Eur. Phys. J. B (2017) 90: 138
DOI: 10.1140/epjb/e2017-80219-0

The European Physical Journal B

Regular Article

Evolutionary fate of memory-one strategies in repeated prisoner's dilemma game in structured populations

Xu-Sheng Liu[1], Zhi-Xi Wu[1,a], Michael Z. Q. Chen[2,b], and Jian-Yue Guan[1]

[1] Institute of Computational Physics and Complex Systems, Lanzhou University, Lanzhou, Gansu 730000, P.R. China
[2] School of Automation, Nanjing University of Science and Technology, Nanjing, Jiangsu 210094, P.R. China

Received 13 April 2017 / Received in final form 22 May 2017
Published online 24 July 2017 – © EDP Sciences, Società Italiana di Fisica, Springer-Verlag 2017

Abstract. We study evolutionary spatial prisoner's dilemma game involving a one-step memory mechanism of the individuals whenever making strategy updating. In particular, during the process of strategy updating, each individual keeps in mind all the outcome of the action pairs adopted by himself and each of his neighbors in the last interaction, and according to which the individuals decide what actions they will take in the next round. Computer simulation results imply that win-stay-lose-shift like strategy win out of the memory-one strategy set in the stationary state. This result is robust in a large range of the payoff parameter, and does not depend on the initial state of the system. Furthermore, theoretical analysis with mean field and quasi-static approximation predict the same result. Thus, our studies suggest that win-stay-lose-shift like strategy is a stable dominant strategy in repeated prisoner's dilemma game in homogeneous structured populations.

1 Introduction

Cooperation exists widely in many biological, social and economic systems [1], and plays an important role in the functionality of these complex systems. However, creatures are considered to be selfish and always attempt to maximize their own benefits in terms of the Darwinian evolution theory [2–6]. Hence, understanding the emergence and maintenance of cooperative behavior among selfish individuals is of paramount importance. In recent decades, evolutionary game theory [2–12] has been introduced and proven to be the canonical theoretical framework to investigate this problem. The most well-known paradigm that describes the one-shot game between two individuals is known as the prisoner's dilemma game (PDG), which has attracted much attention in a great number of theoretical and experimental studies [4,13–16]. In the standard PDG, two players simultaneously decide whether to cooperate (C) or to defect (D), and receive payoffs according to their respective choices. Particularly, for two players with the same strategy results in a benefit R if both cooperate while a punishment P if both defect respectively. For two players with different strategies, the cooperator gets the sucker's payoff S while the defector gets the temptation to defect T. These payoffs satisfy the relation as $T > R > P > S$, such that defection is the absolutely best choice for rational players regardless of the opponent's selection. For repeated PDG $T + S < 2R$ is also presumed [13]. Therefore, in the absence of supporting mechanisms, all the players would get the punishment P instead of the higher reward R provided that they were to cooperate with each other, hence resulting in a dilemma situation.

During the process of development of evolutionary game theory, many interesting strategies are put forward for the PDG. In R. Axelrod's famous tournament [13,17], the tit-for-tat strategy (cooperating in the first round and then doing whatever the other one did at last round) performs excellently, which has stimulated much attention to the study on reactive strategies (where the decision of one player in each round depends on the previous move of the opponent) in both social and biological systems. Although the tit-for-tat strategy is found to be effective in establishing high cooperative level community, it is sensitive to "noise", where an accidental (or mistaking) defection can lead to a long sequence of retaliation for tit-for-tat players. In their 1992 Nature paper [18], Nowak and Sigmund found that the tit-for-tat strategy could be replaced by a more generous strategy, the generous tit-for-tat strategy, which performs better than tit-for-tat for its fault toleration. Particularly, subsequent researches of Nowak and coworkers suggested that the win-stay-lose-shift strategy (repeating previous move whenever you are doing well, but changing otherwise) seems to be a more popular rule for its two advantages over tit-for-tat [19]: it can correct occasional mistakes and exploit unconditional cooperators.

[a] e-mail: wuzhx@lzu.edu.cn
[b] e-mail: mzqchen@outlook.com

Two extented win-stay-lose-shift strategies [20,21] are proposed recently. They both take individuals' aspiration for payoff into consideration and found that the cooperation will be promoted in a proper range of aspirations [20] or the payoff parameters [21]. It is not difficult to find that all these strategies require the players to have one-step memory, i.e. they can be categorized into the class of memory-one strategy. Actually, even "unconditional cooperate" and "unconditional defect" can be considered as memory-one strategies.

Recently, Press and Dyson [22] have coined the term "zero-determinant" (ZD) strategy in evolutionary games, which responds probabilistically in terms of both the players' and the co-players' previous moves. The peculiarity of ZD strategy lies in the fact that it allows players to set their opponents' payoffs unilaterally, i.e., independent of their strategies or responses. This important progress has remotivated people's interest in the memory-one strategy in repeated games in recent years [23–25]. In reference [26], Vukov studied a spatial evolutionary PDG with memory-one strategy in the square lattice population, and found that evolution selects a generous tit-for-tat-like strategy from the memory-one strategy set, which gives rise to a cooperative community with a strikingly high cooperation level for any value of the temptation to defection. In addition, it was found that whether the learning process of the strategies is accurate or not is relevant in establishing cooperation, and the more accurate handling of specific situations is helpful in creating more cooperative societies [26]. Very recently, Baek et al. [27] systematically compared the evolutionary performance of reactive strategies and the memory-one strategies in the PDG in finite, well-mixed populations (which means that each individual interacts with all the others with equal probability). They considered both deterministic strategy and stochastic strategy spaces, and the strategy evolution is performed by the Moran process [28]. It was found that for memory-one strategy, stochasticity may promote or hinder the evolution of cooperation, which depends on the magnitude of the temptation to defection, and the win-stay-lose-shift like strategy (which means the entry values of those strategies are close to win-stay-lose-shift strategy but exist some fluctuations) seems to be of evolutionary robustness in a large range of the parameter regime [27].

We note that reference [26] has treated the competition and evolution of memory-one strategies in population located on a square lattice, whose research is mainly based on computer simulations. The studies in reference [27] have involved both theoretical derivations and Monte Carlo simulations, but it only treated the case of well-mixed populations. In reality, the interactions among individuals are mostly spatially restricted, whose interacting patterns are usually modeled by regular or complex networks (or graphs) [29]. It has been known that the spatial structure is a relevant factor in promoting the evolution of cooperation [5]. In this work, we intend to study the evolutionary fate of a memory-one strategy in a spatially structured population via the approaches of both theoretical analysis and the computer simulations. As to be shown below, we find that when the strategy imitation (or learning) process is accurate enough, the win-stay-lose-shift like strategy will usually stand out from the memory-one strategy set.

2 Model

In our model, we consider the evolutionary dynamics of PDG in homogeneous structured populations, where the players are located on the nodes of either a regular square lattice or on a complex network. Specifically, two types of networks have been considered (a square lattice with periodic boundary conditions and a random regular network). During the evolution, only two possible actions, cooperate (C) or defect (D), can be adopted by the players. After interacting with their neighbors, the individuals acquire their payoffs according to the payoff matrix of the PDG. For simplicity but without loss of generality, the entries of the payoff matrix are formulated as [30]

$$\begin{array}{cc} & \begin{array}{cc} C & D \end{array} \\ \begin{array}{c} C \\ D \end{array} & \begin{pmatrix} 1 & 1-b \\ b & 0 \end{pmatrix} \end{array}, \tag{1}$$

where $b \in [1, 2]$ denotes the temptation to defect.

To take into account of the one-step memory, the strategy of each individual is represented by a vector of quartuple space [5,26]. Let $\boldsymbol{P}_i = (p^i_{cc}, p^i_{cd}, p^i_{dc}, p^i_{dd})$ denotes the strategy of individual i. Elements in $\boldsymbol{P}_i$ represent the probabilities of adopting cooperation of individual i when interacting with one involved neighbor in the next move, subject to the action pair $\{cc, cd, dc, dd\}$ being the outcome of their encounter at the last round, respectively. With this setting, there are more degrees of freedom than the model studied in references [28,31–34] when individuals select their behavior. For instance, strategy $\boldsymbol{P} = (1, 1, 1, 1), (0, 0, 0, 0), (1, 0, 1, 0), (1, 0, 0, 1)$ means unconditional cooperate, unconditional defect, tit-for-tat and win-stay-lose-shift, respectively. Note that, with this strategy configuration, one individual may take different actions to different neighborhoods.

In the beginning, each element in the strategy vector of all individuals are initialized with uniformly distributed random numbers between 0 and 1. All individuals' probabilities of cooperation at their first move are decided by the mean value of the four elements in their strategy vector, namely $(p_{cc} + p_{cd} + p_{dc} + p_{dd})/4$, and thus the initial probability of defection $1 - (p_{cc} + p_{cd} + p_{dc} + p_{dd})/4$. During the evolution, the state of the system is updated in an asynchronous manner. In particular, at each discrete-time step, an edge is randomly selected in the network with two endpoints, i and j. Then the two selected players are allowed to interact with all their neighbors and collect the corresponding payoffs π_i and π_j in terms of their actual actions and the payoff matrix, respectively. Subsequently, the two involved players try to promote their payoffs by imitating (or learning) one of the opponent's strategy. Following previous studies [16], the player i will adopt the strategy of the player j with the probability $W_{i\to j}$, which is assumed to be proportional to their payoff difference $(\pi_j - \pi_i)$. Specifically, the transition probability can be written as [16]

$$W_{i\to j} = \frac{1}{1+\exp[-\beta(\pi_j - \pi_i)]}, \qquad (2)$$

where $\beta > 0$ denotes the uncertainty (noise) in the imitation process. From the viewpoint of biology, β can also be regarded as the strength of selection [35,36]. In this article, we use $\beta = 2.5$ in our Monte Carlo simulations. Obviously, individuals with higher payoffs are imitated by others with higher probability. It is also worth noting that individuals can also imitate their neighbors who yielded lower payoffs, even with somewhat lower probabilities, which indicate that people may make mistakes in the learning process.

It is easy to see that the space of strategy is huge in the considered model. To visit points in the strategy space as much as possible, we consider the fuzzy learning method as in [26], which means that the individuals do not learn their neighbors' strategies with exact accuracy. Whenever the imitation process happens, the elements in the learner's strategy vector are replaced by a normally distributed random number, whose expectation values correspond to the elements of the neighbor's strategy vector and with standard deviation σ. In our current study, we use $\sigma = 0.005$ so that the learning process is sufficiently accurate. The above elementary process is repeated until the system reaches a steady state where the frequency of cooperators ρ_C in the system fluctuates stably.

For simplicity, we discretize the strategy space as what has been done in references [26,27]. To be more specific, we first divide each dimension of the strategy vector into ten parts equally with interval 0.1, and assign each part from small to large with number (0, 1, 2, 3, 4, 5, 6, 7, 8, 9), respectively. Then each dimension of a strategy can be mapped onto an decimal integer number by just taking the integer portion of $p_{xy}/0.1$ (where xy denotes the combination $\{cc, cd, dc, dd\}$), which can be denoted by $\lfloor p_{xy}/0.1\rfloor$. By doing this, each strategy is then represented by a four-digit decimal positive integer, say $\lfloor p_{cc}/0.1\rfloor\times10^3+\lfloor p_{cd}/0.1\rfloor\times10^2+\lfloor p_{dc}/0.1\rfloor\times10+\lfloor p_{dd}/0.1\rfloor$, which corresponds to a hypercube area of four-dimension space with edge length 0.1. For instance, the strategy $\boldsymbol{P} = (0.1, 0.2, 0.3, 0.3)$ can be mapped to the integer array $(1, 2, 3, 3)$, which corresponds to the decimal integer 1233.

The Monte Carlo simulations are carried out for a population of size $N = 100 \times 100$. The simulation results are obtained by averaging over the last 10^5 Monte Carlo time steps of the total 2×10^5. Each data point presented below is the result from an average of over 20−50 independent realizations.

3 Simulation results and theoretical analysis

We first present the simulation results of our model. The average values of each dimension of the strategy vector for all the individuals as a function of the temptation to defect b on the two kinds of networks, i.e., a square lattice with von Neumann neighborhood and a random regular network with degree $k = 4$ are shown in Figure 1. For the case of the square lattice, as is shown in Figure 1a, we observe that $\bar{p}_{cc}$ are close to 1, while $\bar{p}_{cd}$ and $\bar{p}_{dc}$ are close to 0 through out the entire range of the parameter b. Unlike the steadiness of $\bar{p}_{cc}$, $\bar{p}_{cd}$ and $\bar{p}_{dc}$ versus b, $\bar{p}_{dd}$ decreases continuously with the increase of the parameter b. Basically, in the whole range of b, the variant of the win-stay-lose-shift strategy $(1, 0, 0, x)$ (with $x \in [0, 1]$) is favored in the population in the stationary state. Particularly, we notice that $\bar{p}_{dd}$ is always greater than 0.8 when the temptation to defect is not larger than 1.6, which means that most individuals in the system are inclined to adopt the win-stay-lose-shift like strategy for $b \leq 1.6$. The results achieved on the random regular network, as is shown in Figure 1b, are almost the same as those on the square lattice.

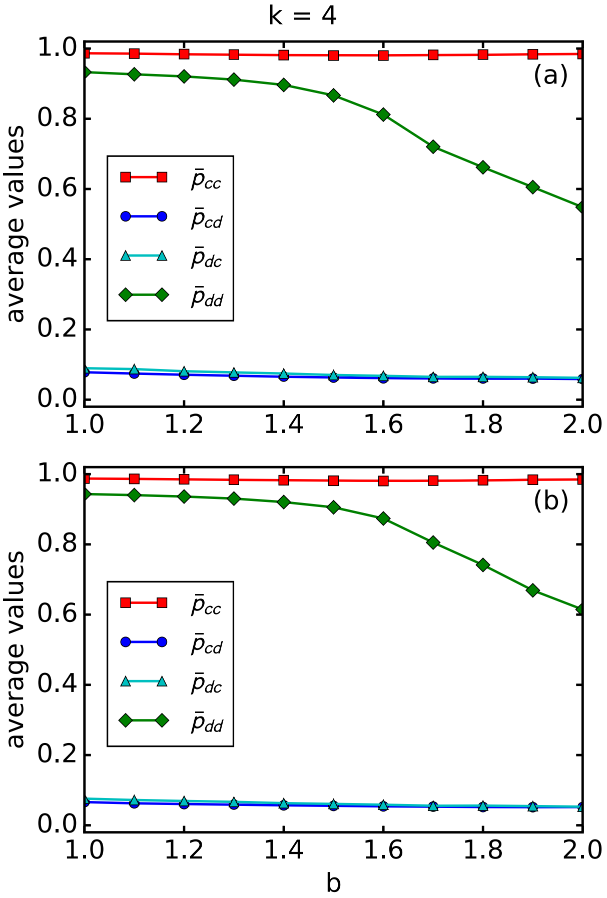

Fig. 1. Average values of each dimension of the strategy vector for all individuals in the stationary state $(\bar{p}_{cc}, \bar{p}_{cd}, \bar{p}_{dc}, \bar{p}_{dd})$, as a function of the temptation to defect b for $\sigma = 0.005$ and $\beta = 2.5$. (a) The results are for the case of the square lattice with von Neumann neighborhood; (b) the results are for the case of the random regular network with degree $k = 4$. The results on the two kinds of networks are almost identical. For $b = 1.4$ in (a), $\bar{p}_{cc} = 0.97$, $\bar{p}_{dd} = 0.90$, $\bar{p}_{cd} = \bar{p}_{dc} = 0.07$. The data points are averaged over 20 independent trials.

Then we intend to measure how the stationary strategies are distributed in the strategy space. The stationary

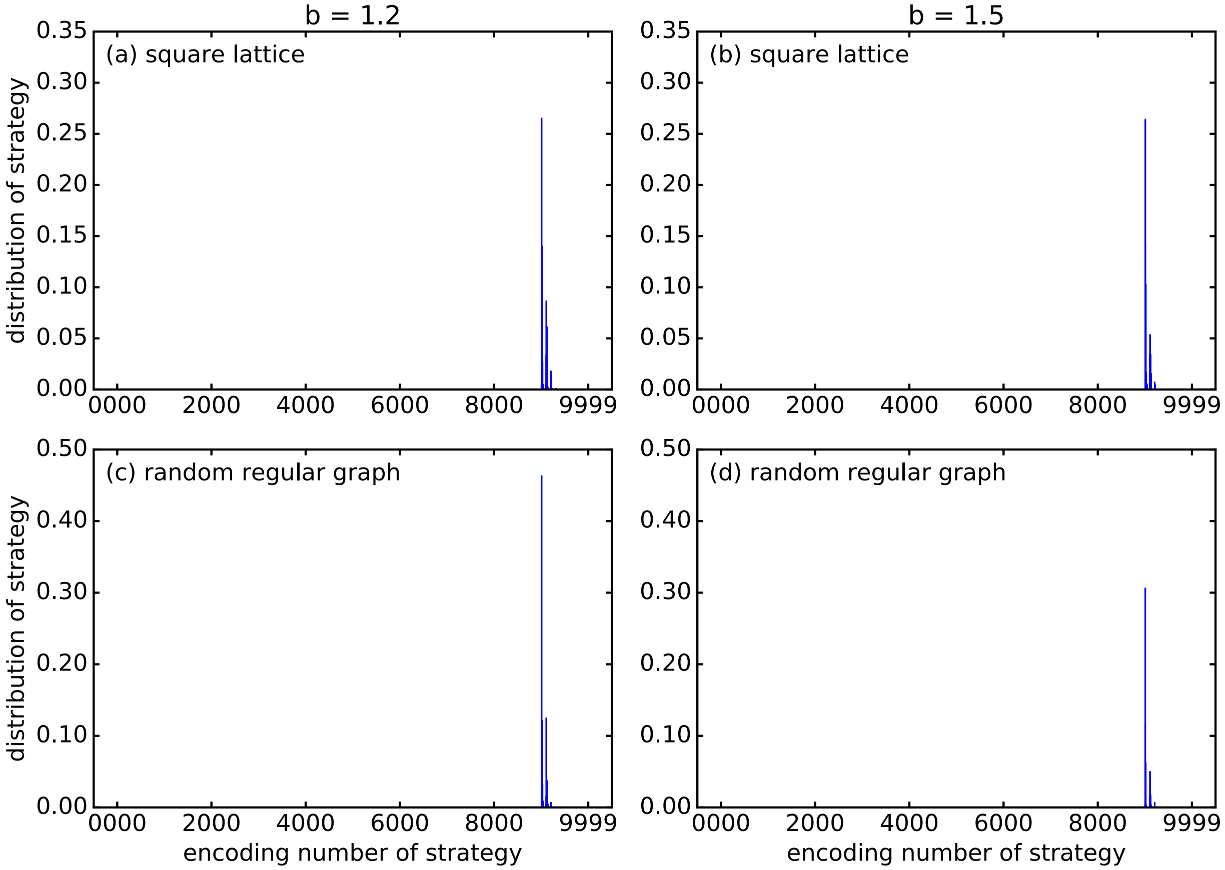

Fig. 2. Statistical distribution of the strategies after the system evolves into the final stable state for $\sigma = 0.005$ and $\beta = 2.5$ on the square lattice with Von Neumann neighborhood (a) and (b), and on the random regular network with degree $k = 4$ (c) and (d). Here, $b = 1.2$ in left panels and $b = 1.5$ in right panels. In this figure, each strategy is mapped onto a corresponding four-digit decimal integer, and the highest peak corresponds to the encoding number 9009. All data are averaged over 50 independent trials.

distribution of the encoded numbers mapped from the strategies are shown in Figure 2 for the evolutionary PDG on the square lattice with von Neumann neighborhood and the random regular network with degree $k = 4$ for $b = 1.2$ and $b = 1.5$, respectively. The highest peak in the distribution corresponds to the encoded number 9009, which is close to the well-known win-stay-lose-shift strategy $\boldsymbol{P} = (1.0, 0.0, 0.0, 1.0)$. We would like to point out that due to the fuzzy learning mechanism, not all strategies emergent in the system are located in the position corresponding to 9009. As is shown in Figure 2, there also exist some lower peaks. In fact, most lower peaks in Figure 2 are close to 9009. The top six encoding numbers in terms of the magnitude of their frequencies, corresponding to Figure 2b, are listed in Table 1. Obviously all these top-ranked encoding number are in the vicinity of the number 9009.

In addition to the case of networks with degree $k = 4$ above, we also study our model on square lattice and random regular networks with degree $k = 8$. The results are presented in Figure 3, which imply that win-stay-lose-shift like strategy still win out of the memory-one strategy set in the stationary state in a wide range of the payoff parameters. The high values of $\bar{p}_{cd}$ and $\bar{p}_{dc}$ in Figure 3b for small value of b is due to the reason that there is always a chance for unconditional cooperate, i.e. (1, 1, 1, 1), rather than win-stay-lose-shift like strategy, to dominate the whole system on random regular network when b is sufficiently small. As is shown in the inset of Figure 3b, the average values of the strategy elements are more and more close to those of the win-stay-lose-shift strategy along with the increasing of b.

Table 1. The top six frequencies of strategy encode numbers after system evolve into equilibrium state on square lattice with von Neumann neighborhood for $b = 1.5$ and $\sigma = 0.005$.

No.	Strategy encode number	Frequency
1	9009	0.2640
2	9008	0.1422
3	9019	0.1030
4	9007	0.0858
5	9018	0.0671
6	9109	0.0534

Though the implementation of simulation in our work is similar to reference [26], our results imply that

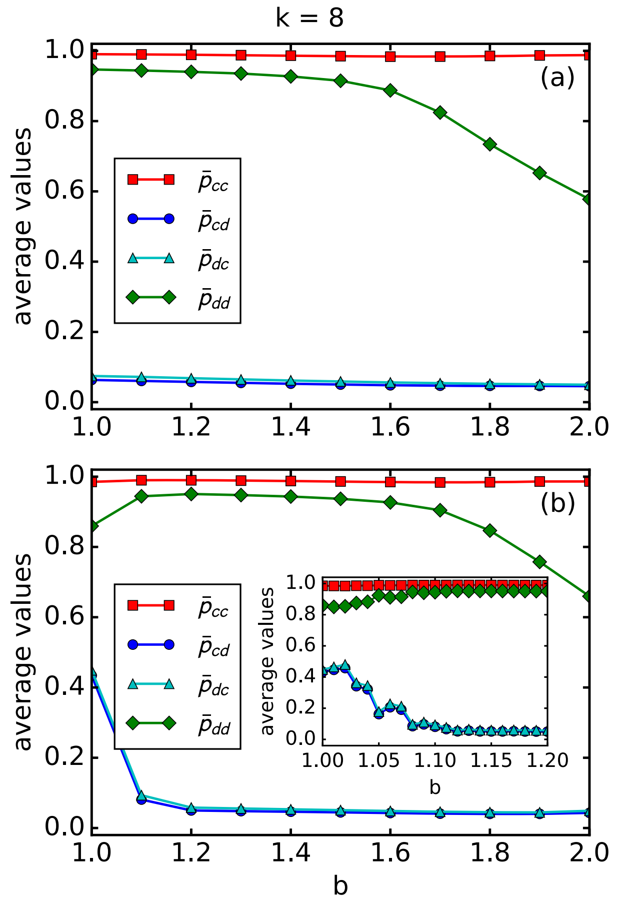

Fig. 3. Average values of each dimension of the strategy vector for all individuals in the stationary state ($\bar{p}_{cc}$, $\bar{p}_{cd}$, $\bar{p}_{dc}$, $\bar{p}_{dd}$), as a function of the temptation to defect b for $\sigma = 0.005$ and $\beta = 2.5$. (a) The results are for the case of the square lattice with Moore neighborhood; (b) the results are for the case of the random regular network with degree $k = 8$ (the inset shows the detailed variance of average values with $b \in [1.0, 1.2]$). The data points are averaged over 50 independent trials.

win-stay-lose-shift- (rather than tit-for-tat-) like strategy will win out of the memory-one strategy set in the stationary state in a wide range of the payoff parameters which is quite different from the conclusion in reference [26], where the tit-for-tat like strategy dominates the system in most parameter space. We would like to point out that our theoretical analysis presented below provides solid support for the correction of our simulation results. Consequently, it would be reasonable to guess tentatively that there could be an error in the author's code which would be very difficult to find out the error, if one exists, without the code the author used. Nevertheless, the results that the more accurate handling of strategies learning process can help create more cooperative societies in reference [26] can be reproduced correctly.

In what follows, we would like to present some approximate theoretical analysis to explain why the strategy 9009 is the most abundant one in the steady state. According to the analytical method used in the previous work [22,24,37], the evolutionary process of the repeated PDG between two players, say i and j with strategies $\boldsymbol{P}_i = (p^i_{cc}, p^i_{cd}, p^i_{dc}, p^i_{dd})$ and $\boldsymbol{P}_j = (p^j_{cc}, p^j_{cd}, p^j_{dc}, p^j_{dd})$, can be depicted by the iteration of the transition matrix, denoted by $\boldsymbol{M}(\boldsymbol{P}_i, \boldsymbol{P}_j)$, acting on some vector $\boldsymbol{V} = (v_{cc}, v_{cd}, v_{dc}, v_{dd})^T$, in which each element v_{xy} denotes the frequency of the arising of action pair $xy \in \{cc, cd, dc, dd\}$. The transition matrix $\boldsymbol{M}(\boldsymbol{P}_i, \boldsymbol{P}_j)$ can be written as

$$\boldsymbol{M}(\boldsymbol{P}_i, \boldsymbol{P}_j) =$$

$$\begin{bmatrix} p^i_{cc}p^j_{cc} & p^i_{cc}(1-p^j_{cc}) & (1-p^i_{cc})p^j_{cc} & (1-p^i_{cc})(1-p^j_{cc}) \\ p^i_{cd}p^j_{dc} & p^i_{cd}(1-p^j_{dc}) & (1-p^i_{cd})p^j_{dc} & (1-p^i_{cd})(1-p^j_{dc}) \\ p^i_{dc}p^j_{cd} & p^i_{dc}(1-p^j_{cd}) & (1-p^i_{dc})p^j_{cd} & (1-p^i_{dc})(1-p^j_{cd}) \\ p^i_{dd}p^j_{dd} & p^i_{dd}(1-p^j_{dd}) & (1-p^i_{dd})p^j_{dd} & (1-p^i_{dd})(1-p^j_{dd}) \end{bmatrix}.$$

In terms of the mathematical property of transition matrix $\boldsymbol{M}$ (which is actually the transition matrix of a Markov chain with non-negative elements), there must exist a stationary vector $\boldsymbol{V}_s$ with a unit eigenvalue satisfies [22]

$$\boldsymbol{V}_s^T \boldsymbol{M} = \boldsymbol{V}_s^T, \tag{3}$$

which corresponds to the equilibrium state of any iterated two-player game. Then we have

$$\boldsymbol{V}_s^T \boldsymbol{M}' = 0, \tag{4}$$

where $\boldsymbol{M}' = \boldsymbol{M} - \boldsymbol{E}$ and $\boldsymbol{E}$ denotes the identity matrix.

In terms of Cramer's rule and the Laplace expansion of matrix $\boldsymbol{M}'$, the dot product between $\boldsymbol{V}_s$ and an arbitrary vector $\boldsymbol{x} = (x_1, x_2, x_3, x_4)^T$ can be given by the determinant of the matrix $\boldsymbol{M}'$ [22]

$$D(\boldsymbol{P}_i, \boldsymbol{P}_j, \boldsymbol{x}) = \det \begin{bmatrix} -1 + p^i_{cc}p^j_{cc} & -1 + p^i_{cc} & -1 + p^j_{cc} & x_1 \\ p^i_{cd}p^j_{dc} & -1 + p^i_{cd} & p^j_{dc} & x_2 \\ p^i_{dc}p^j_{cd} & p^i_{dc} & -1 + p^j_{cd} & x_3 \\ p^i_{dd}p^j_{dd} & p^i_{dd} & p^j_{dd} & x_4 \end{bmatrix}. \tag{5}$$

Consequently, the equilibrium payoff of the individual i from playing against its neighbor j can be calculated as [22]

$$f(\boldsymbol{P}_i, \boldsymbol{P}_j) = \frac{\boldsymbol{V}_s \cdot \boldsymbol{R}}{\boldsymbol{V}_s \cdot \boldsymbol{I}} = \frac{D(\boldsymbol{P}_i, \boldsymbol{P}_j, \boldsymbol{R})}{D(\boldsymbol{P}_i, \boldsymbol{P}_j, \boldsymbol{I})}, \tag{6}$$

where $\boldsymbol{R} = (1, 1-b, b, 0)^T$ is the payoff vector of the individual i by playing against the individual j with the emerged action pair $\{cc, cd, dc, dd\}$, i.e., equation (1). Here $\boldsymbol{I}$ is the vector with all elements 1.

On account of the intrinsic stochasticity of the strategy transformation, accurate prediction of the cooperation probability of an individual in every move is impossible. Here, we alternatively consider the situation that the system evolves sufficiently slowly such that each individual and its neighbors can be assumed to be in their stable state when strategy updating happens, i.e. we regard that the system evolves in a way analogous to the quasi-static process in thermodynamics. Now let $n(\boldsymbol{P}_m, t)$ denote the frequency of the strategy $\boldsymbol{P}_m$ at time t. Based on the above analysis, the time evolution of the probability $n(\boldsymbol{P}_m, t)$ obeys the following master equation

$$\frac{\mathrm{d}n(\boldsymbol{P}_m,t)}{\mathrm{d}t} = \sum_{\boldsymbol{P}_l} n(\boldsymbol{P}_l,t)n(\boldsymbol{P}_m,t)[T(\boldsymbol{P}_l \to \boldsymbol{P}_m) - T(\boldsymbol{P}_m \to \boldsymbol{P}_l)], \tag{7}$$

where $T(\boldsymbol{P}_l \to \boldsymbol{P}_m)$ represents the transition rate

$$T(\boldsymbol{P}_l \to \boldsymbol{P}_m) = \frac{1}{1+\exp[-\beta(\pi_{\boldsymbol{P}_m} - \pi_{\boldsymbol{P}_l})]}. \tag{8}$$

Here $\pi_{\boldsymbol{P}_x}$ $(x \in \{k, l\})$ stands for the mean payoff acquired by using the strategy $\boldsymbol{P}_x$

$$\pi_{\boldsymbol{P}_x} = \sum_{\boldsymbol{P}_y} n(\boldsymbol{P}_y,t) f(\boldsymbol{P}_x,\boldsymbol{P}_y), \tag{9}$$

where $f(\boldsymbol{P}_x, \boldsymbol{P}_y)$ denotes the mean payoff acquired by adopting strategy $\boldsymbol{P}_x$ playing against the strategy $\boldsymbol{P}_y$, i.e., equation (6).

As in the case of Monte Carlo simulations, we slice the the strategy space into discrete points to simplify the problem. In particular, we first divide each dimension of the strategy space into ν parts equally, which will generates $(\nu+1)^4$ strategy points totally. To ensure the transition matrix $\boldsymbol{M}(\boldsymbol{P}_i, \boldsymbol{P}_j)$ to remain nonsingular in the calculation, we just consider all the points except those on the edges, faces and corners of the four-dimensional hypercube of the strategy space. By doing so, we finally have $\nu - 1$ points in each dimension, and we thus have totally $N = (\nu-1)^4$ points. Similar to the way that we have encoded the strategies by four-digit integers in Figure 2, we use $(\nu-1)$-based positional notations to encode these strategy points. The initial frequency of each strategy is set to $1/N$. By solving the equation (7) numerically, we obtain the distribution of the strategies survival in the stationary state. As is shown in Figure 4, the win-stay-lose-shift like strategy, corresponding to the encoding number 9009, is proved to dominate in the system.

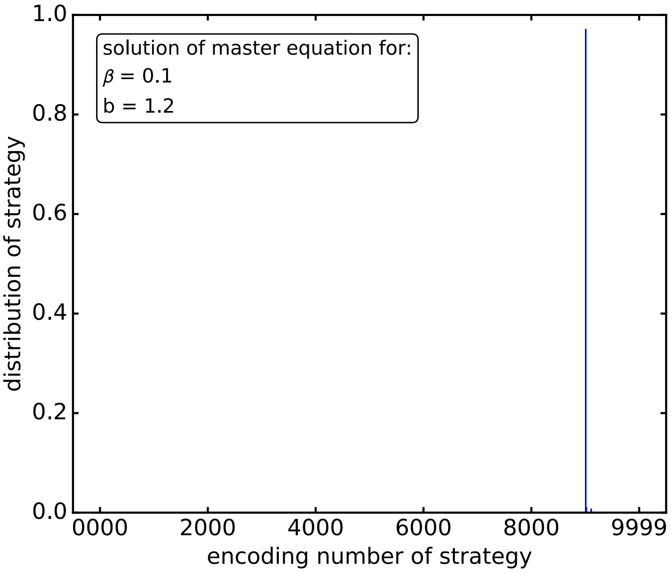

Fig. 4. Statistical distribution of the encoding numbers of the strategies, calculated by solving the master equation numerically for $\beta = 0.1$ and $b = 1.2$. Here we let $\nu = 11$. Note that in our approximate theoretical analysis, we only consider the weak-selection case, i.e., small value of β, so that the evolutionary process can be regarded as a quasi-static process.

4 Conclusions and discussion

In summary, we have studied the evolutionary prisoner's dilemma game in square lattice and random regular networks by taking into account of one-step memory of the players, which is different from the well-mixed case of reference [27]. In our model, strategies are represented by points in a four-dimensional hypercube space. The individuals make decisions according to what they and their opponents have done at the last encounter. Monte Carlo simulation results suggest that the memory-one strategy, which behaves like the classical win-stay-lose-shift strategy, is mostly evolutionary robust in a wide range of the payoff parameters. Our theoretical analysis with mean field and quasi-static approximation show that the win-stay-lose-shift like strategy dominate the whole system in the equilibrium. The quasi-static approximation requires the system evolves sufficiently slowly, which means that what we treat theoretically is actually the weak selection process [5]. Surprisingly, our simulation results under strong selection (large value of β) indicate that the conclusions achieved by theoretical analysis can be extended to the strong selection case. To sum up, both our simulation and theoretical results indicate that the win-stay-lose-shift like strategy is the stable dominant strategy in evolutionary spatial prisoner's dilemma games.

Finally, it is worth mentioning that two different theoretical analysis methods have been used in reference [27]. The first one is to construct a Markov matrix with finite strategy space by computing fixation probability of all strategy pairs in finite populations to find the unique invariant strategy distribution. The other one uses perturbative method to compute exact strategy abundance in the limit of weak selection to determine the most favored stochastic strategy. Their analytical methods are totally different with ours, albeit the main conclusions of both works are similar. Our current work therefore provides another new and efficient analytical method to address the evolutionary fate of memory-one strategies in networked prisoner's dilemma games.

We thank Dr. J.-Q. Zhang and Dr. C.-R. Cai for helpful comments and discussions. This work was supported by the National Natural Science Foundation of China (Grant Nos. 11135001, 11575072, 11475074, and 61374053).

Author contribution statement

Xu-Sheng Liu performed Monte Carlo simulations. Xu-Sheng Liu, Zhi-Xi Wu, Michael Z. Q. Chen and Jian-Yue Guan performed theoretical analysis and wrote the paper. All authors contributed to all aspects of this work.

References

1. J.M. Smith, E. Szathmary, *The major transitions in evolution* (Oxford University Press, 1997)
2. R. Axelrod, *The evolution of cooperation* (Perseus Books, 2006), ISBN 9780465005642
3. J.M. Smith, *Evolution and the theory of games* (Cambridge University Press, 1982)
4. H. Gintis, *Game theory evolving: a problem-centered introduction to modeling strategic behavior* (Princeton University Press, 2000)
5. M. Nowak, *Evolutionary dynamics: exploring the equations of life* (Harvard University Press, Cambridge, 2006)
6. K. Sigmund, *The calculus of selfishness* (Princeton University Press, 2010)
7. M.W. Macy, A. Flache, Proc. Natl. Acad. Sci. **99**, 7229 (2002)
8. B. Skyrms, *The stag hunt and the evolution of social structure* (Cambridge University Press, 2004)
9. A. Szolnoki, M. Perc, Phys. Rev. E **85**, 026104 (2012)
10. H.X. Yang, Z.X. Wu, B.H. Wang, Phys. Rev. E **81**, 065101 (2010)
11. Z. Wang, L. Wang, A. Szolnoki, M. Perc, Eur. Phys. J. B **88**, 124 (2015)
12. M. Perc, A. Szolnoki, Biosystems **99**, 109 (2010)
13. R. Axelrod, W.D. Hamilton, Science **211**, 1390 (1981)
14. A. Szolnoki, N.G. Xie, Y. Ye, M. Perc, Phys. Rev. E **87**, 042805 (2013)
15. M. Doebeli, C. Hauert, Ecol. Lett. **8**, 748 (2005)
16. G. Szabó, C. Tőke, Phys. Rev. E **58**, 69 (1998)
17. R. Axelrod, *The evolution of cooperation* (1984)
18. M.A. Nowak, K. Sigmund, Nature **355**, 250 (1992)
19. M. Nowak, K. Sigmund et al., Nature **364**, 56 (1993)
20. Y. Liu, X. Chen, L. Zhang, L. Wang, M. Perc, PLoS ONE **7**, 1 (2012)
21. M.A. Amaral, L. Wardil, M.C.V. Perc, J.K.L. da Silva, Phys. Rev. E **94**, 032317 (2016)
22. W.H. Press, F.J. Dyson, Proc. Natl. Acad. Sci. **109**, 10409 (2012)
23. D. Hao, Z. Rong, T. Zhou, Phys. Rev. E **91**, 052803 (2015)
24. A.J. Stewart, J.B. Plotkin, Proc. Natl. Acad. Sci. **111**, 17558 (2014)
25. C. Hilbe, L.A. Martinez-Vaquero, K. Chatterjee, M.A. Nowak, Proc. Natl. Acad. Sci. **114**, 4715 (2017)
26. J. Vukov, Phys. Rev. E **90**, 032802 (2014)
27. S.K. Baek, H.C. Jeong, C. Hilbe, M.A. Nowak, Sci. Rep. **6**, 25676 (2016)
28. M.A. Nowak, A. Sasaki, C. Taylor, D. Fudenberg, Nature **428**, 646 (2004)
29. G. Szabó, G. Fáth, Phys. Rep. **446**, 97 (2007)
30. P. Langer, M.A. Nowak, C. Hauert, J. Theor. Biol. **250**, 634 (2008)
31. C. Taylor, D. Fudenberg, A. Sasaki, M.A. Nowak, Bull. Math. Biol. **66**, 1621 (2004)
32. J.M. Pacheco, A. Traulsen, M.A. Nowak, Phys. Rev. Lett. **97**, 258103 (2006)
33. C.P. Roca, J.A. Cuesta, A. Sánchez, Europhys. Lett. **87**, 48005 (2009)
34. M. Brede, Europhys. Lett. **94**, 30003 (2011)
35. A. Traulsen, M.A. Nowak, J.M. Pacheco, Phys. Rev. E **74**, 011909 (2006)
36. A. Traulsen, J.C. Claussen, C. Hauert, Phys. Rev. E **74**, 011901 (2006)
37. C. Hauert, H.G. Schuster, Proc. R. Soc. Lond. B: Biol. Sci. **264**, 513 (1997)