[4] The right-hand side of this inequality is positive for each such x by property (iii) of F since $r/n < 1$.

[5] If $(\partial F(y;0)/\partial y)|_{y=0} = 0$, 0 is convergence stable for any choice of A.

where F_{ij} are the second order partial derivatives of F evaluated at x^*.[6] On the other hand, if $x^* \in (0,E)$ is a NE, then $F_{11} < 0$ and the converse is true if F is concave.

Concavity holds in the following example where it is shown that there is exactly one rest point in $[0,E]$ and it is both a NE and convergence stable.

Example 1. There are many incentive schemes that fit our IR model. We will look in depth at a four player example with parameters $n=4$, $E=20$ and $r=1.6$ since these values are commonly used in experimental games examining cooperation in PGG (e.g. Herrmann et al., 2008; Rand et al., 2009). In these experimental games, incentives are typically implemented through peer decisions (i.e. an individual may be rewarded or punished by other members of his group, often at a cost to themselves) rather than through an institutional scheme as in our model. Here, the reward probabilities are taken as

$$F(x;x_{-1}) = \frac{x+1}{x+3x_{-1}+4}, \tag{9}$$

for which it is straightforward to verify conditions (i)–(iv) above for IR. Moreover, $F(x;x_{-1})$ is a concave function of x for fixed x_{-1}.

In particular, $x/(F(x;0)-F(0;0))$ is an increasing function of x and so 0 is a NE if $A \le (1-r/n)((\partial F(x;0)/\partial x)|_{x=0})^{-1}$ but not if $A > (1-r/n)((\partial F(x;0)/\partial x)|_{x=0})^{-1}$.[7] In fact, 0 is a NE if and only if it is convergence stable. Similarly, $(E-x)/(F(E;E)-F(x;E))$ is an increasing function of x and so E is a NE if and only if it is convergence stable (if and only if $A \ge (1-r/n)((\partial F(x;E)/\partial x)|_{x=E})^{-1} = \frac{16}{3}(1+E)(1-r/n) = 67.2$).

An $x^* \in (0,E)$ is a rest point of the canonical equation if and only if $A((\partial F(x;x^*)/\partial x)|_{x=x^*}) = (1-r/n)$. From (9), $(\partial F(y;x^*)/\partial y)|_{y=x^*} = \frac{3}{16}\cdot 1/(x^*+1)$ and so $x^* = (3A/16)(1-r/n)^{-1} - 1 = 5A/16 - 1$. Thus, x^* is unique and it is in the interval $(0,E)$ if and only if $3.2 < A < 67.2$. Furthermore, $F_{11} = -6/64(x^*+1)^2 = F_{12}$ and so this point is convergence stable and a NE if it exists.

In summary, for each $A > 0$ in this example, there is a unique NE given by

$$x^* = \begin{cases} 0 & \text{if } A \le 3.2 \\ \frac{5A}{16} - 1 & \text{if } 3.2 < A < 67.2 \\ 20 & \text{if } A \ge 67.2 \end{cases} \tag{10}$$

and it is a globally stable rest point of the canonical equation. For small rewards, individuals free-ride. For $A > 3.2$, cooperation increases linearly in the reward until the population is fully cooperative once $A=67.2$.

The replicator equation with institutional incentives is more difficult since the probability an individual receives the reward is not linear either in his contribution or in the contributions of the other group members. Specifically,

$$\pi(y;P) = E + \left(\frac{r}{n}-1\right)y + \frac{r}{n}\bar{x}(n-1) + \int_{[0,E]^{n-1}} F\left(y;\frac{x_2+x_3+\cdots+x_n}{n-1}\right) \times P(dx_2)\cdots P(dx_n) \text{ and}$$

$$\pi(P;P) = E + \bar{x}(r-1) + \int_{[0,E]^n} F\left(x_1;\frac{x_2+x_3+\cdots+x_n}{n-1}\right) \times P(dx_1)P(dx_2)\cdots P(dx_n).$$

A complete analysis of this dynamics is not practical. Instead, we simulate the dynamics in the following example.

Example 2. Continue with the same parameters and reward probabilities as in Example 1 and take $A=20$. From Example 1, $x^* = 5.25$ is the unique NE and it is convergence stable (in fact, it is globally asymptotically stable for the canonical equation). Simulations of the replicator equation are shown in Fig. 1 where the pure strategies are taken as $X_i = i$ for $i = 0,\ldots,20$ (i.e. a finite set). This strategy set models the situation where the initial endowment is made up of 20 indivisible monetary units and so individuals must contribute an integer number of units. From (2) and (3), the dynamics is

$$\dot{p}_i = p_i\left[(1-\tfrac{1.6}{4})(\bar{X}-i) + \sum_{k_1=0}^{20}\cdots\sum_{k_4=0}^{20} A\left[F\left(i;\frac{k_2+k_3+k_4}{3}\right) - F\left(k_1;\frac{k_2+k_3+k_4}{3}\right)\right]p_{k_1}p_{k_2}p_{k_3}p_{k_4}\right] \tag{11}$$

for $i = 0,1,\ldots,20$ where p_i is the frequency of strategy X_i in the large population. Here $\bar{X} = \sum_{i=0}^{20} ip_i$ is the expected contribution of a randomly chosen individual, $F(i;(k_2+k_3+k_4)/3) = (i+1)/(i+k_2+k_3+k_4+4)$ and $F(k_1;(k_2+k_3+k_4)/3) = (k_1+1)/(k_1+k_2+k_3+k_4+4)$. In particular, $F(i;x_{-1})$ is not linear in either i or in $x_{-1} = (k_2+k_3+k_4)/3$.

Fig. 1 also gives the corresponding simulations for the mean-field replicator equation (for this same strategy set) that approximates the replicator equation by assuming the expected payoff of an individual using strategy i is his payoff in a group where the average contribution of the other members is the mean contribution of the population. That is,

$$\frac{dp_i}{dt} = p_i(\pi(i,\bar{X}) - \pi(\bar{X},\bar{X}))$$

$$\pi(i;\bar{X}) = 20 - i + \frac{r}{n}(i + (n-1)\bar{X}) + AF(i;\bar{X})$$

$$\pi(\bar{X},\bar{X}) = 20 - \bar{X} + r\bar{X} + A/n \tag{12}$$

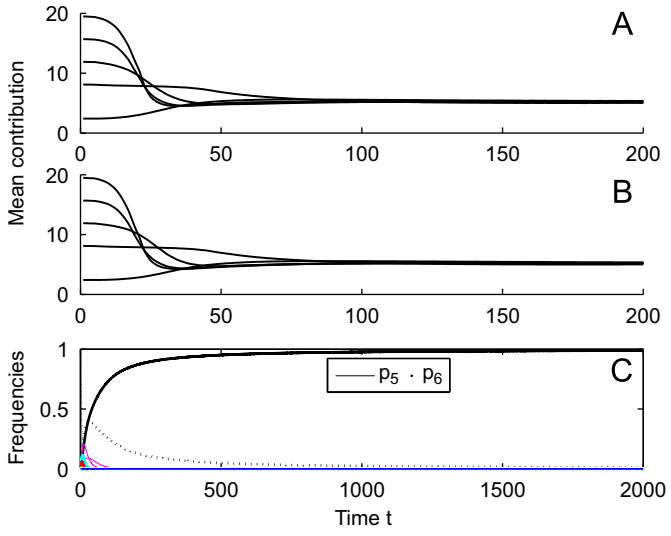

Fig. 1. The replicator equation and its mean-field approximation for IR (Example 2). Panel A shows the mean contribution along five trajectories of the replicator equation whose initial distribution is $p_i=0.0025$ for all $i = 0,1,\ldots,20$ except for one choice of i. These five choices are $p_{20}=0.95$ for the trajectory that has the highest initial mean contribution; $p_{16}=0.95$ for the trajectory with the next highest initial contribution; then $p_{12}=0.95$; $p_8=0.95$; and finally $p_2=0.95$ for the trajectory with the lowest initial contribution. Panel B shows the corresponding five mean contributions of the mean-field replicator equation. Panel C plots $p_i(t)$ for $i = 0,1,\ldots,20$ under the replicator equation with initial distribution $p_i=0.0025$ for all $i = 0,1,\ldots,20$ except $p_2=0.95$. Parameters are taken as in Example 2: $n=4; r=1.6; A=20; E=20$.

[6] Here, we ignore the threshold cases $F_{11}+F_{12}=0$ for convergence stability and $F_{11}=0$ for NE.

[7] From (9), $A = (1-r/n)((\partial F(x;0)/\partial x)|_{x=0})^{-1} = (1-\frac{1.6}{6})\frac{16}{3} = 3.2$.

From Panels A and B, it is clear that the mean field approach approximates the replicator equation well. Panel C indicates that $p_5 \to 1$ (i.e. the population evolves to the monomorphism where all individuals contribute 5). If $x^* = 5.25$ is in the support of the initial distribution P_0, then P_0 evolves to $\delta_{5.25}$ in the simulations (the analytic proof of this result remains an open problem).[8]

3.1.1. Rewarding the highest contributor

Suppose the institution gives the reward A to the highest contributor and, in the case where several group members make the highest contribution, gives each such member an equal share of A. That is,

$$F(x;x_2,x_3,\ldots,x_n)=\begin{cases}1 & \text{if } x>\max\{x_2,x_3,\ldots,x_n\}\\ \dfrac{1}{k+1} & \text{if} x=\max\{x_2,x_3,\ldots,x_n\} \text{ and}\\ & x=x_i \text{ for } k \text{ elements of } \{2,\ldots,n\}\\ 0 & \text{if } x<\max\{x_2,x_3,\ldots,x_n\}\end{cases} \tag{13}$$

First, assume that the population is monomorphic at $x \in [0,E]$. Then an individual contributing y has payoff

$$\pi(y;x)=\begin{cases}E+\left(\frac{r}{n}-1\right)y+\frac{r}{n}x(n-1)+A & \text{if } y>x\\ E+\left(\frac{r}{n}-1\right)y+\frac{r}{n}x(n-1)+\frac{1}{n}A & \text{if } y=x\\ E+\left(\frac{r}{n}-1\right)y+\frac{r}{n}x(n-1) & \text{if } y<x\end{cases} \tag{14}$$

Thus, the explicit form of the canonical equation in (4) is no longer applicable since $\pi(y;x)$ is discontinuous at $y=x$. However, for any reward $A>0$, $\pi(x+\varepsilon;x)>\pi(x;x)>\pi(x-\varepsilon;x)$ when $\varepsilon>0$ is sufficiently small. That is, it is always to an individual's advantage to contribute a small amount more than what everyone else does in order to reap the full reward (and to his disadvantage to contribute a little less). Thus, through trait substitution, adaptive dynamics moves the population in the direction of ever increasing contributions (i.e. there is runaway selection in favor of cooperation (Nakamaru and Dieckmann, 2009)) until the population consists entirely of maximum contributors. That is, $x=E$ is convergence stable.

On the other hand, contributing E is not a NE unless the reward is quite high. For instance, with the parameters from Example 1 where $E=20$, $\pi(0;20)>\pi(20;20)$ whenever $A<48$ and so free-riders can invade the monomorphic population of full contributors in this case. Such situations (i.e. a convergence stable rest point that is not a NE) are often analyzed in adaptive dynamics through evolutionary branching (Geritz et al., 1998; Doebeli and Dieckmann, 2000) whereby there is an initial splitting into a dimorphic population with both strategies near the convergence stable rest point. However, in our model, initial successful invaders must be far away from the rest point 20. We instead consider the replicator dynamics with parameters as in Examples 1 and 2 (in particular, $A=20$) and reward function (13).[9]

For the numerical simulations of (11) in Fig. 2 (Panel B), the initial distribution has 95% of the population contributing 2 and the other 5% evenly distributed among the other 20 strategies $\{X_i=i \,|\, i=0,1,3\ldots,20\}$. Panel B shows an initial successful invasion by full cooperators (i.e. $p_{20}(t)$ is initially increasing fastest). Later free-riding behavior becomes more prevalent (i.e. $p_0(t)$ increases) and so there is an advantage for individuals contributing slightly more whereby we see contributions up to about six appearing in the system. In the long-run, the dynamics converges to a steady-state distribution with a significant share (about 41%) fully cooperative. There are also significant shares for low contributors that diminish as the contribution goes up and disappear for contribution levels between 7 and 19. Interestingly, this steady-state distribution seems to be independent of the initial distributions with support $\{X_i=i \,|\, i=0,\ldots,20\}$.[10] In particular, this occurred for all initial distributions reported in Fig. 2 (Panel A) which shows that the mean contributions are evolving toward the same value.

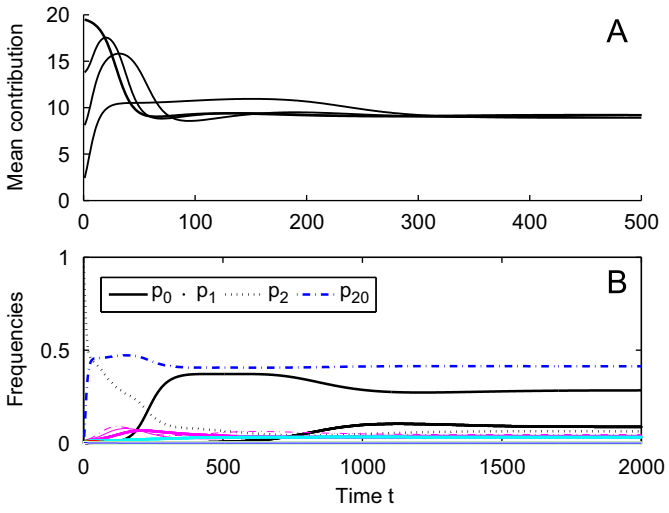

Fig. 2. The replicator equation for IR with discontinuous reward scheme (Section 3.1.1). Panel A shows the mean contribution along four trajectories of the replicator equation whose initial distribution is $p_i=0.0025$ for all $i=0,1,\ldots,20$ except for one choice of i. These four choices are $p_{20}=0.95$ for the trajectory that has the highest initial mean contribution; $p_{14}=0.95$ for the trajectory with the next highest initial contribution; then $p_8=0.95$; and finally $p_2=0.95$ for the trajectory with the lowest initial contribution. Panel B plots $p_i(t)$ for $i=0,1,\ldots,20$ under the replicator equation with initial distribution $p_i=0.0025$ for all $i=0,1,\ldots,20$ except $p_2=0.95$. Parameters are the same as in Fig. 1.

It is clear that no pure strategy x^* can be stable under the replicator equation since monomorphic populations of high contributors can be invaded by free-riders and individuals in monomorphic populations of low contributors have an incentive to contribute slightly more. The argument here is essentially the result that no $x^* \in [0,20]$ is a NE as shown by Corazzini et al. (2010) for a related lottery model with a prize given to the highest contributor. On the other hand, numerical simulations suggest that the limiting distribution shown in Fig. 2 (Panel B) is a polymorphic NE of the game restricted to the strategy set $\{X_i=i \,|\, i=0,\ldots,20\}$ since all strategies with positive shares have (approximately) the same payoffs.

For the remainder of the paper, we again assume that incentives depend continuously on contributions.

3.2. Institutional punishment

Let $G(x;x_{-1})$ be the probability the institution punishes an individual who contributes x in a group whose other members contribute x_{-1} on average. We assume G satisfies the same four properties as F at the beginning of the previous section except that property (iii) is replaced by:

(iii)′ G is a decreasing function of x (i.e. institutions punish low contributors in a group more frequently than high contributors).

[8] Numerical simulations of (11) and (12) were also carried out for other values of A between 0 and 70. In all cases considered, the population evolved to a monomorphism on one side of x^* predicted by (10) for the continuous strategy space.

[9] The mean-field Eq. (12) cannot be used here since the reward function (13) cannot be written in the form $F(x;x_{-1})$ (i.e. the reward to x is not determined uniquely by the average strategy x_{-1} of the rest of the group).

[10] The analytic proof of independence remains an open problem.

One possibility for G is to take the probability of punishment symmetric about the interval $[0,E]$ with the probability of reward in the previous section. That is, the probability that an individual who contributes x is punished in a group whose other members contribute x_{-1} on average is the same probability that an individual who contributes $E-x$ is rewarded in the previous section when the other members of his group contribute $E-x_{-1}$ on average. With this assumption, we have

$$G(x;x_{-1}) = F(E-x;E-x_{-1}) \tag{15}$$

The payoff function is now

$$\pi(x;x_{-1}) = E+\left(\frac{r}{n}-1\right)x+\frac{r}{n}x_{-1}(n-1)-AG(x;x_{-1})$$

where A is now the amount of punishment. Following the steps in Section 3.1, we find that 0 is a NE if and only if[11]

$$A \le \left(1-\frac{r}{n}\right)\frac{x}{G(0;0)-G(x;0)}$$

for all $x \in (0,E]$. Similarly, E is a NE if and only if $A \ge (1-r/n)\times (E-x)/(G(x;E)-G(E;E))$ for all $x \in [0,E]$.

Example 3. Let us continue Example 1 from Section 3.1 by replacing the reward scheme there by its symmetric punishment function. That is, from (9) and (15), $G(x;x_{-1}) = (21-x)/(84-x-3x_{-1})$ which is again a concave function of x. Then 0 is a NE if $A \le ((21\times 16)/3)(1-r/n) = 67.2$ and $E=20$ is a NE if $A \ge (16(21-E)/3)(1-r/n) = 3.2$. That is, punishment must be quite high to overcome the free-riding advantage in a group of free-riders whereas fully cooperative groups are NE already at low levels of punishment. In particular, at least one endpoint is a NE (and convergence stable) for any fixed value of $A>0$. On the other hand, if an interior x^* is a rest point of the canonical equation

$$\frac{dx}{dt} = \left(\frac{r}{n}-1\right)-A\frac{\partial G(y;x)}{\partial y}\bigg|_{y=x} = -0.6+\frac{3A}{16}\cdot\frac{1}{21-x}$$

then $x^* = 21-5A/16$ (cf. Eq. (10)). That is, $x^* \in (0,E)$ if and only if $3.2 < A < 67.2$. However, since the incentive $-AG(x;x_{-1})$ is now convex in x (i.e. $-G_{11}>0$) no such interior x^* is a NE and it is not convergence stable either.[12] In particular, trajectories of the canonical equation evolve monotonically to the endpoint on the same side of x^* as the initial point. For low levels of punishment, x^* is close to 1 and so the NE 0 has the larger basin of attraction. As A increases, the basin of attraction of the NE 20 gets larger until it attracts all initial points when $A \ge 67.2$.

The simulations in Fig. 3 are based on IP with $A=20$ (as in Example 2). For this game, both 0 and 20 are NE and convergence stable. The other (unstable) rest point of the canonical equation is 14.75. Fig. 3 summarizes five trajectories of the replicator equation and its mean field approximation using the method of Example 2 applied to the punishment incentive scheme. Again, the approximation is almost identical to the replicator equation as the mean contribution evolves along each trajectory.[13] In the long run, all trajectories converge to either 0 (i.e. $p_0=1$) or to 20 (i.e. $p_{20}=1$). Interestingly, some trajectories that start with initial mean contribution above 14.75 evolve to 0; whereas they evolve to 20 for the canonical equation. The nonlinear nature of the payoff function results in substantial differences between the outcome predicted by the canonical equation and those from the replicator equation.

[11] Note that $x/(G(0;0)-G(x;0))$ is positive for $x>0$ since G is a decreasing function of x.

[12] This follows from $-G_{11}-G_{12}>0$ (cf. Eq. (8)).

[13] Other simulations can be used to show that trajectories of the replicator equation and its mean-field approximation are not identical. By carefully choosing the initial distribution near a threshold value, the trajectory of the replicator equation evolves to 0 and the mean-field approximation evolves to 20.

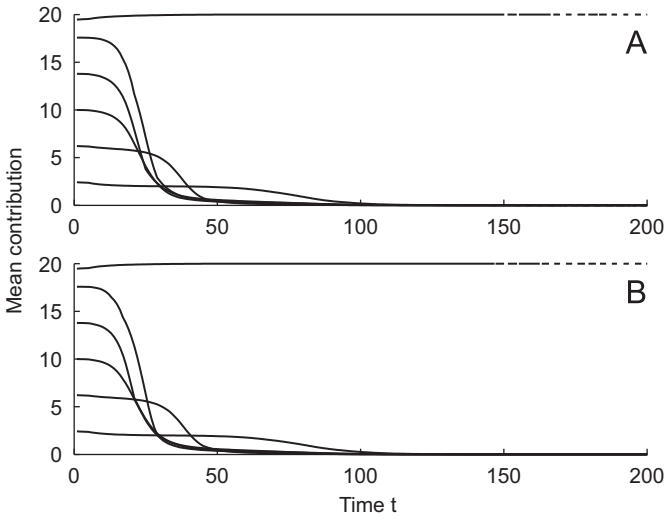

Fig. 3. The replicator equation and its mean-field approximation for IP (Example 12). Panel A shows the mean contribution along each of the five trajectories of the replicator equation whose initial distribution is the same as in Fig. 1. Panel B shows the corresponding results for the mean-field replicator equation. Parameters are the same as in Fig. 1.

Examples 2 and 3 (with $A=20$) show clearly that a positive incentive (reward) can have quite a different effect on NE behavior than a negative incentive (punishment) (cf. Rand et al., 2009). With these incentive schemes, an institution interested in promoting cooperation when cooperation is low is more effective by rewarding its members for better behavior than to punish for worse behavior. However, such rewards are not able to maintain high levels of cooperation whereas punishment is effective at encouraging full cooperation once it is established at intermediate levels in the population. Example 4 of the following section examines the combined effect of these two schemes.

3.3. Institutional reward and punishment

Here, one individual is chosen for the reward and, independently, one individual is chosen to be punished.[14] If the amount of reward is A and the punishment amount B, the payoff function is now

$$\pi(x;x_{-1}) = \pi(x;x_2,x_3,\ldots,x_n) = E+\left(\frac{r}{n}-1\right)x+\frac{r}{n}x_{-1}(n-1)+H(x;x_{-1}) \tag{16}$$

where $H(x;x_{-1}) \equiv AF(x;x_{-1})-BG(x;x_{-1})$.

Example 4. Let us continue the above three examples and assume that $A=B$. The canonical equation is

$$\frac{dx}{dt} = \left(\frac{r}{n}-1\right)+A\frac{\partial F(y;x)}{\partial y}\bigg|_{y=x}-A\frac{\partial G(y;x)}{\partial y}\bigg|_{y=x}$$

where

$$F(y;x) = \frac{y+1}{y+3x+4}$$

$$G(y;x) = \frac{21-y}{84-y-3x}$$

[14] In particular, this could be the same person.

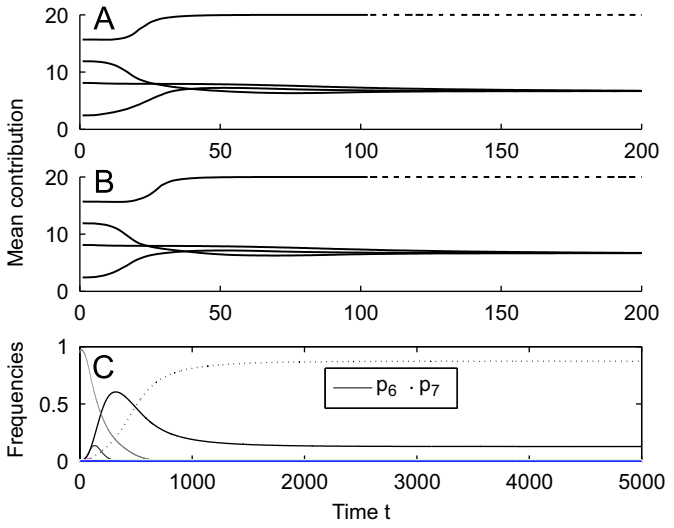

Fig. 4. The replicator equation and its mean-field approximation for IRP (Example 4 with $A=16$). Panel A shows the mean contribution along four trajectories of the replicator equation whose initial distribution is $p_i=0.0025$ for all $i=0,1,\ldots,20$ except for one choice of i. These choices are $p_{16}=0.95$ for the trajectory that has the highest initial mean contribution; $p_{12}=0.95$ for the trajectory with the next highest initial contribution; then $p_8=0.95$; and finally $p_2=0.95$ for the trajectory with the lowest initial contribution. Panel B shows the corresponding results for the mean-field replicator equation. Panel C plots $p_i(t)$ for $i=0,1,\ldots,20$ under the replicator equation with initial distribution $p_i=0.0025$ for all $i=0,1,\ldots,20$ except $p_8=0.95$. Parameters are taken as in Example 4: $n=4;r=1.6;A=16;E=20$.

An interior rest point x is a solution of

$$x^2-20x+K=0$$

where

$$K=\frac{33A}{8}\left(1-\frac{r}{n}\right)^{-1}-21$$

The solutions are

$$x^*_{1,2}=\frac{20\pm\sqrt{400-4K}}{2}=10\pm\sqrt{100-K}$$

Clearly, $x^*_{1,2}\in(0,E)$ if and only if $0\le K<100$ with $x^*_1\ge 10$ and $x^*_2\le 10$.

Since

$$\frac{d}{dx}\left(\frac{dx}{dt}\right)=-\frac{3A}{16}\cdot\frac{2(10-x)}{(21-x)^2(x+1)^2}$$

$x^*_1=10+\sqrt{100-K}$ is unstable and $x^*_2=10-\sqrt{100-K}$ is stable if $0<K<100$. Furthermore, if $0<K<100$, then $x=20$ is convergence stable and $x=0$ is unstable in IRP. If $A=20$ as in Examples 1–3, then $K>100$ and $x=20$ is the only rest point. All trajectories of the canonical equation converge to it as well as all simulations of the replicator equation and its mean-field approximation.

A more interesting case is $A=16$. Then $K=89$, $x^*_1=10+\sqrt{11}\cong 13.32$, and $x^*_2=10-\sqrt{11}\cong 6.68$. The simulations in Fig. 4 (Panels A and B respectively) show that the mean contribution for the replicator equation and its mean-field approximation respectively, all converge to $x=20$ or to $x\cong 6.87$. In Panel C, trajectories are seen to converge to the dimorphic population consisting of most individuals contributing 7 and about 13% contributing 6.[15] This is consistent with the prediction of the canonical equation that has x^*_2 as a convergence stable equilibrium and with the results of Examples 2 and 3. It is also true that the initial points that evolve to a particular mean contribution depend on which evolutionary dynamics is used (in analogy to Example 12 for IP).

[15] This occurs since neither 6 nor 7 is a NE in this IRP game restricted to these two strategies. If x^*_2 is added to the support of the initial distribution whose mean contribution is close-by, simulations show convergence to the monomorphic population at x^*_2.

Table 1
Nash equilibrium, rest points and dynamic stability under the canonical equation in Examples 2–4 where the group size is 4, $E=20$ and $A=20$.

Incentive scheme	Rest point	Nash equilibrium	Dynamic stability
IR	5.25	Yes	Globally stable
IP	0	Yes	Locally stable
	14.75	No	Unstable
	20	Yes	Locally stable
IRP	20	Yes	Globally stable

4. Discussion

From the above analysis, institutional incentives are an effective means to promote cooperation in the public goods game. Furthermore, the NE structure of PGG with incentives (that depend continuously on contribution levels) predicts the eventual contribution level under the standard evolutionary dynamics (i.e. the canonical equation and the (mean-field) replicator equation) of this multi-player game that has a continuous strategy space. In order to obtain stable positive contribution levels, the incentive amount A must be large enough to overcome the individual free-riding advantage in PGG.

The four examples considered in detail show that, when group size is 4, $E=20$ and $A=20$, institutional reward (IR) destabilizes free-riding behavior and cooperation can invade. However, IR is unable to maintain full cooperation (i.e. contributing $E=20$) since the globally stable NE behavior has contribution level near 5.25. On the other hand, cooperation cannot invade an initially free-riding population under institutional punishment (IP) but, once established, full cooperation can be maintained (i.e. 0 and 20 are both NE). Finally, IRP (institutional reward and punishment) combines both desired features of IR and IP leading to the prediction that the population will become fully cooperative in the long run. Table 1 illustrates these results by characterizing stability of the rest points of the canonical equation for all three incentive schemes.

In these examples, the probability an individual receives the reward (or is punished) is a concave function of his contribution level. There are many other possible institutional incentive schemes. For instance, rather than a fixed incentive amount, it may be more realistic to assume that these amounts decrease as the group's average contribution increases (e.g. businesses may feel the costs associated to an incentive scheme are unwarranted when all employees are performing well). It could also be argued that the amounts increase in group average contribution to model situations where a business increases bonuses when it has a good year (due to high performance levels of employees). Another realistic change to the incentive scheme is to reward or punish more than one individual, especially as group size increases. The approach taken in this paper can be generalized to such incentive schemes.

On the other hand, it is also of interest to analyze rational behavior in public goods games (with or without incentives) repeated among the same group of players. Real-life situations related to dilemmas between individual rationality and collective advantage are often of this sort. Experimental results on repeated PGG (Schram, 2000) show that individuals in a group tend to adjust their contribution between rounds to be closer to the group's current average contribution. Such "reactive" strategies

that are beyond the scope of the present model, warrant more analysis in future theoretical research.

Acknowledgments

The authors appreciate the constructive comments from reviewers of the original version. Financial assistance from the Natural Sciences and Engineering Research Council of Canada and The National Basic Program (973) (No. 2007CB109107) of the People's Republic of China is gratefully acknowledged.

References

Bomze, I.M., 1991. Cross entropy minimization in uninvadable states of complex populations. J. Math. Biol. 30, 73–87.
Christiansen, F.B., 1991. On conditions for evolutionary stability for a continuously varying character. Am. Nat. 138, 37–50.
Corazzini, L., Faravelli, M., Stanca, L., 2010. A prize to give for: an experiment on public good funding mechanisms. Econ. J. 120, 944–967.
Cressman, R., 2009. Continuously stable strategies, neighborhood superiority and two-player games with continuous strategy spaces. Int. J. Game Theory 38, 221–247.
Cressman, R., Hofbauer, J., 2005. Measure dynamics on a one-dimensional continuous strategy space: theoretical foundations for adaptive dynamics. Theor. Popul. Biol. 67, 47–59.
Dieckmann, U., Law, R., 1996. The dynamical theory of coevolution: a derivation from stochastic ecological processes. J. Math. Biol. 34, 579–612.
Doebeli, M., Dieckmann, U., 2000. Evolutionary branching and sympatric speciation caused by different types of ecological interactions. Am. Nat. 156, S77–S101.
Fehr, E., Gächter, S., 2000. Cooperation and punishment in public goods experiments. Am. Econ. Rev. 90, 980–994.
Geritz, S.A.H., Kisdi, É, Meszéna, G., Metz, J.A.J., 1998. Evolutionarily singular strategies and the adaptive growth and branching of the evolutionary tree. Evol. Ecol. 12, 35–57.
Hauert, C., Haiden, N., Sigmund, K., 2004. The dynamics of public goods. Discrete Contin. Dyn. B 4, 575–587.
Herrmann, B., Thoni, C., Gächter, S., 2008. Antisocial punishment across societies. Science 319, 1362–1367.
Hofbauer, J., Sigmund, K., 1998. Evolutionary Games and Population Dynamics. Cambridge University Press, Cambridge.
Morgan, J., 2000. Financing public goods by means of lotteries. Rev. Econ. Studies 67, 761–784.
Nakamaru, M., Dieckmann, U., 2009. Runaway selection for cooperation and strict-and-severe punishment. J. Theor. Biol. 257, 1–8.
Oechssler, J., Riedel, F., 2001. Evolutionary dynamics on infinite strategy spaces. Econ. Theory 17, 141–162.
Rand, D.G., Dreber, A., Ellingsen, T., Fudenberg, D., Nowak, M.A., 2009. Positive interactions promote public cooperation. Science 325, 1272–1275.
Sandholm, W., 2010. Population Games and Evolutionary Dynamics. MIT Press, Cambridge, MA.
Schram, A., 2000. Sorting out the seeking: the economics of individual motivations. Public Choice 103, 231–258.
Sigmund, K., 2010. The Calculus of Selfishness. Princeton University Press, Princeton.
Sigmund, K., Hauert, C., Nowak, M.A., 2001. Reward and punishment. Proc. Nat. Acad. Sci. USA 98, 10757–10762.