SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

November 2013, Vol. 56 112201:1–112201:9
doi: 10.1007/s11432-013-4857-y

How cooperation arises from rational players?

MU YiFen* & GUO Lei

Institute of Systems Science, Academy of Mathematics and Systems Science, Chinese Academy of Sciences, Beijing 100190*, China*

Received May 27, 2013; accepted August 4, 2013

Abstract In classical control systems, the plant to be controlled does not have intention to gain its payoff or benefit, which is obviously not the case in various aspects of social and economic systems(or subsystems). In the latter case, competition and cooperation between players who will optimize their own payoffs turn out to be an important feature, and a fundamental problem is how to achieve cooperation from these rational players. In this paper, we present a neat way to lead to cooperation in dynamical Prisoner's Dilemma game. In our scenario, the two players are heterogenous with hierarchical roles as the 'leader' and the 'follower' respectively. It is shown that the system will co-evolve into and stay at the cooperation state if and only if the leader is restricted not to take the dominating strategies. For the special case of 1-step-memory, the optimal strategies for the leader and follower are 'Tit for Tat' and 'ALL C' respectively. In this framework, both the heterogeneity of the players' roles and the multiplicity of time-scales are crucial for cooperation, which are quite natural settings from the view point of control theory. Besides, the boundary for cooperation also turns out to depend on the relative payoffs of the players.

Keywords cooperation, Prisoner's Dilemma, leader-follower game, finite-memory strategy, optimization

Citation Mu Y F, Guo L. How cooperation arises from rational players? Sci China Inf Sci, 2013, 56: 112201(9), doi: 10.1007/s11432-013-4857-y

1 Introduction

Over the past half century, a great deal of research effort have been devoted to adaptive control systems, and much progress has been made in both theory and applications (see, e.g. [1]). In the traditional parametric adaptive control, the controller may be regarded as a single agent acting on the system based on the information or measurement received, and the time-varying parameter process may be regarded as the strategy of another 'agent', save that the action of this 'agent' usually does not depend on the control actions, and that it has no intention to get its 'payoff'.

However, in many systems such as those in social and economic systems, there exist more complex interactions: the controller, which might be called the 'leader' in the system, stands on a position to manage or to control the system by making rules, while the plant, which might be called the 'follower', can respond to these rules. This interaction can be seen from many phenomena in the real world, like the interaction among the law, the government, the enterprises, and the individuals. Different from the

*Corresponding author (email: mu@amss.ac.cn) classical control systems, the follower does have its intension and rationality to optimize its payoff and thus will adapt its action to the leader's action. Then the system can just evolve under the actions of both players. Then, in these cases, competition and cooperation among agents with different positions and roles in the system turn out to be a significant feature, since in many cases, cooperation is good for both the players and the system, but not easy to achieve.

To study this new kind of system, we need to extend the traditional control framework. One feasible and reasonable way is to model the interactions between agents in such systems as games between heterogeneous/hierarchical players, which is clearly different from the research either in classical game theory [2], or in the differential game theory [3]. As a starting point, we derived a non-equilibrium dynamical game between two heterogeneous players called a 'machine' and a 'human' and considered the optimization of the 'human' while the 'machine' was assumed to take a fixed strategy [4]. Following [4], in the current paper, we will further consider the optimization of both players and formulate the problem as a repeated game between a leader and a follower. Then we will find that the cooperation just emerges. This leader-follower way is different from the setting in the existing related literature.

One notable related work was accomplished by Axelrod [5,6]. In the computer tournament Axelrod designed, the simplest 'Tit for Tat' strategy (which just copies the action of its opponent in the last round) won surprisingly. In another simulation ([7], Chapter 3) where the strategies evolved by Generic Algorithm, even better strategies emerged.

Another related direction is to study the cooperation in the population with spacial structures, such as on lattices or networks, which are pretty productive over the past decades [8,9]. Some other researchers have also tried to achieve cooperation by letting the finite automaton play games [10,11], or by introducing an ε-Nash Equilibrium [12] or the 'good' strategy [13].

Although much work [14] has been devoted to the study of emergence and maintenance of cooperation, the framework in this paper is still different from them. In the above mentioned work, the players are assumed to be homogenous on their roles. In our framework, the players possess hierarchical roles (so they must be heterogeneous), which is a quite natural setting from the view-point of control theory.

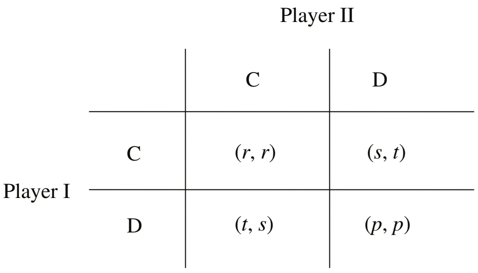
This hierarchical formulation appears to be rare in the existing literature and the most related work seems to be [15]. However, [15] still differs with the current paper in several aspects: the original motivation of [15] was not to achieve cooperation but to reach a state which is better than mutual defection, the follower's strategy was restricted to a special class, the method was enumerating rather than analytical in essence, the win-loss criterion was not introduced and used, and finally neither other games besides the Prisoner's Dilemma game nor the case for general memory length were considered. These will all be considered in the current paper.

In this paper, the repeated Prisoner's Dilemma game will be played between a 'leader' and a 'follower'. Then we will prove that, under some conditions, the rational players will go into and stay at the cooperation state. Moreover, for the 1-step-memory case, the 'Tit for Tat' strategy is best for the leader and it is robust to the payoff parameters and the initial states, while the 'ALL C' strategy is best for the follower. This is a simple, direct and precise way to lead to cooperation and for the first time, the optimality of 'Tit for Tat' is proven analytically. For general k-step-memory strategies, the necessary and sufficient condition is given for cooperation, which shows that the relative payoff for the players is just indispensable. Additionally, we will give two claims that to study cooperation in 2×2 symmetric games, only three games need to be considered and the condition for cooperation holds for all of them. Parts of this paper were presented in [16].

The remainder of the paper is organized as follows. In Section 2, we give the problem formulation and in Section 3, we state the main results and the proofs will be given in Section 4. Section 5 will conclude the paper with some remarks.

2 The problem

The Prisoner's Dilemma (PD) game has been mostly used in the study of evolution of cooperation and can be presented in Figure 1 below, where action 'C' means the player cooperates with the partner, and



		Player II	
		C	D
Player I	C	(r, r)	(s, t)
	D	(t, s)	(p, p)

Figure 1 The payoff matrix of the PD game.

'D' means the player defects the partner. The parameters r, s, t, p denote the payoffs of the players under each action profile. For instance, the payoff profile (t, s) means that under the action profile (D,C), i.e., when Player I's action is D and Player II's action is C, Player I will get a payoff t and Player II will get a payoff s. The parameters satisfy the standard condition [5] as in (1) below.

$$t > r > p > s, \quad r > \frac{t+s}{2}. \tag{1}$$

In the PD game, mutual defection (D,D) is the unique Nash equilibrium, while mutual cooperation (C,C) is better for both. In fact, for any finitely repeated PD game, (D,D) is the unique Nash equilibrium [2]. Much research has been carried out to study how to achieve cooperation. In this paper, following the scenario of [4], we will let the game be played in the leader-follower way and rigorously prove that cooperation can be achieved.

In [4], we derived the dynamical game where the players were heterogenous: one being a 'human' and the other being a 'machine'. Though the two players will take their actions simultaneously as usually assumed, the machine acts according to a fixed k-step memory strategy, while the human can identify the machine's strategy and optimize his own payoff. However, the optimization problem of the 'machine' was not considered in [4].

In this paper, we will further let the 'machine' also optimize its payoff. Thus the machine is like a 'leader' who sets its strategy first, and the human acts like a 'follower' who will react to the leader's strategy. Next, we will formulate this idea precisely and use the notion 'leader' (called a 'she' below) and 'follower' (called a 'he' below) instead of the 'machine' and 'human'.

As in [4], the game is infinitely played at $t = 0, 1, 2, \ldots$. Let $l(t)$ and $f(t)$ denote the actions of the leader and the follower at t. The action 'C' is denoted by 0 and 'D' by 1 and the action set is denoted by $\mathbb{A} = \{0, 1\}$.

Assume the leader's strategy has k-step memory as follows:

$$l(t+1) = g(l(t-k+1), f(t-k+1); \ldots; l(t), f(t)), \tag{2}$$

where g represents a certain mapping. In contrast, the follower's strategy can be any mapping $g_F : H \to \mathbb{A}$, where $H = \bigcup_t H_t$, and $H_t \triangleq \{(l(0), f(0); ...; l(t-1), f(t-1))\}$ is the history of actions up to time $t-1$.

For the analysis to follow, we define the system **state** as

$$s(t) \triangleq \sum_{l=0}^{k-1} \left\{2^{2l+1} \cdot l(t-l) + 2^{2l} \cdot f(t-l)\right\} + 1, \tag{3}$$

which establishes a one-to-one correspondence between the vector set $\{0,1\}^{2k}$ and the integer set $\{1, 2, \ldots, 2^{2k}\}$. Hereinafter, when $s(t) = i$, we will denote it by $s(t) = s_i$, $i = 1, 2, \ldots, 2^{2k}$. Here we use s_i to represent a certain state so as to avoid the possible misunderstanding when only i is used. Then, the leader's action at time $t+1$ under a strategy with k-step-memory can be written as

$$l(t+1) = \sum_{i=1}^{2^{2k}} a_i I_{\{s(t)=s_i\}}(s(t)), \tag{4}$$

where I represents the indicator function

$$I_{\{s(t)=s_i\}}(s(t)) = \begin{cases} 1, & \text{if } s(t) = s_i; \\ 0, & \text{if } s(t) \neq s_i. \end{cases}$$

Then the leader's strategy can be denoted by a vector $A = (a_1, a_2, \ldots, a_{2^{2k}})$. The strategies constitute a set $\mathcal{A}_k = \{\text{k-step-memory-strategy}\}$. Obviously, $\mathcal{A}_k$ increase with k, i.e., $\mathcal{A}_k \subset \mathcal{A}_{k+1}$.

Now, given any initial state $s(0)$, any leader's strategy g as in (2) (or A as in (4)) and any follower's strategy g_F (as defined following (2)), the game will be realized and both the players will get their one-step payoff value $p_L(t)$ and $p_F(t)$. The relative payoffs at t for both players can also be defined as $w_L(t) \triangleq \text{sgn}\{p_L(t) - p_F(t)\}$, and $w_F(t) \triangleq -w_L(t)$, which indicate whether the leader or the follower wins or not at time t.

The overall payoff on the infinite time horizon is defined in the average sense (called the averaged payoff in the following): $P_i \triangleq \overline{\lim}_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} p_i(t),\ i = L, F$. The averaged relative payoff $W_i, i = L, F$ is defined similarly. When the leader and follower choose their strategies A and g_F respectively, their resulting averaged payoffs and relative payoffs will be denoted by $P_i(s(0), A, g_F),\ W_i(s(0), A, g_F),\ i = L, F$, where $s(0)$ is a given initial state. In the case where $s(0)$ is not concerning, we simply denote them by $P_i(A, g_F),\ W_i(A, g_F),\ i = L, F$.

Let the players optimize their averaged payoffs and averaged relative payoffs lexicographically: If there are two strategies which give the player the same averaged payoff, he/she will choose the one which gives a bigger averaged relative payoff. If two strategies give the player the same averaged payoff and the same averaged relative payoff, the player chooses either one. Then by Proposition 1 in next section, this choice will not influence the final result of the game.

By Theorem 2.1 in [4], for any given $A_i \in \mathcal{A}_k$, there exists $g_F = B_i^* \in \mathcal{A}_k$ which is optimal for the follower. Then, neglecting the initial state here, the lexicographical optimization of the players can be expressed as

$$B_i^* = \underset{B\in\mathcal{A}_k}{\text{argmax}}(P_F(A_i, B), W_F(A_i, B)), \quad A_{i^*} = \underset{A_i\in\mathcal{A}_L\subseteq\mathcal{A}_k}{\text{argmax}} (P_L(A_i, B_i^*), W_L(A_i, B_i^*)). \tag{5}$$

By (5), the follower will choose B_i^* so that he can optimize his averaged payoff $P_F(A_i, B)$ and averaged relative payoff $W_F(A_i, B)$ lexicographically knowing the leader's strategy A_i; meanwhile, the leader will choose A_{i^*} so that she can optimize her averaged payoff and averaged relative payoff $P_L(A_i, B_i^*)$, $W_L(A_i, B_i^*)$ lexicographically knowing the follower's optimal choice B_i^*. In (5), $\mathcal{A}_L$ is a subset of $\mathcal{A}_k$. In next section, we will see that in order to induce cooperation, we must have $\mathcal{A}_L \neq \mathcal{A}_k$ for $k \geqslant 2$.

Now, two basic and further questions remain to be addressed: Knowing that the follower will optimally react to her strategy, which strategy should the leader choose in order to optimize her own averaged payoff and relative payoff lexicographically? What will happen to the whole dynamic game system? i.e., What is the solution to (5) and what is the system state under the solution? These will be answered in the following section.

3 Main results

3.1 Cooperation in the leader-follower Prisoner's Dilemma game

First we give the following proposition, from which one can see that if there are two strategies giving the same averaged payoff and the same relative payoff to a player, then he/she can choose any one without influencing the result of the game.

Proposition 1. Given two pairs of strategies of the leader and follower (A_1, B_1) and (A_2, B_2), if the systems under them lead to the same averaged payoff and the same averaged relative payoff for the follower(or for the leader), i.e. $P_F(A_1, B_1) = P_F(A_2, B_2)$ and $W_F(A_1, B_1) = W_F(A_2, B_2)$ (or the respective equations for the leader), then the system will give the same averaged payoff and the same averaged relative payoff for the leader (or for the follower), i.e. $P_L(A_1, B_1) = P_L(A_2, B_2)$ and $W_L(A_1, B_1) = W_L(A_2, B_2)$ (or the respective equations for the follower).

Now, we can answer the questions raised at the end of Section 2 for the simplest case with $k = 1$.

Theorem 1. Consider the leader-follower Prisoner's Dilemma game defined by (5). Let $k = 1$ and $\mathcal{A}_L = \mathcal{A}_1$. Then, the optimal strategy profile $(A_{i^*}, B^*_{i^*})$ exists. Moreover, under $(A_{i^*}, B^*_{i^*})$, the system will reach and stay at the cooperation state (C,C) from any initial state $s(0)$.

Proposition 2. Consider the leader-follower Prisoner's Dilemma game defined by (5). Let $k = 1$ and $\mathcal{A}_L = \mathcal{A}_1$. Then

(i) The 'Tit for Tat' strategy, i.e, $A_T = (0, 1, 0, 1)$ is the solution for A_{i^*} for any initial state $s(0)$ and any parameters (s, p, r, t) satisfying (1). The corresponding optimal strategy $B^*_{i^*}$ for the follower is the 'ALL C' strategy $A_C = (0, 0, 0, 0)$.

(ii) If $(t + p)/2 \geqslant r$, the 'Tit for Tat' strategy is the unique solution for A_{i^*} as in (i).

Remark 1. The optimal strategy profile (TFT, ALL C) ('Tit for Tat' will be abbreviated as TFT below) can be seen as the 'equilibrium' in the leader-follower game, by which the Prisoner's Dilemma game can reach cooperation in a neat way. This is different from the previous understanding of the homogeneous scenario [17], where the 'ALL C' strategy is obviously dominated by 'ALL D'.

Remark 2. Proposition 2 implies the robustness of TFT with respect to the perturbation of payoff parameters and the initial states.

For the case with general k, cooperation cannot be reached in the game defined by (5) if $\mathcal{A}_L = \mathcal{A}_k$. However, the necessary and sufficient condition with respect to $\mathcal{A}_L$ can be given in order to reach cooperation. Before giving the condition, we define two kinds of strategies for the leader.

For simplicity, in the following, denote the leader's strategy by L and the follower's by F. The averaged payoff and averaged relative payoff under the strategy profile (L, F) are denoted by $P_i(L, F)$ and $W_i(L, F)$, $i = L, F$. Given $L \in \mathcal{A}_k$, denote by $F(L) = \operatorname{argmax}_{F \in \mathcal{A}_k}(P_F(L, F), W_F(L, F))$ the optimal strategy for the follower.

Then, a leader's strategy L is called a **cooperating strategy**, if under L and $F(L)$, the leader's averaged payoff satisfies $P_L(L, F(L)) = r$, $W_L(L, F(L)) = 0$ where r is the payoff for the player at the cooperation state (C,C) (see the payoff matrix in Figure 1). The set of all cooperating strategies is denoted by $\mathcal{A}_{cc}$. Note that for any k, the 1-step-memory-strategy 'Tit for Tat' belongs to the set $\mathcal{A}_{cc}$ (by Theorem 1), such that we always have $\mathcal{A}_{cc} \neq \emptyset$.

Then, $\mathcal{A}_{cc}$ defines the cooperation state (C,C): on one hand, given $(L, F(L))$, if the system will reach and stay at the cooperation state (C,C), the leader's average payoff must be $P_L(L, F(L)) = r$, $W_L(L, F(L)) = 0$; on the other hand, if the leader's averaged payoff is $P_L(L, F(L)) = r$, $W_L(L, F(L)) = 0$, then the system must be at the cooperation state (C,C), since by (1), any other cycle of states cannot give the leader such a payoff.

A leader's strategy L is called a **dominating strategy**, if under L and $F(L)$, the leader's averaged payoff satisfies $W_L(L, F(L)) > 0$, $P_L(L, F(L)) \geqslant r$. The set of all dominating strategies is denoted by $\mathcal{A}_d$.

Intuitively, the dominating strategy is preferred by the leader but not by the follower: on one hand, if the leader takes a dominating strategy $L \in \mathcal{A}_d$, then whatever the follower optimizes, the leader's averaged payoff will always be strictly better than at the cooperation state; on the other hand, the follower's averaged payoff will always be $P_F \leqslant r$ (since at any state $p_L + p_F \leqslant 2r$), $W_F < 0$ (since at any state $w_L + w_F = 0$), i.e. the leader dominates the follower and the game now.

Here we note that both $\mathcal{A}_{cc}$ and $\mathcal{A}_d$ are defined relative to $\mathcal{A}_k$. When k varies, both $\mathcal{A}_{cc}$ and $\mathcal{A}_d$ will varies too. Then, we have

Theorem 2. Consider the leader-follower Prisoner's Dilemma game defined by (5). Let k be general and $\mathcal{A}_L \subseteq \mathcal{A}_k$. Then the system will reach cooperation if and only if $\mathcal{A}_{cc} \bigcap \mathcal{A}_L \neq \emptyset$ and $\mathcal{A}_d \bigcap \mathcal{A}_L = \emptyset$.

From Theorem 2, the sets $\mathcal{A}_{cc}$ and $\mathcal{A}_d$ define a boundary between the cooperation and non-cooperation in the system defined by (5). Here we note that to describe $\mathcal{A}_{cc}$ and $\mathcal{A}_d$, both the payoff and relative payoff for the players must be considered, and the relative payoff is just indispensable.

To see this, define four subsets of $\mathcal{A}_k$ as below: $\mathcal{A}_k^w \triangleq \{L \in \mathcal{A}_k : W_L(L, F(L)) < 0\}$, $\hat{\mathcal{A}}_k^w \triangleq \{L \in \mathcal{A}_k : W_L(L, F(L)) \leqslant 0\}$, $\mathcal{A}_k^p \triangleq \{L \in \mathcal{A}_k : P_L(L, F(L)) < r\}$, $\hat{\mathcal{A}}_k^p \triangleq \{L \in \mathcal{A}_k : P_L(L, F(L)) \leqslant r\}$. Then, we have $\mathcal{A}_{cc} = (\hat{\mathcal{A}}_k^w \setminus \mathcal{A}_k^w) \bigcap (\hat{\mathcal{A}}_k^p \setminus \mathcal{A}_k^p)$. Also, define $\mathcal{A}_c \triangleq \mathcal{A}_k \setminus \mathcal{A}_d$, which is the largest set that $\mathcal{A}_L$ can be taken in (5) in order to reach cooperation. Then, $\mathcal{A}_c = \hat{\mathcal{A}}_k^w \bigcup ((\hat{\mathcal{A}}_k^w)^c \bigcap \mathcal{A}_k^p) = \hat{\mathcal{A}}_k^p \setminus ((\hat{\mathcal{A}}_k^w)^c \bigcap (\hat{\mathcal{A}}_k^p \setminus \mathcal{A}_k^p)) = \mathcal{A}_k \setminus ((\hat{\mathcal{A}}_k^w)^c \bigcap (\mathcal{A}_k^p)^c)$. So, neither $\mathcal{A}_{cc}$ nor $\mathcal{A}_d$ can be expressed completely by the payoff or the relative payoff only.

It is interesting to investigate the sets $\mathcal{A}_{cc}$, $\mathcal{A}_d$ in detail. Unfortunately, to describe them clearly is not easy: on one hand, the number of elements in $\mathcal{A}_k$ is $2^{(4^k)}$, which makes searching over $\mathcal{A}_k$ quite hard when k is large; on the other hand, the popular mathematics lose their power here. For $k = 2$, the simulation result shows that $\mathcal{A}_d$ is not big, which seems helpful for cooperation. For $k \geqslant 2$, we can prove that $\mathcal{A}_d$ is nonempty.

Proposition 3. Let $k \geqslant 2$. If the parameters r, s, t, p in (1) also satisfy $(2s+t)/3 > p$ and $(2t+s)/3 > r$, then there always exists a leader's strategy $L =$ '2 Tits for 1 Tat', i.e., $k = 2$, $A = (a_1, a_2, \ldots, a_{16})$, $a_{11} = 0$, $a_i = 1, \forall i \neq 11$, belonging to $\mathcal{A}_d$.

Remark 3. Proposition 3 coincides with Axelrod's simulation result where the strategies were evolved by Generic Algorithm ([7], Chapter 3) and some new strategies performing better than TFT emerged. This fact also reveals the difference between the solution for the leader-follower game and the Nash equilibrium: given a big enough strategy space, the leader do have the advantage over the follower.

3.2 Cooperation in 2×2 symmetric games

There are two other interesting game models: the Snowdrift and the Staghunt game, which can also be used to study the emergence and maintenance of cooperation [18,19]. They can also be presented by the payoff matrix in Figure 1, where for the Snowdrift game, the parameters satisfy $t > r > s > p$, $r \geqslant (t+s)/2$, and for the Staghunt game, $r > t \geqslant p > s$. An interesting question is: are there some other games for studying cooperation between 2 players? The answer is NO.

To see this, consider a general symmetric game in which there are two players and either has two actions 'C' and 'D'. And the payoff matrix is still presented as Figure 1. Then if the cooperation state (C,C) is good for both players but hard to reach, the parameters must satisfy:

(i) $t > s$, i.e. the defector gets more payoff than the cooperator in the profile (C,D);
(ii) $r > s$, i.e. the cooperator gets more payoff against a cooperator than against a defector;
(iii) $r > p$, i.e. mutual cooperation is better than mutual defection for both;
(iv) $t \geqslant p$, i.e. the defector gets no less payoff against a cooperator than against a defector;
(v) $2 \cdot r \geqslant s + t$, i.e. the sum of their payoffs at mutual cooperation is no less than at (C,D) or (D,C).

Now the relations between r and t, s and p are uncertain, which might be:
$r > t \Leftrightarrow$ the best action for one player against C of the opponent is C;
$r < t \Leftrightarrow$ the best action for one player against C of the opponent is D;
$s > p \Leftrightarrow$ the best action for one player against D of the opponent is C;
$s < p \Leftrightarrow$ the best action for one player against D of the opponent is D.

Denoting by (x, y) that x is the best action against C of the opponent, y is the best action against D. Then if $(x, y) = (\mathrm{C}, \mathrm{C})$ (which implies $r > t,\ s > p$), the cooperation can be reached trivially. In contrast, if the cooperation is hard to reach, there must hold $(x, y) = (\mathrm{C}, \mathrm{D})$, or $(x, y) = (\mathrm{D}, \mathrm{C})$, or $(x, y) = (\mathrm{D}, \mathrm{D})$, which correspond to $r > t,\ s < p$, or $r < t,\ s > p$, or $r < t,\ s < p$, i.e. the Staghunt, the Snowdrift, the Prisoner's Dilemma game respectively. So, we have

Claim 1. When considering cooperation in 2×2 symmetric games, it suffices to consider the Prisoner's Dilemma, the Snowdrift and the Staghunt game.

Now, for the Snowdrift and the Staghunt game, when $k = 1$, can the leader-follower game defined by (5) reach cooperation when $L \in \mathcal{A}_1$? It is easy to check that:

(i) For the Snowdrift game, the solution to (5) is (ALL D, ALL C), which will not reach cooperation.

(ii) For the Stag-hunt game, the solution to (5) is (ALL C, ALL C) or (TFT, ALL C) which will reach cooperation.

However, if the leader's strategy set $\mathcal{A}_L$ is restricted as in Theorem 2, cooperation can be reached from the game defined by (5). Thus, we have

Claim 2. For the 2×2 symmetric games, Theorem 2 gives the necessary and sufficient condition for cooperation.

3.3 Some further remarks

In the above sections, we proposed the framework for the game-based control system by modeling it as repeated games between a leader and a follower, and stated the main results that under some necessary and sufficient condition, cooperation can emerge from rational players who optimize their own payoffs.

In reality, the game can be played in the following way:

(a) First, the leader chooses a strategy.

(b) Then, the follower optimizes by choosing his actions at each step, and both players get their payoffs. The leader's strategy will be used for finite but long enough steps (say T_0 steps).

(c) After T_0 steps in step(b), the leader can change to a new strategy. Then the game begins from step(a) again and step(b) will be carried out for another T_0 steps. This process will not be ended until either player will not change his/her strategy any more.

Then, we remark that in this framework, there are two ingredients which we think of as important for cooperation: first, the players are hierarchical on their roles — a 'leader' and a 'follower': the leader will set her strategy first; second, the time scales of the players to improve their strategies are different (this point is hidden): the follower is assumed to be fast and flexible, while the leader is slower when changing her strategy to a better one so that the follower can have enough time (e.g. in [4], this time is at least 7 steps when $k = 1$) to identify the leader's strategy.

Otherwise, if the players are homogenous on their roles, then the repeated Prisoner's Dilemma game($k = 1$) has more than 1 equilibrium while TFT is not an equilibrium strategy. On the other hand, if the players are homogenous on the time scales, the weakness of TFT must be considered [17] too: it cannot correct a mistake of the opponent, i.e. once the action of the opponent is D, TFT will retaliate immediately, leaving no chance for repentance of its opponent and thus lead to a cycle of (D, D) or an oscillation between (C, D) and (D, C), and the system may not reach cooperation.

But with heterogeneity on the roles and the time-scales, TFT is the optimal strategy for the leader and its weakness can be easily overcome.

4 Proofs of the main results

Proof of Proposition 1: Given the players' strategy profile (A_i, B_i), the system will enter a cycle of states. Denote the cycles resulted by (A_1, B_1) and (A_2, B_2) by C_1, C_2. Denote the number of state s_1, s_2, s_3, s_4 (defined for $k = 1$) included in the cycles C_1, C_2 by m_1, m_2, m_3, m_4 and n_1, n_2, n_3, n_4 respectively. Then we have $P_F(C_1) = P_F(C_2),\ W_F(C_1) = W_F(C_2)$, i.e.

$$\frac{m_1 \cdot r + m_2 \cdot t + m_3 \cdot s + m_4 \cdot p}{m_1 + m_2 + m_3 + m_4} = \frac{n_1 \cdot r + n_2 \cdot t + n_3 \cdot s + n_4 \cdot p}{n_1 + n_2 + n_3 + n_4},$$
$$\frac{m_1 \cdot 0 + m_2 \cdot 1 + m_3 \cdot (-1) + m_4 \cdot 0}{m_1 + m_2 + m_3 + m_4} = \frac{n_1 \cdot 0 + n_2 \cdot 1 + n_3 \cdot (-1) + n_4 \cdot 0}{n_1 + n_2 + n_3 + n_4}.$$

Then by some simple calculations, we get

$$\frac{m_1 \cdot r + m_2 \cdot s + m_3 \cdot t + m_4 \cdot p}{m_1 + m_2 + m_3 + m_4} = \frac{n_1 \cdot r + n_2 \cdot s + n_3 \cdot t + n_4 \cdot p}{n_1 + n_2 + n_3 + n_4},$$

$$\frac{m_1\cdot 0+m_2\cdot(-1)+m_3\cdot 1+m_4\cdot 0}{m_1+m_2+m_3+m_4}=\frac{n_1\cdot 0+n_2\cdot(-1)+n_3\cdot 1+n_4\cdot 0}{n_1+n_2+n_3+n_4},$$

i.e. $P_L(C_1)=P_L(C_2),\ W_L(C_1)=W_L(C_2)$. That completes the proof.

Proof of Theorem 1: $k=1$, then $|\mathcal{A}_1|=2^4=16$. So obviously the optimal strategy $(A_i^*,B_{i^*}^*)$ exists by finiteness.

By Theorem 2.2 in [4], for $k=1$, the follower will never lose; thus $\mathcal{A}_d=\emptyset,\mathcal{A}_c=\mathcal{A}_k$. Additionally, the set $\mathcal{A}_{cc}=\{L:P_L=r,W_L=0\}$ is nonempty: the 'Tit for Tat' belongs to it: if $L=$ TFT, then $F(L)=$ ALL C, and the system will enter and stay at the cooperation state where $P_L=r,\ W_L=0$. Then Theorem 1 can be proved immediately from Theorem 2.

Proof of Proposition 2: Denote $A_{i^*}=(a_1^*,a_2^*,a_3^*,a_4^*)$. (i) is obvious.

Now we prove (ii). By Theorem 1, the cycle under the solution of (5) includes only the mutual cooperation state s_1. Thus $a_1^*=0$.

Next, we prove $a_2^*=1$ by contradiction argument. If $a_2^*=0$, the follower who optimizes his averaged payoff will identify this fact and then makes use of it by always acting a D, and then the system will stay at s_2. This is a contradiction.

Now, since A_{i^*} must satisfy (5) for any initial state, the State Transfer Graph (see [4] for the detailed definition) of the strategy A_{i^*} must be strongly connected. Then a_3^* and a_4^* cannot be both 1. So A_{i^*} might be $(0,1,0,1)$, $(0,1,1,0)$ or $(0,1,0,0)$.

If $(t+p)/2\geqslant r$, on the State Transfer Graphs of the strategies $(0,1,1,0)$ and $(0,1,0,0)$, there exists a cycle $\{s_2,s_4\}$ where $P_F(\{s_2,s_4\})\geqslant P_F(\{s_1\})=r$ and $W_F(\{s_2,s_4\})>W_F(\{s_1\})=0$, i.e. B_{i^*} for the follower will not lead to the cooperation state. This contradicts to Theorem 2. This proves the uniqueness of TFT when $(t+p)/2\geqslant r$.

Proof of Theorem 2: In the proof, $P_L(L,F(L))$ and $W_L(L,F(L))$ are abbreviated as P_L and W_L. Under the strategy profile $(L,F(L))$, the system state (and thus the actions of players) will enter and stay in a cycle on which the leader's averaged payoff and averaged relative payoff might be: $[P_L>r,W_L>0]$, $[P_L>r,W_L=0]$, $[P_L>r,W_L<0]$, $[P_L=r,W_L>0]$, $[P_L=r,W_L=0]$, $[P_L=r,W_L<0]$, $[P_L<r,W_L>0]$, $[P_L<r,W_L=0]$, $[P_L<r,W_L<0]$.

First, we prove that the sets $\{L:P_L>r,W_L<0\}$, $\{L:P_L>r,W_L=0\}$, $\{L:P_L=r,W_L<0\}$ are all equal to $\emptyset$, i.e. no leader's strategy gives her the averaged payoff as $[P_L>r,W_L<0]$, $[P_L>r,W_L=0]$ or $[P_L=r,W_L<0]$.

In fact, if $W_L<0$, in the cycle, the number of $(l(t),f(t))=(1,0)$ (which is s_3 when $k=1$) is strictly smaller than the number of $(l(t),f(t))=(0,1)$ (which is s_2 when $k=1$), and vice versa. Since the PD game is symmetric, we have $W_L<0\Leftrightarrow P_L<P_F$ and similarly $W_L=0\Leftrightarrow P_L=P_F$. Additionally, there always is $P_L+P_F\leqslant 2\cdot r$, thus we have $W_L<0\Rightarrow P_L<r$ and $W_L\leqslant 0\Rightarrow P_L\leqslant r$. So no leader's strategy will lead to $[P_L>r,W_L<0]$, $[P_L>r,W_L=0]$, or $[P_L=r,W_L<0]$, i.e. the sets $\{L:P_L>r,W_L<0\}$, $\{L:P_L>r,W_L=0\}$ and $\{L:P_L=r,W_L<0\}$ are all equal to $\emptyset$.

So if and only if $\mathcal{A}_L\bigcap\mathcal{A}_d=\emptyset$, the leader's averaged payoff can be: $[P_L=r,W_L=0]$, $[P_L<r,W_L>0]$, $[P_L<r,W_L=0]$ or $[P_L<r,W_L<0]$, where obviously $[P_L=r,W_L=0]$ is best for the leader and thus $L\in\{L:P_L=r,W_L=0\}=\mathcal{A}_{cc}$ will be chosen if and only if $\mathcal{A}_L\bigcap\mathcal{A}_{cc}$.

Second, if and only if $(L,F(L))$ leads to a cycle with $[P_L=r,W_L=0]$, i.e., $L\in\mathcal{A}_{cc}$, the system is at the cooperation state $(l(t),f(t))=(0,0)$, i.e. (C,C). This is because $P_L=r,\ W_L=0$ implies $P_F=P_L=r$, and only the cooperation state gives them the largest payoff sum $P_L+P_F=2r$. This completes the proof.

Proof of Proposition 3: Proposition 3 can be proved directly from Remark 2.1 in [4]. If the leader takes the strategy '2 Tits for 1 Tat", then if $(2s+t)/3>p$, the optimal cycle for the follower is $\{s_{10},s_7,s_{11}\}$, which corresponds to a cycle of action profiles like

$$\begin{pmatrix} l(t)\\ f(t)\end{pmatrix}=\begin{pmatrix}1\ 1\ 0\ 1\ 1\ 0\ \cdots\\ 0\ 0\ 1\ 0\ 0\ 1\ \cdots\end{pmatrix}.$$

And the leader can get a payoff $(2t+s)/3>r$, i.e. $P_L>r,\ W_L>0$. That completes the proof.

5 Concluding remarks

From a new view point, the controller and the plant in the traditional control framework can be regarded as two 'agents' while the plant has no intention and capability to get its payoff or benefit. This is not the case in the real world. In the real world, the plant and the controller can both optimize their own payoffs. For instance, if we regard the law or policy as a controller while regard the individual as a plant, then both of them will optimize their payoffs which can be well defined. This kind of interactions may be modeled as games between hierarchical players and the systems can be seen as game-based control systems. In these systems, competition and cooperation between the players are a focus in research, and how to achieve cooperation is a fundamental problem.

In this paper, we have studied the dynamical Prisoner's Dilemma game between a leader and a follower, which are different both in their roles and time-scales. The leader's strategy is restricted to be in a given set and both the players would optimize their averaged payoffs and averaged relative payoffs lexicographically. Then we can prove that if we want the system to evolve into and stay at the cooperation state, then the leader are not allowed to take the dominating strategies and a necessary and sufficient condition for cooperation can be built. This appears to be a neat and rigorous way to achieve cooperation.

Of course, there are many interesting problems in this framework worthy of further investigation. For example: how can the problem be properly formulated when there are many leaders or many followers? Can the leader-follower structure still promote cooperation then? What will happen if the follower does not optimize but sub-optimize his own payoff? These appear to be more complicated problems and belong to further investigation.

Acknowledgements

The work was supported by National Natural Science Foundation of China (Grant No. 60821091).

References

1 Guo L. Adaptive systems theory: some basic concepts, methods and results. J Syst Sci Complex, 2003, 16: 293–306

2 Fudenberg D, Tirole J. Game Theory. Cambridge: MIT Press, 1991

3 Isaacs R. Differential Games: a Mathematical Theory with Applications to Warfare and Pursuit, Control and Optimization. New York: Wiley, 1965

4 Mu Y, Guo L. Optimization and identification in nonequilibrium dynamical games. In: Proceedings of the 48th IEEE Conference on Decision and Control, Shanghai, 2009. 5750–5755

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

6 Axelrod R. The Complexity of Cooperation: Agent-Based Models of Competition and Collaboration. New Jersey: Princeton University Press, 1997

7 Davis L. Genetic Algorithms and Simulated Annealing. London: Morgan Kaufman Publishers, Inc., 1987

8 Nowak M A, Bonhoeffer S, May R M. Spatial games and the maintenance of cooperation. Proc Natl Acad Sci USA, 1994, 91: 4877–4881

9 Ohtsuki H, Hauert C, Lieberman E, et al. A simple rule for the evolution of cooperation on graphs and social networks. Nature, 2006, 441: 502–505

10 Rubinstein A. Finite automata play the repeated Prisoner's Dilemma. J Econ Theor, 1986, 39: 83–96

11 Neyman A, Okada D. Two-person repeated games with finite automata. Int J Game Theory, 2000, 29: 309–325

12 Radner R. Can bounded rationality resolve the Prisoner's Dilemma. In: Mas-Colell A, Hildenbrand W, eds. Essays in Honor of Gerard Debreu. Amsterdam: North-Holland, 1986. 387–399

13 Smale S. The Prisoner's Dilemma and synamical systems asociated to noncooperative games. Econometrica, 1980, 48: 1617–1634

14 Nowak M A. Five rules for the evolution of cooperation. Science, 2006, 314: 1560–1563

15 Kleimenov A F, Semenishchev A A. Repeated Prisoner's Dilemma: Stackelberg solution with finite memory. In: Proceedings of the 11th IFAC workshop of Control Applications of Optimization, St. Petersburg, 2000, 2: 567–572

16 Mu Y, Guo L. How cooperation arises from rational players? In: Proceedings of the 49th IEEE Conference on Decision and Control, Atlanta, 2010. 6149–6154

17 Imhofa L A, Fudenberg D, Nowak M A. Tit-for-tat or win-stay, lose-shift? J Theor Biol, 2007, 247: 574–580

18 Souzaa M O, Pachecob J M, Santosc F C. Evolution of cooperation under N-person snowdrift games. J Theor Biol, 2009, 260: 581–588

19 Skyrms B. The Stag Hunt and the Evolution of Social Structure. Cambridge: Cambridge University Press, 2004