Evolutionary Game Dynamics of Multi-agent Cooperation Driven by Self-learning

Jinming Du, Bin Wu, and Long Wang
Center for Systems and Control, College of Engineering, Peking University, Beijing 100871, China

***Abstract*—Multi-agent cooperation problem is a fundamental issue in the coordination control field. Individuals achieve a common task through association with others or division of labor. Evolutionary game dynamics offers a basic framework to investigate how agents self-adaptively switch their strategies in accordance with various targets, and also the evolution of their behaviors. In this paper, we analytically study the strategy evolution in a multiple player game model driven by self-learning. Self-learning dynamics is of importance for agent strategy updating yet seldom analytically addressed before. It is based on self-evaluation, which applies to distributed control. We focus on the abundance of different strategies (behaviors of agents) and their oscillation (frequency of behavior switching). We arrive at the condition under which a strategy is more abundant over the other under weak selection limit. Such condition holds for any finite population size of $N \geq 3$, thus it fits for the systems with finite agents, which has notable advantage over that of pairwise comparison process. At certain states of evolutionary stable state, there exists "ping-pong effect" with stable frequency, which is not affected by aspirations. Our results indicate that self-learning dynamics of multi-player games has special characters. Compared with pairwise comparison dynamics and Moran process, it shows different effect on strategy evolution, such as promoting cooperation in collective risk games with large threshold.**

I. Introduction

Cooperation phenomenon among multiple individuals are omnipresent in complex systems. It has been widely investigated during the last decades. The emergence and persistence of cooperation among individuals, whatever human beings, other creatures, or agents in complex systems are all focused by researchers [1]–[6]. Agents achieve a common task through association with others or division of labor. They self-adaptively switch their strategies in accordance with various targets. Different theories have been proposed to explain why agents help others at a cost to themselves. Evolutionary game dynamics, since proposed by John Maynard Smith and George Price in 1970s [7]–[10], has been applied in various areas, from biology to social economics. It has been shown as a powerful framework to investigate the evolution of interactive behaviors.

Studies on evolutionary game dynamics involve two categories. On the one hand, such framework has been widely used to depict the frequency-dependent genetic evolution of microbial communities or other animal populations, such as the Moran process [10] and the Wright-Fisher process [11]. On the other hand, evolutionary game dynamics has been used to interpret culture evolution in humans, which contains pairwise comparison process [1], [12], [13] and a kind of non-imitation learning process, namely self-learning process [14]–[16].

It is shown that the dynamics of pairwise comparison process and the Moran process share similar properties [13], [17]. For example, in a well-mixed population of finite size N, consider a game between two strategies, A and B. For 2×2 games, if two A players interact, both get payoff a_1. If A interacts with B, then A gets a_0 and B gets b_1. If two B players interact, both get b_0. These interactions are represented by payoff matrix

$$\begin{array}{c} \\ A \\ B \end{array} \begin{array}{c} \begin{array}{cc} A & B \end{array} \\ \begin{pmatrix} a_1 & a_0 \\ b_1 & b_0 \end{pmatrix} \end{array}. \tag{1}$$

It has been found that A is more abundant than B if $a_1(N-2) + a_0 N > b_1 N + b_0(N-2)$ for low mutation rate and under weak selection [1], [18]. Then, Antal et al. extend this result to arbitrary mutation rates and any intensities of selection [19]. They found that the condition holds for a large variety of stochastic selection processes provided the transition probabilities depends only on the payoff differences. For general two strategy d player game, the above condition can be generalized to $\sum_{k=0}^{d-1}(Na_k - a_{d-1}) > \sum_{k=0}^{d-1}(Nb_k - b_0)$. This is valid for imitation-exploration dynamics for any selection intensity. In particular, when the population size is large enough, it reduces to $\sum_{k=0}^{d-1}(a_k - b_k) > 0$.

Compared with such well studied processes, the dynamics of self-learning process has rarely been analytically investigated, though it has been intensively investigated on social networks via simulations [14]–[16]. Self-learning process has some advantage over pairwise comparison process. On the one hand, extinct from pairwise comparison process, it is a low-information setting. Players can evaluate and make decision of their strategies based on the difference between its own payoff and the aspiration, while for the pairwise comparison rule, to update its own strategy, individuals are supposed to know the payoff of another one. On the other hand, players are more likely to adopt strategies that are popular and/or successful in the pairwise comparison process. Under this mechanism, once the deleterious strategy is inhibited to quite a low extent, it will be hard for such focal strategy to exist, flourish, or dominate. Whereas, for self-learning dynamics, even in a homogenous population, there is a positive probability that an individual can switch to some other strategies due to the unsatisfied payoff it obtains. This facilities the system to escape from the "absorbing" states.

Besides, some recent researches find that there exists a

978-1-4673-5769-2/13/$31.00 ©2013 IEEE

vibration of cooperation level for some payoff aspirations after evolutionary process. Such interesting phenomenon was regarded as the "ping-pong effect" [14], [15]. It describes a system switching back and forth between two different states. It also indicates the behavior switching of agents in the system. It has practical meanings in the evolutionary games, especially when the population is in a globally defective state. Cooperative behaviors could be partly retrieved by it. Then the system may escape from the defective trap.

Thus we ask under what condition one strategy is more abundant than the other? Is the condition consistent with that of the Moran process with mutations and the pairwise comparison process with exploration? If not, given a multi-player game, which process is more likely to make a specified strategy more abundant? To this end, we study a self-learning model for the general two strategy multi-player games. Since most of the games take place within a group of individuals [20]–[22]. We consider two aspects: one is the evolution of strategies, especially the abundance of strategies; the other is the frequency or cycle of the strategy switch in evolutionary process. In particular, we explore this through the method of spectrum analysis in statistical signal processing.

The rest of the paper is organized as follows: In Section II, we propose a multi-player self-learning game model. In Section III, we analytically study the evolutionary process of the strategies in the model. In Section IV, We show the results of different aspects. In Section V, the conclusions are drawn.

II. Model

A. Example: Flight formation

The organized flight of flying subjects is easily observed, yet challenging to study, phenomena [23]–[25]. Both migratory birds, such as flying-geese, and fighting-planes, unmanned aerial vehicles (UAVs), and etc., always fly together in formation, such as a V formation. The scientific questions about these groups usually concerned with how synchrony is achieved. Early researchers were primarily biologists, but more recently aeronautical engineers, mathematicians, computer scientists and, currently, researchers in control theory have been attracted to the study of organized flight.

Scientists [26] believe that birds flying in this way save a significant amount of energy, for the reason that the shape of the formation reduces the drag force. Other studies have estimated that a flock of 25 birds in formation can fly as much as 70% further than a solo bird using the same amount of energy. However, not all birds benefit equally. The birds in the middle of each line of V experience less drag than either the lead bird or the bird at the end of each line. The lead bird has to work the hardest since it flies into undisturbed air. The upwash it creates improves the aerodynamics of the two behind it, and these two further improve conditions for the next two birds in line. When the lead bird tires, it will drop out of the lead position. Another bird from further back will rapidly move forward to take the leading position and maintain the formation. The two birds in the furthest trailing positions also tire more rapidly and should be rotated frequently. This cyclical rearrangement gives more birds the responsibility of being the leader, as well as a chance to enjoy the maximum benefits of being in the middle of the formation.

What we are focused on are how many agents become cooperators (leaders) during flying; and how often are agents as cooperators (leaders) by turn. Based on these, we interpret the flight formation behavior from a game perspective. Since the agents in the formation benefit unequally. The agents at some certain positions consume more, while others save more energy. If we deem the former behavior as cooperation, and the latter behavior as defection. Thus, the agents play normal form games (e.g. collective risk game) with everyone else in the system. The strategy switch (position change) is determined by each agent's fitness and destination (aspiration). Agents own their payoffs corresponding to their strategies (roles). When the payoff of agent is inferior to its aspiration, its strategy (role) may be altered. During whole flying process, the system unceasingly change formation. The number of cooperators in the population increase or decrease correspondingly. Soon after that, the population will revert back to the state it was. Interestingly, this process shows a periodic phenomenon, which is similar to "ping-pong effect".

B. Game model

We consider an evolutionary game in a well-mixed population with finite size N. In the game, players have two strategies, A and B. Every d individuals interact simultaneously to get their payoffs, i.e., they are in a two strategy d player game. Denote a_i and b_i as the payoffs of a strategy A and B player obtains, respectively, when facing i other A individuals within the group of size d, where i ranges from 0 to $d-1$.

$$\begin{array}{lcccccc} \text{Opposing } A & d-1 & d-2 & \dots & k & \dots & 0 \\ A & \multicolumn{6}{c}{\multirow{2}{*}{$\begin{pmatrix} a_{d-1} & a_{d-2} & \dots & a_k & \dots & a_0 \\ b_{d-1} & b_{d-2} & \dots & b_k & \dots & b_0 \end{pmatrix}.$}} \\ B & & & & & & \end{array} \quad (2)$$

In a population of size N with i A players, the probability of choosing a group that consists of k A players and $d-1-k$ B players is given by a hypergeometric distribution [27]. The probability can be given by $\frac{C_{i-1}^{k}C_{N-i}^{d-1-k}}{C_{N-1}^{d-1}}$, where $d \leq i$. The symbol C_n^k denotes combinatorial notation, which is the number of ways to choose a k element subset from an n element set. Thus, the average payoffs for A and B are

$$\pi_A(i) = \sum_{k=0}^{d-1} \frac{C_{i-1}^{k}C_{N-i}^{d-1-k}}{C_{N-1}^{d-1}} a_k, \quad (3)$$

$$\pi_B(i) = \sum_{k=0}^{d-1} \frac{C_{i}^{k}C_{N-i-1}^{d-1-k}}{C_{N-1}^{d-1}} b_k. \quad (4)$$

The strategy update rule, which plays an important role in the evolution of strategies [28], is as follows. Initially, each individual is randomly assigned a strategy between A and B. In our model, we introduce a parameter α as the aspiration level of the players. Aspiration level provides the benchmark used to evaluate how greedy an individual is. The larger the aspiration is, the greedy the focal individual is. We randomly choose an individual, namely x, from the entire population of size N. The payoff it obtains is denoted as π_x. It changes its current strategy to the opposite one with a probability depending on the difference between the aspiration level and its payoff. The wider the gap between the aspiration and the payoff is, the more the switching possibility is. The probability function is

$$\frac{1}{1+e^{-\omega(\alpha-\pi_x)}}, \tag{5}$$

where $\omega > 0$ denotes the selection intensity, measuring how the decision depends on the difference between the aspiration level and the payoff. For $\omega \to 0$, the difference has little impact on the decision. Individuals switch strategies almost randomly. This implies that individuals are irrational and do not know what they are doing. For $\omega \to \infty$, the difference determines all. In this case, individuals are purely rational: once they are not satisfied with their payoffs, they will change their strategies. This is similar to the previous "win-stay lose-shift" strategy [29], [30]. In this paper, we concentrate on the issues under weak selection limit, which has a long standing history in population genetics and now is introduced to social learning. Recent experiment results suggest the imitation intensity, with which we adjust our strategies, is small [28].

III. Analysis

By the update rule we proposed, each time step, the number of strategy A, i.e., i, can only increase by one, stay the same or decrease by one. When the number of strategy A increases by one, two subsequent events happen: first, a B strategy individual is selected from the population, then it is unsatisfied with the payoff it obtains and switches to strategy A. Similar process holds for the number of strategy B. Therefore the probability that the number of A individuals changes from i to $i \pm 1$ in one time step is

$$T_i^+ = \frac{N-i}{N}\frac{1}{1+e^{-\omega[\alpha-\pi_D(i)]}}, \tag{6}$$

$$T_i^- = \frac{i}{N}\frac{1}{1+e^{-\omega[\alpha-\pi_C(i)]}}, \tag{7}$$

while $T_i^0 = 1 - T_i^+ - T_i^-$.

This Markov process is one dimensional birth-death process with reflecting boundaries, thus it satisfies the detailed balance condition [31], [32]

$$\psi_{j-1}T_{j-1}^+ = \psi_j T_j^- \quad \text{for } 1 \leq j \leq N. \tag{8}$$

where $(\psi_0, \psi_1, \cdots, \psi_j, \cdots, \psi_N)$ is the stationary distribution of the Markov chain. Here ψ_j is the probability that the average proportion of time that the system is in state j. Since $\sum_{j=0}^N \psi_j = 1$, according to the above recursion formula (8), the stationary distribution is given by

$$\psi_j = \frac{\frac{\prod_{i=0}^{j-1} T_i^+}{\prod_{i=1}^{j} T_i^-}}{1+\sum_{k=0}^{N-1}\frac{\prod_{i=0}^{k} T_i^+}{\prod_{i=1}^{k+1} T_i^-}} \quad \text{for } 1 \leq j \leq N. \tag{9}$$

We basically study the average abundance of cooperation strategy under weak selection, $\langle X_A(\omega)\rangle$, which can be calculated as follows

$$\langle X_A(\omega)\rangle = \sum_{j=0}^{N}\frac{j}{N}\psi_j(\omega). \tag{10}$$

The stationary distribution $\psi_j(\omega)$ can be expanded approximately by

$$\psi_j(\omega) \approx \psi_j(\omega)\big|_{\omega=0} + \left(\frac{d}{d\omega}\psi_j(\omega)\bigg|_{\omega=0}\right)\omega, \tag{11}$$

where

$$\psi_j(\omega)\big|_{\omega=0} = \psi_j(0) = \frac{C_N^j}{2^N}. \tag{12}$$

And we have

$$\begin{aligned}\frac{d}{d\omega}\psi_j(\omega) =& \frac{(\frac{\prod_{i=0}^{j-1} T_i^+}{\prod_{i=1}^{j} T_i^-})'(1+\sum_{k=0}^{N-1}\frac{\prod_{i=0}^{k} T_i^+}{\prod_{i=1}^{k+1} T_i^-})}{(1+\sum_{k=0}^{N-1}\frac{\prod_{i=0}^{k} T_i^+}{\prod_{i=1}^{k+1} T_i^-})^2}\\ &-\frac{(\frac{\prod_{i=0}^{j-1} T_i^+}{\prod_{i=1}^{j} T_i^-})(1+\sum_{k=0}^{N-1}\frac{\prod_{i=0}^{k} T_i^+}{\prod_{i=1}^{k+1} T_i^-})'}{(1+\sum_{k=0}^{N-1}\frac{\prod_{i=0}^{k} T_i^+}{\prod_{i=1}^{k+1} T_i^-})^2},\end{aligned} \tag{13}$$

where,

$$(T_i^+)' = \frac{N-i}{N}\frac{\{e^{-\omega[\alpha-\pi_B(i)]}\}[\alpha-\pi_B(i)]}{\{1+e^{-\omega[\alpha-\pi_B(i)]}\}^2}, \tag{14}$$

$$(T_i^-)' = \frac{i}{N}\frac{\{e^{-\omega[\alpha-\pi_A(i)]}\}[\alpha-\pi_A(i)]}{\{1+e^{-\omega[\alpha-\pi_A(i)]}\}^2}. \tag{15}$$

Since $\omega \to 0$,

$$(T_i^+)'\big|_{\omega=0} = \frac{N-i}{4N}[\alpha-\pi_B(i)], \tag{16}$$

$$(T_i^-)'\big|_{\omega=0} = \frac{i}{4N}[\alpha-\pi_A(i)], \tag{17}$$

$$(\prod_{i=0}^{j-1} T_i^+)\bigg|_{\omega=0} = \prod_{i=0}^{j-1}\frac{N-i}{2N} = \frac{N!}{(N-j)!(2N)^j}, \tag{18}$$

$$(\prod_{i=1}^{j} T_i^-)\bigg|_{\omega=0} = \prod_{i=1}^{j}\frac{i}{2N} = \frac{j!}{(2N)^j}, \tag{19}$$

$$[\sum_{i=0}^{j-1}(T_i^+)'(\prod_{k=0,k\neq i}^{j-1} T_k^+)]\bigg|_{\omega=0} = \frac{N!\sum_{i=0}^{j-1}[\alpha-\pi_B(i)]}{2(N-j)!(2N)^j}, \tag{20}$$

$$[\sum_{i=1}^{j}(T_i^-)'(\prod_{k=1,k\neq i}^{j} T_k^-)]\bigg|_{\omega=0} = \frac{j!\sum_{i=1}^{j}[\alpha-\pi_A(i)]}{2(2N)^j}. \tag{21}$$

Then, we have

$$\begin{aligned}\frac{d}{d\omega}\psi_j(\omega)\bigg|_{\omega=0} =& \frac{C_N^j}{2^{(2N+1)}}2^N\sum_{k=1}^{j}[\pi_A(k)-\pi_B(k-1)]\\ &-\frac{C_N^j}{2^{(2N+1)}}\sum_{k=1}^{N}C_N^k\sum_{i=1}^{k}[\pi_A(i)-\pi_B(i-1)].\end{aligned} \tag{22}$$

where $\pi_A(i)$ and $\pi_B(i-1)$ are shown in (3) and (4), hence,

$$\pi_A(i) - \pi_B(i-1) = \sum_{k=0}^{d-1} \frac{C_{i-1}^{k} C_{N-i}^{d-1-k}}{C_{N-1}^{d-1}} (a_k - b_k). \qquad (23)$$

For $\omega \to 0$, the first order estimation of average abundance is independent of α. However, up to the second order, the expansion of $\psi_j(\omega)$ will include terms containing polynomial $[\pi_A(i) - \alpha]$, and etc., thus the exact stationary distribution is dependent on α.

Based on this, we deduce the criterion of $\langle X_A(\omega) \rangle > \frac{1}{2}$ for general multi-player game. Since we have the relation

$$\sum_{j=0}^{N} \frac{j}{N} \psi_j(0) = \sum_{j=0}^{N} \frac{j}{N} \frac{C_N^j}{2^N} = \frac{1}{2}. \qquad (24)$$

Considering the approximate expansion (11), the criterion is:

$$\sum_{j=0}^{N} \frac{j}{N} \Big(\frac{d}{d\omega} \psi_j(\omega) \Big|_{\omega=0} \Big) \omega > 0. \qquad (25)$$

Inserting (22) and (23), the criterion equals to

$$\frac{\omega}{4N(2^N)} \Big[\sum_{j=1}^{N} (2j - N) C_N^j \sum_{i=1}^{j} \sum_{k=0}^{d-1} \frac{C_{i-1}^{k} C_{N-i}^{d-1-k}}{C_{N-1}^{d-1}} (a_k - b_k) \Big] > 0. \qquad (26)$$

We can prove that the above inequality leads to a general criterion as follows

$$\frac{\omega}{4(2^d)} \sum_{k=0}^{d-1} [C_{d-1}^{k} (a_k - b_k)] > 0. \qquad (27)$$

Since such condition should hold for any choice of $(a_k - b_k)$s, and with the identity $\sum_{j=1}^{N} \sum_{i=1}^{j} = \sum_{i=1}^{N} \sum_{j=i}^{N}$, we only need to demonstrate a simplified equivalent condition as

$$\sum_{i=1}^{N} C_{i-1}^{k} C_{N-i}^{d-1-k} \sum_{j=i}^{N} (2j - N) C_N^j = 2^{N-d} N C_{N-1}^{d-1} C_{d-1}^{k}. \qquad (28)$$

This can be easily proved through mathematical induction. Therefore, we have that $\langle X_A(\omega) \rangle > \frac{1}{2}$, or strategy A is favored by selection when

$$\sum_{k=0}^{d-1} \left[C_{d-1}^{k} (a_k - b_k) \right] > 0. \qquad (29)$$

This condition holds for any two strategy d player games ($d \geq 2$) in the whole population of $N \geq 3$. Consider a two-player game ($d = 2$), the above criterion reduces to $a_1 + a_0 > b_1 + b_0$, which can be deemed as risk dominance. This condition is equivalent to $x^* < \frac{1}{2}$, where x^* is the internal equilibrium of the replicator equation [1], [9]. In particular, for two strategies which are the best replies to themselves, we find that selection can favor A replacing B for small ω, if the initial frequency of A exceeds the invasion barrier $x^* = \frac{1}{2}$. Above results are independent of population size N. Hence, it is not necessary for the constraint limit of large population size, which is widely used in former studies.

IV. Results and Discussion

A. Theoretical Results

Most previous studies on self-learning mechanism is investigated via simulation, however analytical research is lacking. We explore the evolutionary dynamics of strategies driven by aspiration in a well-mixed population of finite size. We study the strategy abundance of self-learning process, which is not trivial, since the update rule of strategies is not dependent on the payoff differences between strategies. Therefore, such process does not fulfill the requirements of birth-death processes mentioned in previous papers [19]. The criterion therein can not be applied any longer. Besides, different from the cases in former works, the self-learning process has no absorbing state even without the introduction of exploration or mutation. Thus we can no longer compare the strategies through calculating fixation probabilities as former papers [1], [19]. We utilize the stationary distribution to analyze the abundance of strategies.

When both A and B are Nash equilibria, then it comes to the problem of equilibrium selection, or "trembling hand" [33]. In deterministic replicator dynamics of infinite populations, such 2×2 game admits an unstable equilibrium given by $x^* = \frac{b_0 - a_0}{a_1 - a_0 - b_1 + b_0}$. If the initial frequency of A is less than this value, then it will be eliminated by natural selection. A can only replace B if its initial frequency exceeds this invasion barrier. The same evolutionary dynamics holds for B competing with other cooperative strategies. Other researches [1], [19] studied evolutionary game dynamics in finite populations. It has been shown that A is more abundant than B, if $a_1(N-2) + a_0 N > b_1 N + b_0(N-2)$.

In this paper, we study the self-learning process in finite populations. We derive the basic condition that allows selection for A replacing B. The exact expression for the solution of Nth-order polynomials is complicated, however, the following surprisingly simple result holds. For any $N \geq 3$ and sufficiently weak selection, we have gotten a criterion of $\langle X_A(\omega) \rangle > \frac{1}{2}$ for general multi-player game, namely $\sum_{k=0}^{d-1} [C_{d-1}^{k} (a_k - b_k)] > 0$. Thus selection favours A replacing B if such condition satisfies. In such case, the first order estimation of average abundance is independent of α, which implies that pairwise interaction is the effective intrinsic of self-learning dynamics to some approximate extent. While the higher orders of the expansion is not independent of aspiration level α, but more complicated. This is the special character of self-learning dynamics. In particular, the special case for two players ($d = 2$) of this condition is $a_1 + a_0 - b_1 - b_0 > 0$.

Multi-player games offer a more general framework of studying interactions among agents in the system. In previous works, either imitation dynamics or the Moran process as well as its variation [22], it is the sum of different strategies' payoff differences accounts for whether a strategy can dominate, $\sum_{k=0}^{d-1} (a_k - b_k) > 0$. While in our model, the criterion for self-learning dynamics has an additive weight before the differences of payoffs, $\sum_{k=0}^{d-1} [C_{d-1}^{k} (a_k - b_k)] > 0$. Hence the numbers of different strategy holders become more important, which determines the corresponding weights of payoff differences. Thus, the contribution of $a_k - b_k$ for the dominance of a focal strategy varies for different k. Since the middle k offers the biggest weight, when the numbers of two strategies are almost the same, such payoff difference has more impact. This phenomenon leads to the result that the items in the payoff matrix are no longer equivalent. The items at some certain positions appear to be more critical.

To compare the different effect on cooperation of self-learning dynamics with imitation dynamics, we study typical multi-player game, collective risk game. In collective risk games, players collect public goods through cooperators' contribution. Such amount will be magnified by a gain factor, r, and then be equally distributed to everyone whatever his or her strategy is. While there exists a threshold, which requires at least m players to cooperate in order that individuals can get their payoffs, otherwise, everyone's payoff is zero. The payoff matrix is shown in (30)

$$\begin{array}{c} \begin{array}{ccccccc} d-1 & ... & k & ... & m-1 & ... & 0 \end{array} \\ \left(\begin{array}{ccccccc} \frac{drc}{d} & ... & \frac{(k+1)rc}{d} & ... & \frac{mrc}{d} & ... & 0 \\ \frac{(d-1)rc}{d}+c & ... & \frac{krc}{d}+c & ... & 0 & ... & 0 \end{array}\right). \end{array} \quad (30)$$

Therefore, $\sum_{k=0}^{d-1}[C_{d-1}^{k}(a_k - b_k)] = \sum_{k=m}^{d-1} C_{d-1}^{k}(\frac{rc}{d} - c) + C_{d-1}^{m-1}\frac{mrc}{d}$, while $\sum_{k=0}^{d-1}(a_k - b_k) = c(r + m - d)$. Thus the criterion of self-learning dynamics can be written as $r > \frac{d\sum_{k=m}^{d-1} C_{d-1}^{k}}{\sum_{k=m}^{d-1} C_{d-1}^{k} + mC_{d-1}^{m-1}}$, whereas the latter leads to $r > d - m$ For sufficiently large threshold m, the criterion condition o imitation dynamics is more strict than self-learning dynamics Namely, it requires more reward of contribution (bigger r for imitation dynamics to promote cooperation. Therefore self-learning model is more likely to promote cooperation i collective risk games with high threshold. Whereas, for the games with small threshold, where the case is almost like that of public goods games, the conclusion may be reverse.

B. Simulation Results

Evolutionary results are as follows. Aspiration level α has a significant effect on the cooperation level (see Fig. 1). Smaller α is in favor of cooperators. With the increase of α, such promotion effect decreases, until α rises to a value between the payoffs of cooperators and defectors. That's a balance. After that defectors dominate, until the aspiration is not available to reach. Then the dynamics of each strategy is mainly driven by random drift. Thus each of the two is around $1/2$ of the whole population. Selection intensity also affects the results. Under weak selection, the change of α will not evidently affect the cooperation level. With the increase of selection intensity, the cooperation level becomes more and more sensitive to the increasing aspiration.

C. Spectral Analysis

There exists coherent vibration in the abundance of strategies during their evolutionary process, which was regarded as "ping-pong effect" (see Fig. 2). It reflects the fact that players are drifting with the tide. In order to obtain the maximized benefits, they switch strategies continually. We can utilize the methods of signal processing to study the period or frequency [34].

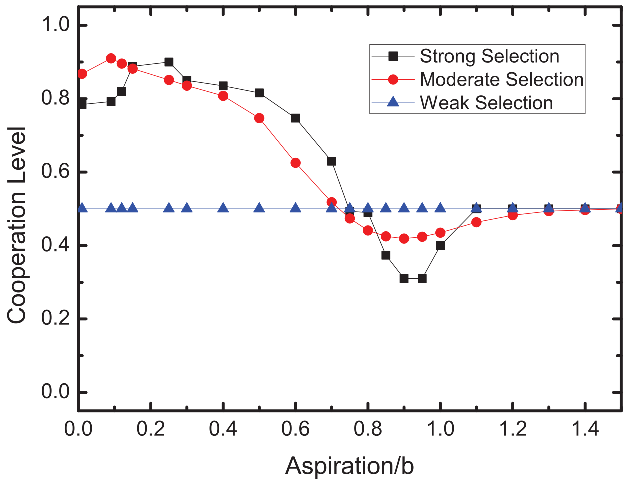
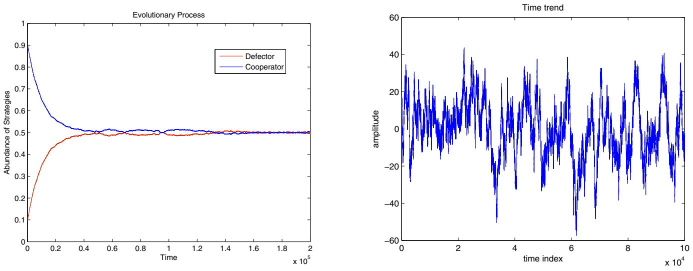


Fig. 1: **Cooperation level with respect to Aspiration.** For different selection intensities, average cooperation levels are shown. Under weak selection, the cooperation level shows little relation with α. Under strong selection, it decreases after increase, then a valley may exist when α increases.



(a) Evolutionary process

(b) "ping-pong" effect

Fig. 2: **Evolutionary process of cooperative level.** For arbitrary initial state, population will end up with evolutionary stable state. An oscillation phenomenon emerges after evolutionary stable.

We take the stable data during evolutionary process as sampling signal (see Fig. 2b). Since the evolution of strategies is a Markov process, which is non-stationary process. Thus, we can treat such sampling signal as cyclo-stationary signal [35]. Since analysis of such signal has strict analytical basis [36]. After removing the noise effect from the collected data, the effective signal can be used to calculate its power spectrum. We utilize Wavelet Transform, which is appropriate for both stationary and non-stationary signals.

The power spectra are shown in Fig. 3. We can find there exists a low-frequency peak in the power spectrum. That means a stable period. which is independent on time trend. For different aspirations, the peak is identical. That is to say, whatever the aspiration is, the period of "ping-pong effect" is unchanged, which means frequency stability.

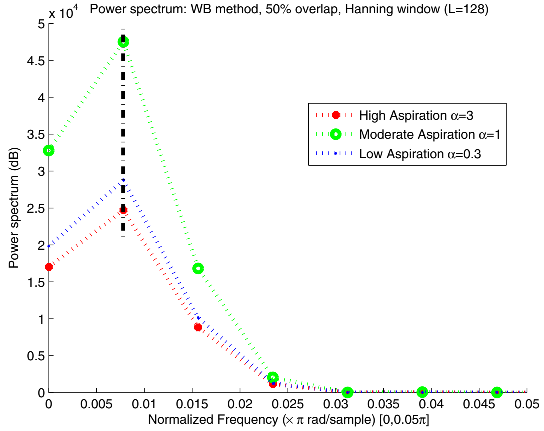


Fig. 3: **Power Spectra for different aspirations.** Comparing the spectra of different α, we find an interesting result that the peaks are obtained at identical frequency, whatever α is.

V. Conclusion

In a well-mixed population of finite size, the dynamics of self-learning process shows different effect on strategy evolution from pairwise comparison process and the Moran process. Changing the dynamics affects the outcome of an evolutionary process. For games of different types, when a certain strategy needs to be promoted, different dynamics should be chosen for the purpose. Besides, at certain states of evolutionary stable, there exists "ping-pong effect" with stable frequency. Our results hold for weak selection, which is useful for agent interactions, but it is also meaningful to derive simple results that hold for any intensity of selection. Similarly, results for multi-strategies games and those in structured populations are deserved to be further studied. Furthermore, our model can be used to explain the evolutionary dynamic characters of self-learning multi-agent system in many areas, such as UAVs formate, fish swarm formation, etc.

Acknowledgment

This work is supported by the National Basic Research Program of China (973 Program) under Grant No. 2012CB821203 and the National Natural Science Foundation of China (NSFC) under Grants No. 61020106005 and No. 10972002.

References

[1] M. A. Nowak, A. Sasaki, C. Taylor, and D. Fudenberg, "Emergence of cooperation and evolutionary stability in finite populations," Nature, vol. 428, pp. 646-650, April 2004.

[2] J. A. Fax and R. M. Murray, "Information flow and cooperative control of vehicle formations," IEEE Trans. Automat. Contr., vol. 49, pp. 1465-1476, September 2004.

[3] K.-S. Hwang and S.-W. Tan and C.-C. Chen, "Cooperative strategy based on adaptive Q-learning for robot soccer systems," IEEE Trans. Fuzzy Syst., vol. 12, pp. 569-576, August 2004.

[4] R. Olfati-Saber, J. A. Fax and R. M. Murray, "Consensus and cooperation in networked multi-agent systems," Proc. IEEE, vol. 95, pp. 215-233, January 2007.

[5] W. Dong and J. A. Farrell, "Cooperative control of multiple nonholonomic mobile agents," IEEE Trans. Automat. Contr., vol. 53, pp. 1434-1448, July 2008.

[6] J. Du, B. Wu, and L. Wang, "Evolution of global cooperation driven by risks," Phys. Rev. E, vol. 85, pp. 056117, May 2012.

[7] J. M. Smith and G. R. Price, "The logic of animal conflict," Nature, vol. 246, pp. 15-18, November 1973.

[8] J. M. Smith, *Evolution and the Theory of Games*, Cambridge, U. K.: Cambridge University Press, 1982.

[9] J. W. Weibull, *Evolutionary Game Theory*, Cambridge, MA: The MIT Press, 1995.

[10] M. A. Nowak and K. Sigmund, "Evolutionary dynamics of biological games," Science, vol. 303, pp. 793-799, February 2004.

[11] L. A. Imhof and M. A. Nowak, "Evolutionary game dynamics in a Wright-Fisher process," J. Math. Biol., vol. 52, pp. 667-681, May 2006.

[12] G. Szabó and C. Tőke, "Evolutionary prisoner's dilemma game on a square lattice," Phys. Rev. E, vol. 58, pp. 69-73, July 1998.

[13] A. Traulsen, J. M. Pacheco, and M. A. Nowak, "Pairwise comparison and selection temperature in evolutionary game dynamics," J. Theor. Biol., vol. 246, pp. 522-529, June 2007.

[14] K. Gao, W.-X. Wang, and B.-H. Wang, "Self-questioning games and ping-pong effect in the BA network," Physica A, vol. 380, pp. 528-538, July 2007.

[15] X. Chen and L. Wang, "Promotion of cooperation induced by appropriate payoff aspirations in a small-world networked game," Phys. Rev. E, vol. 77, pp. 017103, January 2008.

[16] C. P. Roca and D. Helbing, "Emergence of social cohesion in a model society of greedy, mobile individuals," Proc. Natl. Acad. Sci. U. S. A., vol. 108, pp. 11370-11374, July 2011.

[17] B. Wu, P. M. Altrock, L. Wang, and A. Traulsen, "Universality of weak selection," Phys. Rev. E, vol. 82, pp. 046106, October 2010.

[18] M. Kandori, G. J. Mailath, and R. Rob, "Learning, mutation, and long run equilibria in games," Econometrica, vol. 61, pp. 29-56, January 1993.

[19] T. Antal, M. A. Nowak, and A. Traulsen, "Strategy abundance in 2×2 games for arbitrary mutation rates," J. Theor. Biol., vol. 257, pp. 340-344, March 2009.

[20] G. Hardin, "The tragedy of the commons," Science, vol. 162, pp. 1243-1248, December 1968.

[21] M. Milinski, R. D. Sommerfeld, H.-J. Krambeck, F. A. Reed, and J. Marotzke, "The collective-risk social dilemma and the prevention of simulated dangerous climate change," Proc. Natl. Acad. Sci. U. S. A., vol. 105, pp. 2291-2294, February 2008.

[22] C. S. Gokhale and A. Traulsen, "Evolutionary games in the multiverse," Proc. Natl. Acad. Sci. U. S. A., vol. 107, pp. 5500-5504, March 2010.

[23] J. M. Fowler, "A formation flight experiment," IEEE Contr. Syst., vol. 23, pp. 35-43, October 2003.

[24] I. L. Bajec and F. H. Heppner, "Organized flight in birds," Anim. Behav., vol. 78, pp. 777-789, August 2009.

[25] F. Dörfler and B. Francis, "Geometric analysis of the formation problem for autonomous robots," IEEE Trans. Automat. Contr., vol. 55, pp. 2379-2384, October 2010.

[26] H. Weimerskirch, J. Martin, Y. Clerquin, P. Alexandre, and S. Jiraskova, "Energy saving in flight formation," Nature, vol. 413, pp. 697-698, October 2001.

[27] R. L. Graham, D. E. Knuth, and O. Patashnik, *Concrete Mathematics*, 2nd ed. Massachusetts: Addison-Wesley Publishing Company, 1994.

[28] A. Traulsen, D. Semmann, R. D. Sommerfeld, H.-J. Krambeck, and M. Milinski, "Human strategy updating in evolutionary games," Proc. Natl. Acad. Sci. U. S. A., vol. 107, pp. 2962-2966, February 2010.

[29] M. Milinski and C. Wedekind, "Working memory constrains human cooperation in the Prisoner's Dilemma," Proc. Natl. Acad. Sci. U. S. A., vol. 95, pp. 13755-13758, November 1998.

[30] M. Posch, A. Pichler, and K. Sigmund, "The efficiency of adapting aspiration levels," Proc. R. Soc. B, vol. 266, pp. 1427-1435, July 1999.

[31] N. G. van Kampen, *Stochastic Processes in Physics and Chemistry*, 3rd ed. Amsterdam: Elsevier Science, 2007.

[32] C. W. Gardiner, *Handbook of Stochastic Methods for Physics, Chemistry, and the Natural Sciences*, 3rd ed. New York: Springer, 2004

[33] R. Selten, "Reexamination of the perfectness concept for equilibrium points in extensive games," Int. J. Game Theory, vol. 4, pp. 25-55, 1975.

[34] D. G. Manolakis, V. K. Ingle and S. M. Kogon, *Statistical and adaptive signal processing: spectral estimation, signal modeling, adaptive filtering, and array processing*, Norwood, MA: Artech House, 2000.

[35] G. B. Giannakis, *Cyclostationary Signal Analysis*, 2nd ed. Boca Raton, FL: CRC Press, 2010.

[36] S. Qian, *Introduction to Time-Frequency and Wavelet Transforms*, London, England: Prentice Hall, 2001.