Neurocomputing 513 (2022) 104–113

Contents lists available at ScienceDirect

Neurocomputing

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

Reinforcement learning facilitates an optimal interaction intensity for cooperation

Zhao Song [a,b], Hao Guo [b], Danyang Jia [b], Matjaž Perc [c,d,e], Xuelong Li [b], Zhen Wang [a,b,f,*]

[a] School of Mechanical Engineering, Northwestern Polytechnical University, Xi'an 710072, China
[b] School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China
[c] Faculty of Natural Sciences and Mathematics, University of Maribor, Maribor, Slovenia
[d] Department of Medical Research, China Medical University Hospital, China Medical University, Taichung, Taiwan
[e] Complexity Science Hub Vienna, Vienna, Austria
[f] School of Cybersecurity, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China

ARTICLE INFO

Article history:
Received 10 January 2022
Revised 20 July 2022
Accepted 15 September 2022
Available online 24 September 2022
Communicated by Zidong Wang

Keyword:
Evolutionary game theory
Cooperation
Reinforcement learning theory
Interaction intensity

ABSTRACT

Our social interactions vary over time and they depend on various factors that determine our preferences and goals, both in personal and professional terms. Researches have shown that this plays an important role in promoting cooperation and prosocial behavior in general. Indeed, it is natural to assume that ties among cooperators would become stronger over time, while ties with defectors (non-cooperators) would eventually be severed. Here we introduce reinforcement learning as a determinant of adaptive interaction intensity in social dilemmas and study how this translates into the structure of the social network and its propensity to sustain cooperation. We merge the iterated prisoner's dilemma game with the Bush-Mostelle reinforcement learning model and show that there exists a moderate switching dynamics of the interaction intensity that is optimal for the evolution of cooperation. Besides, the results of Monte Carlo simulations are further supported by the calculations of dynamical pair approximation. These observations show that reinforcement learning is sufficient for the emergence of optimal social interaction patterns that facilitate cooperation. This in turn supports the social capital hypothesis with a minimal set of assumptions that guide the self-organization of our social fabric.

1. Introduction

In nature and society, there are various kinds of interactions to support the stability of biological, social, and technological systems. On the one hand, through the interactions, agents could compete for resources and pursue the goal of maximizing their interests, which is consistent with the evolution theory in terms of the survival of the fittest. On the other hand, however, there extensively exists cooperative interaction behavior between agents, which could not be well explained by the evolution theory. Therefore, understanding the emergence of cooperation becomes a key to solving some social dilemmas [1,2]. Aim to this issue, the evolutionary game theory has provided an effective framework because it can well capture the relationship between interactions and the evolutionary dynamics of cooperation [3–5].

With the evolutionary game theory, the prisoner's dilemma game (PDG) is frequently adopted as a standard framework to reflect the interest conflicts between agents and groups [6–8]. In this basic model, two agents need to choose a strategy from cooperation and defection at the same time. Thus, there are four different strategy combinations, wherein mutual cooperation is the best case for the collective benefit, while defection is the best decision to maximize agent's interest regardless of the strategy of the opposite. In traditional cases, the iterated PDG is usually used in a well-mixed infinite population, where agents will interact with the rest in an equal way. However, it is difficult for cooperation to survive under such a scenario. Aim to resolve this issue, hundreds of scenarios have been proposed both theoretically and experimentally. Typically, Nowak summaries five rules for the evolution of cooperation, including the kin selection, direct and indirect reciprocity, group selection, and network reciprocity [9–11]. Besides, many other mechanisms are demonstrated to promote the evolution of cooperation, for instance, coevolution of strategy and structure [12], and mobility[13]. In particular, network reciprocity indicates that cooperators will resist the invasion of defectors by forming

* Corresponding author at: School of Cybersecurity, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China.
E-mail address: zhenwang0@gmail.com (Z. Wang).

https://doi.org/10.1016/j.neucom.2022.09.109
0925-2312/

spatial clusters in networks [14–16]. Along this line, one can study agents' communications and their relationship formed via daily routines on a network, in which vertexes represent agents and edges reflect the relationship between them [17–20]. Taking advantage of networks' property, the research of cooperation on various networks arises to figure out the topological effects on cooperation [21–27]. Besides, some other mechanisms nourishing cooperation in structured population have also been put forward [28–39].

In spite of great progress, the vast majority of posting works are based on a prior assumption that agents interact with all the nearest neighbors. In other words, agents are supposed to have extremely strong interaction intensities to interact with their nearest neighbors during a whole evolutionary process, which is inconsistent with realistic observations. In reality, agents may not necessarily interact with all their nearest neighbors, and the interaction with neighbors will also change under different conditions. In this sense, the absolute interaction relationship between agents can be replaced by a stochastic probability, namely, the interaction intensity, which leads to diverse interaction patterns for agents, and thus various evolutionary dynamics [40–43]. Although it has been demonstrated that adopting such interaction probability in the iterated PDG is able to improve cooperation in structured populations, these works mainly focused on a constant interaction probability. It is apparent that the interaction intensity of an agent will change according to the specific environment. Thus, the adaptive interaction intensity needs to be studied, with which agents could change their interactions adaptively.

In recent years, the reinforcement learning (RL) methods, rooted in psychology and neuroscience, have become another approach to understanding the cooperation in evolutionary processes [44,45]. Ordinary RL algorithms contain the Monte Carlo method [46], Sarsa [47], dynamic programming [48], temporal differences method[49], Q-learning [50]. The core idea of RL methods is that the action an agent takes in the next step is affected by the reward, which is the feedback of the environment to its current state. For example, positive feedback, like a high reward, will lead to the enhancement of its current action when encountering the same state the next time, and vice verse [51,52]. Motivated by this mechanism, works that combine reinforcement learning and evolutionary game for solving social dilemmas have attracted much research interest [53,54]. These studies have established a reasonable framework, under which the strategy (cooperation or defection), is similar to the action in RL, and the payoff can be regarded as the reward. The Bush-Mostelle (BM) model, another classic RL algorithm, enables agents to adaptively change their current actions according to their rewards from the current states, which is helpful to depict the self-regarding process [55,56]. Under the BM model, the larger the reward exceeds an agent's aspiration, the more possibly the agent will continue adopting its current action. Based on this, we could study the dynamical process of the evolutionary game with the adaptive interaction intensity. Specifically, agents could strengthen (or weaken) the current strategies (interaction intensities) when receiving a higher (or lower) payoff than aspiration.

In this paper, we focus on exploring the effect of the adaptive interaction intensity on cooperation in both structured and well-mixed populations. We also investigate the co-evolution progress where the interaction intensity of agents changes as their strategies update. The main contributions of this paper can be summarized as i) We propose an adaptive interaction intensity model to optimize population benefits and inspire agent behaviors. ii) We construct a valid dynamical pair approximation framework for adaptive interaction intensity, which fills the gap in this direction. iii) We demonstrate the self-organization behavior pattern in the interaction fabric, which reflects the internal factors of social harmony. In the following part, we first show that the introduction of adaptive interaction intensity improves the cooperative level of iterated PDG on square lattices. We analyze the influence of the adaptive interaction intensity on cooperation from the view of co-evolution processes and microscopic mechanisms. Furthermore, we use the pair approximation method to obtain an analytical result. Finally, we validate the robustness of the proposed mechanism on a well-mixed population.

2. Preliminaries

Game theory provides a framework to investigate human behaviors in interactions. In a two-agent two-strategy game, each agent has two optional strategies, cooperation (C) or defection (D). After the interaction, agents obtain payoffs according to the payoff matrix:

$$\begin{array}{c} \\ C \\ D \end{array} \begin{array}{c} \begin{array}{cc} C & D \end{array} \\ \begin{pmatrix} R & S \\ T & P \end{pmatrix} \end{array}, \tag{1}$$

where R, P, S, and T denote the reward for mutual cooperation, punishment for mutual defections, sucker's payoff, and the defective temptation when one agent chooses cooperation and the other chooses defection, respectively. These parameters decide the game type. If the parameters satisfy the condition $T > R > P > S$, it is a prisoner's dilemma game; if the parameters satisfy the condition $T > R > S > P$, it is a snowdrift game; if the parameters satisfy the condition $R > T > P > S$, it is a stag hunt game; if the parameters satisfy the condition $R > T$ and $S > P$, it is a coordinate game.

Every agent in a game pursues the maximum of its own payoff. Nash equilibrium refers to a strategy combination that contains the optimal strategies of two agents. The Nash equilibrium is (D, D) in prisoner's dilemma game, (C, D) and (D, C) in snowdrift game, (C, C) and (D, D) in stag hunt game, and (C, C) in coordinate game.

The Bush-Mostelle (BM) method, one of the classic reinforcement learning algorithms, describes the self-regarding process based on the current reward and action. It describes the probability that an agent takes the specific action next time. Assume agent's action (C or D) and reward in time t as a_t and r_t, the probability p_{t+1} of taking action C in $t+1$ is described as:

$$p_{t+1} = \begin{cases} p_t + (1-p_t)s_t, & a_t = C, s_t \geq 0 \\ p_t + p_t s_t, & a_t = C, s_t < 0 \\ p_t - p_t s_t, & a_t = D, s_t \geq 0 \\ p_t - (1-p_t)s_t, & a_t = D, s_t < 0 \end{cases}, \tag{2}$$

where s_t $(-1 < s_t < 1)$ is the stimulus in t and is defined as:

$$s_t = \tanh[\beta(r_t - A)], \tag{3}$$

where A is the aspiration level and β $(\beta > 0)$ is the sensitivity to $r_t - A$. Positive s_t will increase the probability of taking action C and reduce the probability of taking action D, while negative s_t will reduce the probability of taking action C and increase the probability of taking action D. Specially, when $t = 0, p_t$ obeys the uniform density on $[0, 1]$, independently for different agents.

3. Model and method

We consider evolutionary games on square lattices with periodic boundaries and Von Neumann neighborhood, where each vertex denotes an agent who can interact with four nearest neighbors Ω along edges. Agents play the pairwise prisoner's dilemma game with their neighbors. To simplify yet without losing generality, we

adopt the weak prisoner's dilemma game as our main model in the following discussions, where $S = P = 0, R = 1$ and $T = b$ $(b \geqslant 1)$.

Then, we define $p_{x\to y}$ as the willingness that agent x would like to interact with neighbor y. In our model, we assume the willingness of an agent to interact with different neighbors is independent, and so does the willingness of different agents. Thus, the interaction intensity, i.e., the probability that agent x and her neighbor y would successfully interact, can be denoted as:

$$p_{xy} = p_{x\to y} \times p_{y\to x}. \tag{4}$$

Based on this defination, p_{xy}, the interaction intensity between x and y, depends on the interacting willingness of x towards $y, p_{x\to y}$, and the interacting willingness of y towards $x, p_{y\to x}$. Note $p_{xy} = p_{yx}$ because of the symmetry. We assume that each pair of agents interacts following the interaction intensity with probability $1-\epsilon$, where ϵ denotes the probability that agents would randomly choose to interact or not. (In this paper, the parameter ϵ will be set as 0.1 for all simulations.)

For convenience, we divide the four nearest neighbors into two sets based on whether the interaction happens, including the interactive neighbors Ω_i and non-interactive neighbors Ω_n. In each Monte Carlo step, agent x plays the game with her interactive neighbors and obtains an accumulated payoff Π_x at the step t. With all these definitions, we can then describe the dynamics of our model as two processes. First, agent x updates strategy according to the "imitation rule", where uses Fermi function is a well-known method to calculate the imitation probability:

$$f_{y\to x} = \frac{1}{1+\exp\left[(\Pi_x - \Pi_y)/K\right]}, \tag{5}$$

where Π_y is the accumulated payoff of the randomly selected neighbor y and the noise K is set to be 0.1. Then, the interactive willingness of agent x towards y will adaptively change according to the BM model:

$$p_{x\to y}(t+1) = \begin{cases} p_{x\to y}(t) + \left(1-p_{x\to y}(t)\right)s_{xy}(t), a_{xy}(t) = 1, s_{xy}(t) \geq 0 \\ p_{x\to y}(t) + p_{x\to y}(t)s_{xy}(t), a_{xy}(t) = 1, s_{xy}(t) < 0 \\ p_{x\to y}(t) - p_{x\to y}(t)s_{xy}(t), a_{xy}(t) = 0, s_{xy}(t) \geq 0 \\ p_{x\to y}(t) - \left(1-p_{x\to y}(t)\right)s_{xy}(t), a_{xy}(t) = 0, s_{xy}(t) < 0 \end{cases}, \tag{6}$$

where $a_{xy} = 1$ and $a_{xy} = 0$ represent interaction and non-interaction between agent x and y at the step t, respectively, and $s_{xy}(t)$ denotes an stimulus that can be described as:

$$s_{xy}(t) = \tanh\left[\beta\left(r_{xy}(t) - A_x(t)\right)\right], \tag{7}$$

where $r_{xy}(t)$ is the payoff of agent x obtained from neighbor $y, A_x(t) = \Pi_x/4$ is the aspiration level averaged over the four nearest neighbors, and β represents the sensitivity of $s_{xy}(t)$ to $r_{xy}(t) - A_x(t)$. Finally, we can give detailed aspirations and payoffs of focal cooperators and defectors in different configurations in Fig. 1. Note that $r_{xy}(t) = 0$ when there is no interaction between two agents or the neighbor (opposite) y adopts the defection strategy, which means agents can only obtain a payoff from interaction with cooperative neighbors. In addition, when $a_{xy}(t) = 0, r_{xy} = 0$ and s_{xy} is either less than or equal to zero. However, the BM model normally contains four scenarios. Thus, without loss of generality, we still describe the whole four scenarios in Eq. (7).

Initially, each agent randomly chooses cooperation or defection with the same probability and her willingness to interact with each nearest neighbor subjects to a uniform distribution $[0, 1]$. To avoid the willingness being too small or large during the evolutionary process (the evolution process may be frozen), we set the minimum of the willingness as 0.01 and the maximum as 1. We conduct all the Monte-Carlo simulations (MCS) on a square lattice with $L = 300$ to obtain the evolution data. We take an average of the last 5,000 steps (with a total being 3×10^4) to represent the steady states, and average over ten independent simulation runs for a fixed set of parameter values.

4. Pair approximation

As shown in Fig. 2, a randomly selected 2-site configuration of a square lattice is the basic structure of the pair approximation method and it can contribute to the pair evolutionary dynamics, where the focal agent B is the randomly chosen nearest neighbor of the focal agent A in the strategy updating process. The payoff of A (Π_A) and B (Π_B) are determined by their nearest neighbors x, y, z, B and u, v, w, A, respectively. For example, if A adopts cooperation, B adopts defection, and A learns the strategy of B, as a result, the pair $c-c$ and $c-d$ will decrease while the pair $d-c$ and $d-d$ will increase.

In our model, the accumulated payoff of the focal agent depends on the actual interactive neighbors. In order to calculate the payoffs of focal agents, we need to know the interaction intensity of agents toward cooperators (interaction intensity towards defectors is omitted because it produces no payoffs). For a focal cooperator agent A, her neighbors are divided into interactive neighbors and non-interactive neighbors according to whether the interaction happens. Then, the overall $c-c$ link denotes the link between A and all her cooperator neighbors, and the actual $c-c$ link denotes the link between A and her interactive cooperator neighbors. Thus, for the whole population, we use the overall $c-c$ link $n_{c,c}(t)$ to denote links between the focal cooperator agents and their cooperator neighbors regardless of the interaction intensity, and the actual $c-c$ link $n_{c,c'}(t)$ to denote links between the focal cooperator agents and their interactive cooperator neighbors. In other words, $n_{c,c'}(t)$ is the subset of $n_{c,c}(t)$. In this paper, these parameters are obtained from the Monte-Carlo simulation results. Specifically, the ratio of actual interacting $c-c$ links $n_{c,c'}(t)$ to the overall $c-c$ links $n_{c,c}(t)$ in the last m steps of the total M steps is used to repre-

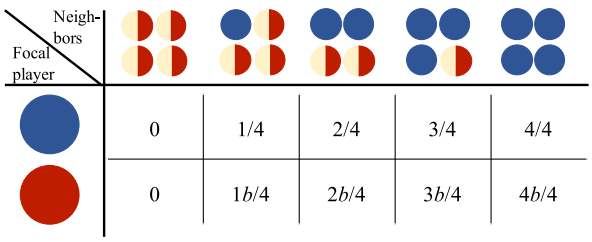


Fig. 1. Aspirations of cooperators and defectors. Aspiration is defined as the average payoffs of agents. Blue and red circles represent cooperators and defectors, respectively, and circles colored with both yellow and red represent non-interaction or defectors. Agents can only obtain payoffs by interaction with cooperators.

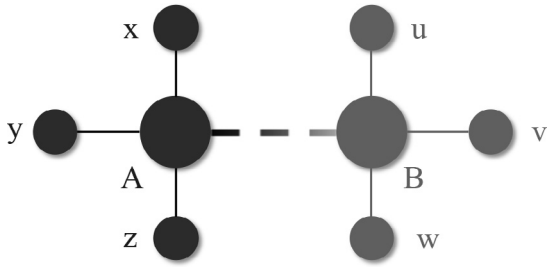


Fig. 2. 2-site configuration. A randomly selected 2-site configuration in square lattice, where A and B are two focal agents with nearest neighbors x, y, z, B and u, v, w, A, respectively. Agent A learns the strategy of agent B according to the "imitation rule".

sent the average interaction intensity of cooperators towards cooperators, which can be denoted as $\bar{I}_{c,c}$:

$$\bar{I}_{c,c} = \frac{\sum_{t=M-m}^{M} n_{c,c\prime}(t)}{\sum_{t=M-m}^{M} n_{c,c}(t)}. \tag{8}$$

In the same way, the average interaction intensity of defectors towards cooperators $\bar{I}_{d,c}$ can be written as:

$$\bar{I}_{d,c} = \frac{\sum_{t=M-m}^{M} n_{d,c\prime}(t)}{\sum_{t=M-m}^{M} n_{d,c}(t)}. \tag{9}$$

For a given 2-site configuration in which A is a cooperator and B is a defector, the focal agent A accumulates a payoff:

$$\Pi_c(x,y,z,\bar{I}_{c,c}) = \bar{I}_{c,c} \times n_c(x,y,z) \times 1, \tag{10}$$

and the focal agent B accumulates a payoff:

$$\Pi_d(u,v,w,\bar{I}_{d,c}) = \bar{I}_{d,c} \times n_c(u,v,w) \times b, \tag{11}$$

where $n_c(x,y,z)$ and $n_c(u,v,w)$ denote the number of cooperators among neighbors x,y,z and u,v,w of focal agent A and B, respectively. In a given configuration, when A is a defector and B is a cooperator, the focal agent A's payoff can be calculated similarly.

The pair evolution with interaction intensity is determined by:

$$\begin{aligned}\dot{p}_{c,c} = &\sum_{x,y,z}[n_c(x,y,z)+1]p_{d,x}p_{d,y}p_{d,z}\sum_{u,v,w}p_{c,u}p_{c,v}p_{c,w}f_{d(x,y,z,\bar{I}_{d,c})\to c(u,v,w,\bar{I}_{c,c})}\\ &+\sum_{x,y,z}[-n_c(x,y,z)]p_{c,x}p_{c,y}p_{c,z}\sum_{u,v,w}p_{d,u}p_{d,v}p_{d,w}f_{c(x,y,z,\bar{I}_{c,c})\to d(u,v,w,\bar{I}_{d,c})},\end{aligned} \tag{12}$$

$$\begin{aligned}\dot{p}_{c,d} = &\sum_{x,y,z}[1-n_c(x,y,z)+1]p_{d,x}p_{d,y}p_{d,z}\sum_{u,v,w}p_{c,u}p_{c,v}p_{c,w}f_{d(x,y,z,\bar{I}_{d,c})\to c(u,v,w,\bar{I}_{c,c})}\\ &+\sum_{x,y,z}[n_c(x,y,z)-2]p_{c,x}p_{c,y}p_{c,z}\sum_{u,v,w}p_{d,u}p_{d,v}p_{d,w}f_{c(x,y,z,\bar{I}_{c,c})\to d(u,v,w,\bar{I}_{d,c})},\end{aligned} \tag{13}$$

where $n_c(x,y,z)$ denotes the number of cooperators among neighbors x,y,z of the focal agent. Inserting Eq. (11) and (12), $f_{*\to *}$ subjects to the Fermi function as described in Eq. (6)

$$\begin{aligned}f_{d(x,y,z,\bar{I}_{d,c})\to c(u,v,w,\bar{I}_{c,c})} &= \frac{1}{1+\exp[(\Pi_d(x,y,z,\bar{I}_{d,c})-\Pi_c(u,v,w,\bar{I}_{c,c}))/K]}\\ f_{c(x,y,z,\bar{I}_{c,c})\to d(u,v,w,\bar{I}_{d,c})} &= \frac{1}{1+\exp[(\Pi_c(x,y,z,\bar{I}_{c,c})-\Pi_d(u,v,w,\bar{I}_{d,c}))/K]}.\end{aligned} \tag{14}$$

And $p_{c,c}$ denotes the frequency of overall $c-c$ links in the population, $p_{c,d}$ denotes the frequency of overall $c-d$ links in the population, $p_{d,c}$ denotes the frequency of overall $d-c$ links in the population, and $p_{d,d}$ denotes the frequency of overall $d-d$ links in the population. Note that we have omitted the multiplier factor $\frac{2p_{c,d}}{p_c^2 p_d^2}$ because it has no influence on the equilibrium. One can also see that $p_{c,d}=p_{d,c}$ and $p_{c,c}+p_{c,d}+p_{d,c}+p_{d,d}=1$ because of the symmetry and the natural constraint. Then, the frequency of cooperation can be acquired as $f_c = p_{c,c}+p_{c,d}$. Apparently, when the interaction intensity $\bar{I}_{*,*}$ equals to 1, it becomes the same case as the traditional pair approximation method. In other words, our model makes an extension of the previous work in terms of the evolutionary process and thus provides a more comprehensive framework [57].

5. Results

In weak prisoner's dilemma game, the defective temptation b is a pivotal parameter that influences the frequency of cooperation, while in the BM model, β plays an important role in deciding the stimulus, which is related to the dynamics of interaction intensity. So we start by exploring how the frequency of cooperation varies as a function of the parameter b and β. As shown in Fig. 3, obviously, cooperation is promoted by the introduction of the adaptive interaction intensity when compared to the traditional case (i.e. $\beta=0$). That is, the critical values at which cooperation becomes extinct are enlarged. In particular, cooperation is significantly enhanced when β is large. For instance, when $\beta=2$, defectors cannot dominate the population until b is larger than 1.5. While for small β, the enhancement of cooperation is relatively slight. For example, when $\beta=0.1$, cooperators can barely survive when b is around 1.05. In addition, it is worth noticing that there is a sharp transition from the full cooperation phase to the full defection phase in the population. Given a certain β, as b increases, cooperators first dominate the population (marked as the blue area in Fig.3), then transiently co-exist with defectors (marked as the narrow white area), and finally die out with the dominance of defectors (marked as the red area).

We then adopt the extended pair approximation method to theoretically demonstrate the promotion of cooperation and the sharp transition of the cooperation frequency. Take the cases of $\beta=0.1$ and $\beta=2$, where cooperation is fragile and competitive in face of the invasion of defection, as an example. In Fig. 4, we plot the cooperation frequency f_c as a function of b based on both simulation and theoretical results. It shows the same transition trends in both simulation and theoretical analysis, although there exists little difference in the specific values since the evolutionary process on lattice networks is very complicated and often beyond the consideration of the pair approximation method. Comparison between simulation and theoretical results also confirms that the adaptive interaction intensity has a beneficial effect on cooperation. Namely, large β extends the range of b where cooperators survive.

To evaluate the above-observed phenomena, Figs. 5 (β=2) and 6 (β=0.1) show the distribution of interactive neighbors and strategies as the MCS step increases. Several characteristic spatial pat-

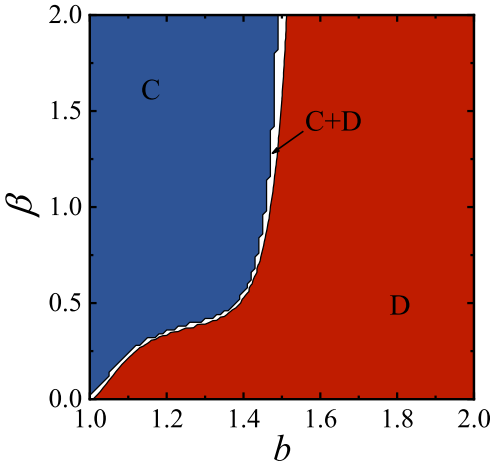


Fig. 3. Frequency of cooperation in $b-\beta$ panel. Steady state frequency of cooperation is plotted as a function of b and β. The panel is separated into three phases. Blue, red and white represent full cooperation phase, full defection phase and the mixed strategy phase, respectively. We obtain the results with a random distribution for strategies.

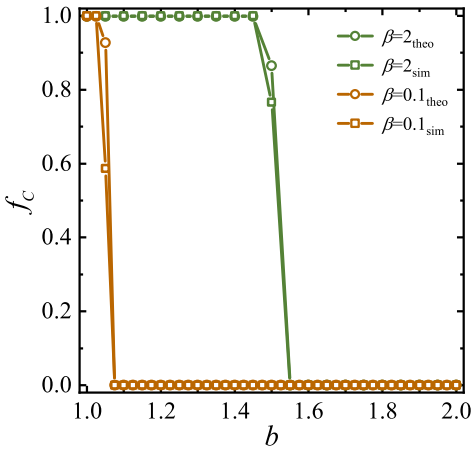


Fig. 4. Frequency of cooperation as a function of b with various β. Squares and circles represent the simulation and theoretical results, respectively. Orange and green represent results obtained with $\beta = 0.1$ and $\beta = 2$, respectively. We obtain the results with a random distribution for the strategy in simulations.

terns of the entire evolutionary process with a prepared initial distribution are used to demonstrate the microscopic evolutionary dynamics under the influence of adaptive interaction intensity. There are three types of agents on the network, i) cooperator (blue) with at least one interactive neighbor; ii) defector (red) with at least one interactive neighbor; iii) isolated agent (yellow) with no interactive neighbors. As shown in Figs. 5(a) and 6(a), the initially prepared distribution divides the whole population into cooperative and defective areas. Then, because of the randomly initialized interaction willingness of agents, there are scattered yellow spots in blue and red stripes. In addition, since the early evolution (the first two columns) of different b seems to be similar, only some representative snapshots for all cases are plotted. In particular, we can observe the enduring (END) period where cooperation endures the invasion of defection and the expanding (EXD) period where cooperation spreads [58,59].

For the case of $\beta = 2$, isolated agents surrounded by cooperators disappear very quickly in the early stage, which means they have built at least one interaction with their cooperative neighbors. However, in defective areas, isolated agents hardly establish interaction with neighbors. In a word, agents are more willing to interact with cooperators than defectors. This leads to a strong interaction intensity within the former and a weak interaction intensity within the latter. As evolution proceeds, there will be different snapshots for different cases of b. When $b = 1.45$, the EXP period that features the expanding of cooperation shows up, and cooperation finally dominates the population. While the END period that features the endurance of cooperation is popular when $b = 1.55$ [60]. $b = 1.5$ leads to the coexistence of cooperation and defection in the population. Of particular interest, the invasion of defection always brings isolated agents, which means that agents prefer to maintain interaction with cooperative neighbors instead of defective neighbors. In addition, agents are motivated to build new interactions with the cooperative neighbors and cut off the interaction with the defective neighbors. Therefore, during the evolution process, agents will dynamically adjust their interaction relationship with neighbors and gradually reach an optimal interaction intensity. Similar evolutionary patterns are reflected in the case of $\beta = 0.1$. But there exists some difference. Specifically, the evolutionary process is slower in the case of $\beta = 0.1$. For example,

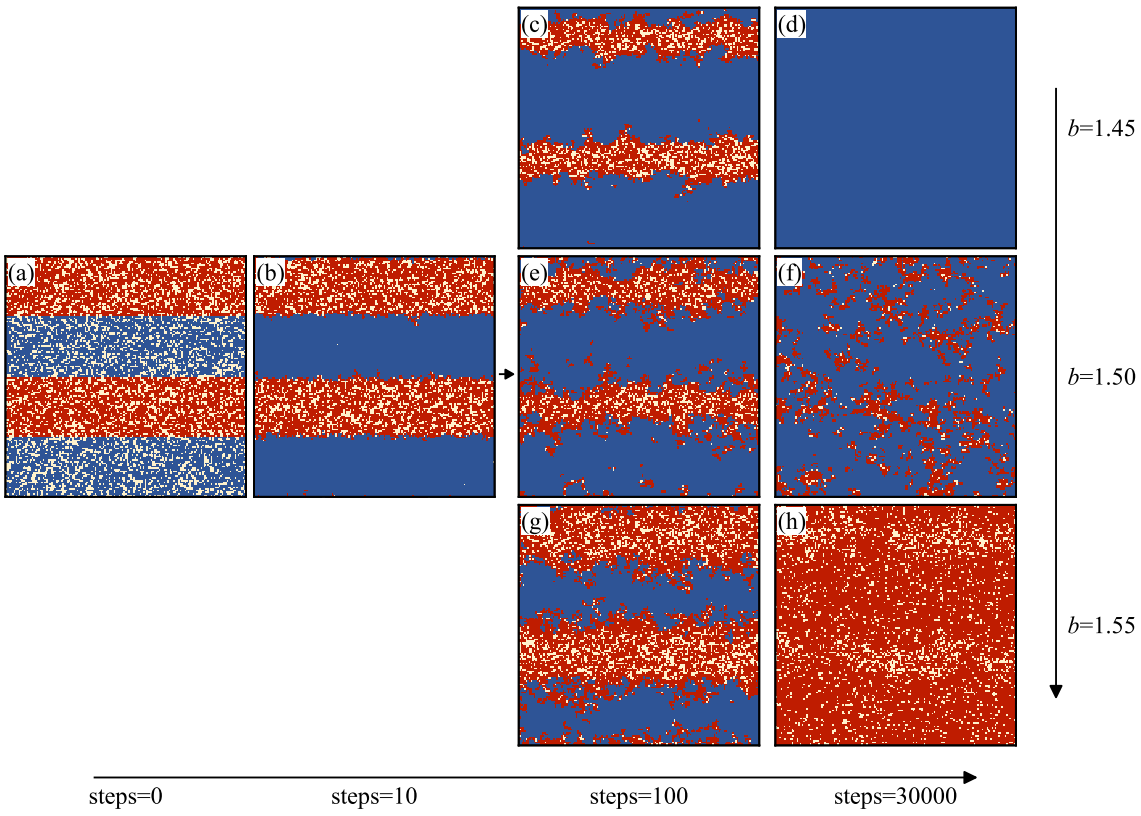


Fig. 5. Snapshots of the strategy distribution in evolutionary process. Snapshots are taken at step 0, 10, 100, and 30000, respectively with $b = 1.45$, $b = 1.5$ and $b = 1.55$. The layout follows the direction of arrows. Fist and second columns show the representative initial evolution process when $\beta = 2$ for different b. Blue, red, and yellow represent cooperators, defectors, and isolated agents, respectively.

from Fig. 6 (a) to (b), isolated agents surrounded by cooperators need more time to build interaction with neighbors. Furthermore, the invasion of defection causes less emergence of isolated agents, which means the behavior of cutting off the interaction with defectors becomes slower, too. These results imply that a small β will slow down the evolution process for agents to obtain the optimal interaction intensity.

To further explore the impact of the interaction intensity on the evolutionary process, Fig. 7 presents the evolution characteristic of cooperation and the interaction intensity. For different β, it is clear

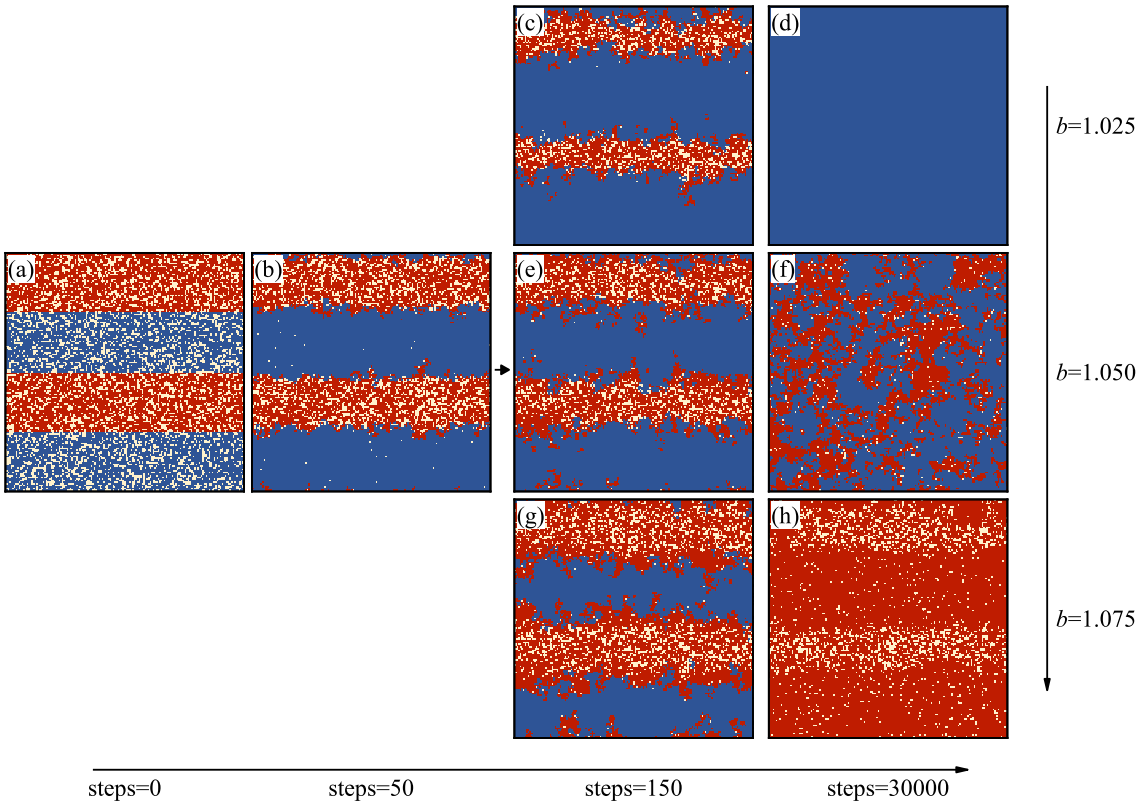


Fig. 6. Snapshots of the strategy distribution in evolutionary process. Snapshots are taken at step 0, 10, 150, and 30000, respectively with $b = 1.025, b = 1.05$ and $b = 1.075$. The layout follows the direction of arrows. Fist and second columns show the representative initial evolution process when $\beta = 0.1$ for different b. Blue, red, and yellow represent cooperators, defectors, and isolated agents, respectively.

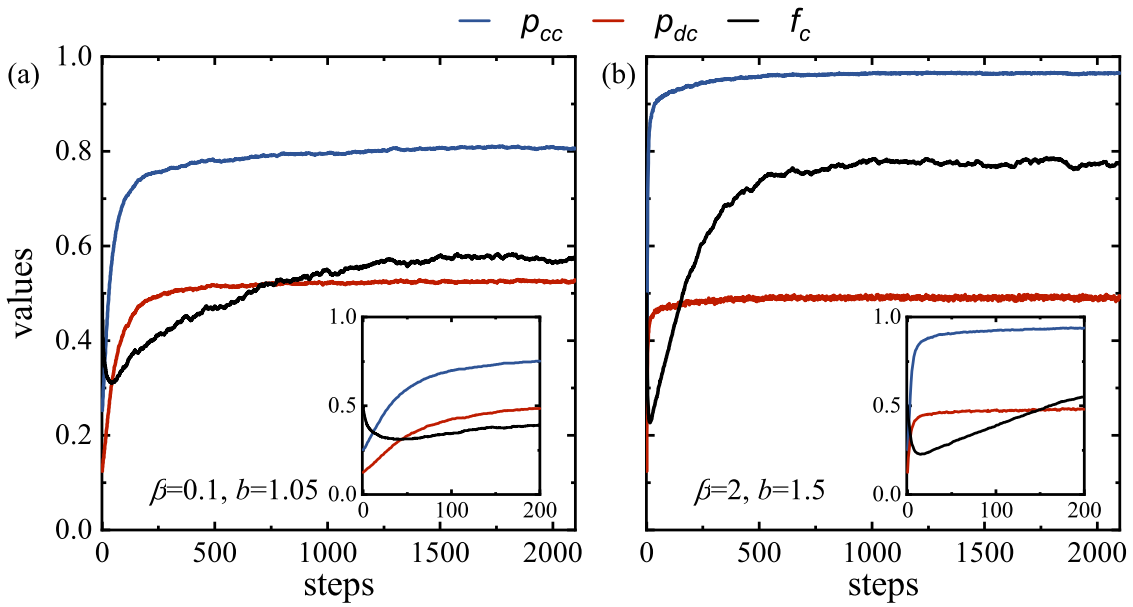


Fig. 7. Evolution of interaction intensity and cooperation frequency. The average values of the interaction intensity between cooperators p_{cc}, between defectors and cooperators p_{dc}, and the frequency of cooperation f_c evolve with time, which are represented by blue curve, red curve, and black curve, respectively. From left to the right, parameters are set as $\beta = 0.1, b = 1.05$ and $\beta = 2, b = 1.5$, respectively. The insets are the amplifications of the first 200 steps of whole evolution.

that the overall evolutionary patterns are similar, except for some notable differences. When β is small, the interaction intensity evolves relatively slowly. Consequently, it leads to the inability to provide the optimal interaction intensity for the evolution of cooperation, to effectively promote cooperation. However, when β is large, the interaction intensity evolves very fast, and it already reaches a steady state before the EXP period. This implies for large β, the population will efficiently establish an optimal interaction relationship. In addition, this optimal interaction relationship can benefit the cooperation of the population. These results are consistent with observations in Fig. 3. (See Fig. 8).

The number of interactive neighbors distribution for both cooperators and defectors is various for different cases of β and b. In the full cooperation phase (Fig. 8(a) and (d)), agents establish more interactions with their neighbors to obtain higher payoffs. While in the full defection phase (Fig. 8(c) and (f)), agents cannot receive payoffs whether they interact with neighbors or not, so they will finally form a random interaction relationship. It is worth noticing that in Fig. 8(b) and (e), where the cooperators and defectors coexist at a steady-state, the distributions vary for different β. The number of interactive neighbors of cooperators under $\beta = 2$ is larger than the case of $\beta = 0.1$. On the contrary, the number of interactive neighbors of defectors under $\beta = 2$ is smaller than the case of $\beta = 0.1$. This indicates that a relatively larger β is more beneficial for cooperation. More importantly, one can find when β is large, the distribution of both cooperation and defection is closer to that in the phases of full cooperation and defection, respectively. This means when β is large, the interaction intensity evolves faster, so agents will efficiently establish a steady interaction relationship that is closer to the optimal interaction intensity. Thus, it becomes more clear that the interaction intensity is enhanced by cooperation, which in return promotes cooperation in the population.

At last, we test whether the model presents a similar effect on the well-mixed network. Fig. 9 shows the cooperation frequency at the steady-state on a well-mixed network with 25 nodes. Initially, cooperators and defectors distribute randomly with the same probability, and the interaction willingness follows a uniform distribution. Since the final cooperation frequency will always be 0 or 1, the cooperation frequency is an average of 50 simulations for each combination of parameters β and b. The introduction of the

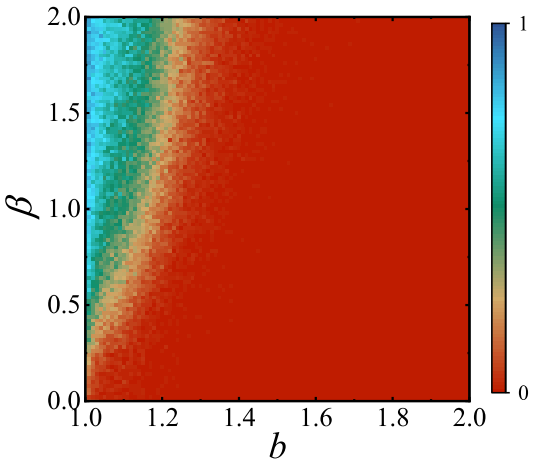


Fig. 9. Frequency of cooperation in $b - \beta$ panel. The panel shows the average cooperation frequency of 50 independent simulations as a function as b and β. Red represents full defection phase in all 50 simulations, and blue represents full cooperation phase in all 50 simulations. The initial interaction willingness follows the uniform distribution in $[0, 1]$.

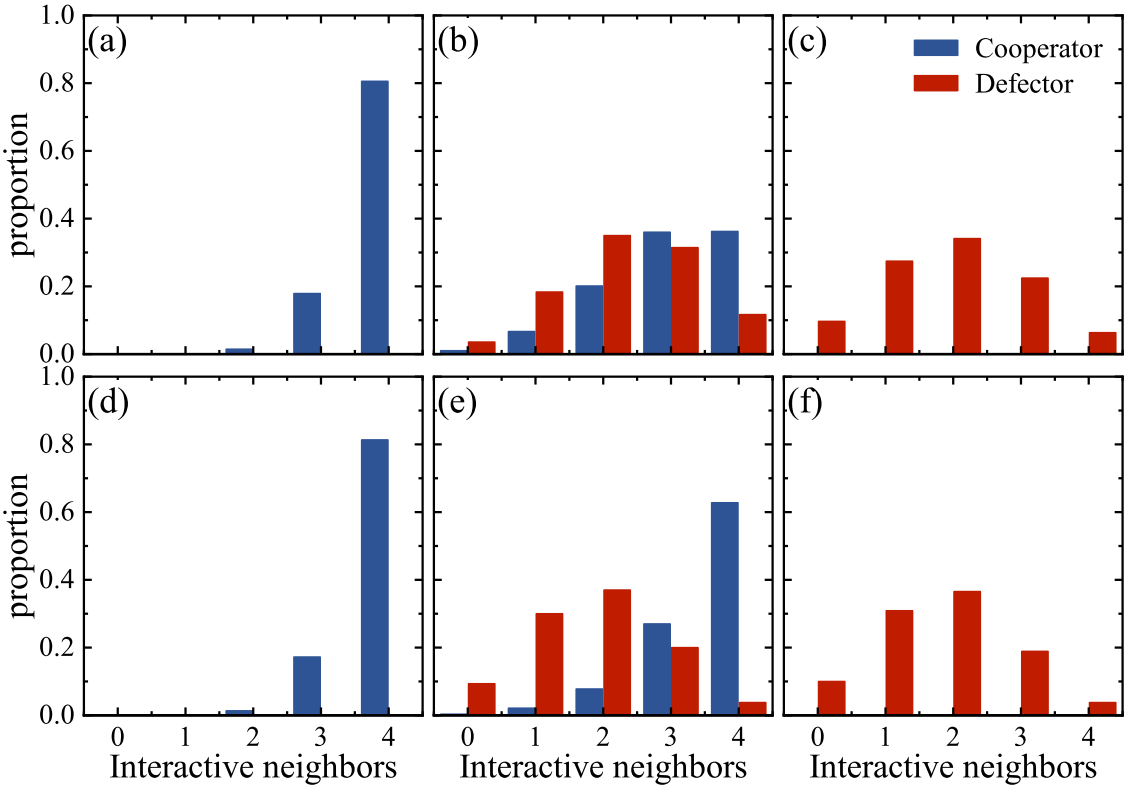


Fig. 8. Distribution of interactive neighbors of cooperators and defectors. Parameters are set as $\beta = 0.1$, and b = 1.025, 1.05, 1.075 in (a)(b)(c), while $\beta = 2$, and b = 1.45, 1.5, 1.55 in (d)(e)(f). Results of cooperators and defectors are denoted by blue and red columns, respectively.

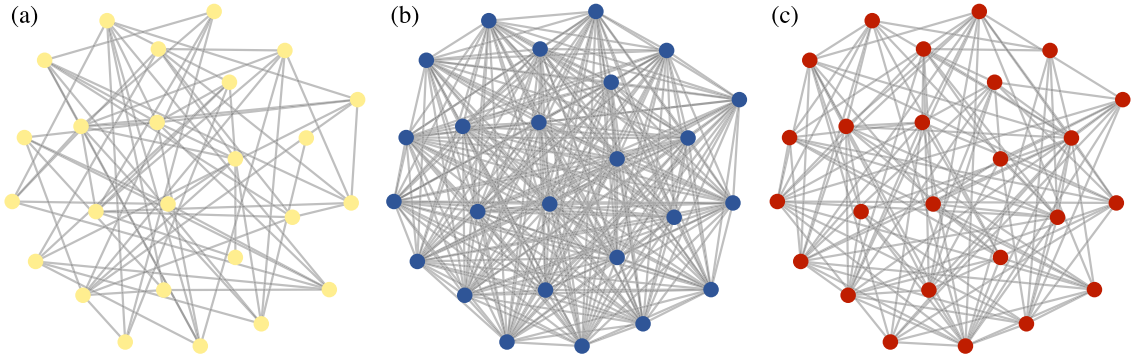


Fig. 10. Illustrations of the well-mixed networks. (a), (b), and (c) represent the initial network, network in full cooperation phase, and network in full defection phase, respectively. Grey lines denote the interaction between agents. Yellow nodes denote agents at the initial state regardless of their strategy. Blue and red nodes denote cooperators and defectors, respectively. We obtain the results with parameters $\beta = 2, b = 1.2$ in (b), and $b = 1.6$ in (c).

adaptive interaction intensity improves cooperation. Moreover, as β becomes large, the critical value at which cooperation goes extinct is also enlarged.

The interaction networks of the full cooperation phase and full defection phase are given in Fig. 10 (b) and (c), respectively. As the initial interaction willingness of agents follows a uniform distribution, there will only be a quarter of edges having successful interactions in the initial network, as shown in Fig. 10(a). In the full cooperation phase, nearly all pairs between two cooperators have successful interactions. While in the full defection phase, the interactions are heavily sparse. Therefore, the results on both lattice and well-mixed networks prove the interaction intensity is enhanced between cooperators, and the optimal interaction intensity will promote cooperation.

6. Discussion

We investigate the evolutionary process of the iterated prisoner's dilemma game with adaptive interaction intensity based on the reinforcement learning method. Different from the traditional evolution process, agents with adaptive interaction intensity can decide whether to interact with nearest neighbors or not. The results demonstrate that the adaptive interaction intensity enhances the cooperation in the population and is beneficial for solving social dilemmas. The evolution of the interaction intensity and cooperation is affected by the sensitivity parameter of the BM model. When this parameter is small, the interaction intensity and the strategy will mutually influence each other for a relatively long period. On the contrary, when the parameter is large, the interaction intensity will fast evolve to an optimal point, which enlarges the critical value of cooperation becoming extinct. In other words, with a relatively large sensitivity, the adaptive interaction intensity makes cooperation more competitive to survive. In addition, we obtain a universal conclusion in the well-mixed population, which excludes the influence of network structure. That is, cooperative behavior enhances interaction, which conversely improves cooperation in groups. The conclusions we obtained are also theoretically confirmed by the extended pair approximation.

Inspired by the reality that the interactions between agents change over time, we use the BM model to realize the adaptive interaction intensity. The simulation and theoretical results demonstrate adaptive interaction intensity can promote cooperation greatly. One may find that since this mechanism allows agents to timely terminate the interactions with poor-behaving neighbors, it leads to the appearance of isolated agents who have no interactive neighbors. However, we would stress that timely stopping the interaction with poor-behaving neighbors and establishing new interactions with good-behaving neighbors still is a wise choice in the long term. Because it promotes cooperation in the population, which is beneficial for not only agents themselves but also the whole population [57].

Our work is a simple attempt to introduce adaptive behavior into evolutionary games. In addition to the adaptability of the agents' interacting behavior, it is also worth considering the environmental interventions and constraints in the further works. Along the line of our work, there could be some future endeavors. For example, agents might be able to switch the type of the game according to self-regarding skills and judgments (e.g., Q-learning process [46]). We hope this work will be an inspiring exploration that sheds light on solving social dilemmas.

CRediT authorship contribution statement

Zhao Song: Software, Formal analysis, Investigation, Visualization, Writing - original draft. **Hao Guo:** Software, Writing - review & editing. **Danyang Jia:** Writing - review & editing. **Matjaž Perc:** Writing - review & editing. **Xuelong Li:** Methodology, Conceptualization, Writing - review & editing. **Zhen Wang:** Supervision, Writing - review & editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

This research was supported by the National Science Fund for Distinguished Young Scholars (No. 62025602), the National Natural Science Foundation of China (Nos. 11931015, U1803263 and 81961138010), Fok Ying-Tong Education Foundation, China (No. 171105), Technological Innovation Team of Shaanxi Province (No. 2020TD-013), Fundamental Research Funds for the Central Universities (No. D5000211001), the Slovenian Research Agency (Grant Nos. P1-0403 and J1-2457) and the Tencent Foundation and XPLORER PRIZE.

References

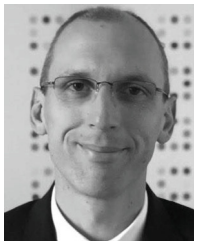
[1] J. Hofbauer, K. Sigmund, et al., Evolutionary games and population dynamics, Cambridge University Press, 1998.
[2] J. Tanimoto, Fundamentals of evolutionary game theory and its applications, Springer, 2015.

[3] G.S. Van Doorn, M. Taborsky, The evolution of generalized reciprocity on social interaction networks, Evolution: International Journal of Organic Evolution 66 (3) (2012) 651–664.
[4] M. Perc, A. Szolnoki, Coevolutionary games-a mini review, BioSystems 99 (2) (2010) 109–125.
[5] J. Tanimoto, Evolutionary games with sociophysics, Evolutionary Economics.
[6] M. Perc, J.J. Jordan, D.G. Rand, Z. Wang, S. Boccaletti, A. Szolnoki, Statistical physics of human cooperation, Physics Reports 687 (2017) 1–51.
[7] Z. Wang, S. Kokubo, M. Jusup, J. Tanimoto, Universal scaling for the dilemma strength in evolutionary games, Physics of life reviews 14 (2015) 1–30.
[8] J. Tanimoto, H. Sagara, Relationship between dilemma occurrence and the existence of a weakly dominant strategy in a two-player symmetric game, BioSystems 90 (1) (2007) 105–114.
[9] M.A. Nowak, Evolutionary dynamics: exploring the equations of life, Harvard University Press, 2006.
[10] M.A. Nowak, Five rules for the evolution of cooperation, science 314 (5805) (2006) 1560–1563.
[11] H. Ohtsuki, C. Hauert, E. Lieberman, M.A. Nowak, A simple rule for the evolution of cooperation on graphs and social networks, Nature 441 (7092) (2006) 502–505.
[12] J.M. Pacheco, A. Traulsen, M.A. Nowak, Coevolution of strategy and structure in complex networks with dynamical linking, Physical review letters 97 (25) (2006) 258103.
[13] S. Meloni, A. Buscarino, L. Fortuna, M. Frasca, J. Gómez-Gardeñes, V. Latora, Y. Moreno, Effects of mobility in a population of prisoner's dilemma players, Physical Review E 79 (6) (2009) 067101.
[14] E. Lieberman, C. Hauert, M.A. Nowak, Evolutionary dynamics on graphs, Nature 433 (7023) (2005) 312–316.
[15] X. Xu, Z. Rong, Z. Tian, Z. Wu, Timescale diversity facilitates the emergence of cooperation-extortion alliances in networked systems, Neurocomputing 350 (2019) 195–201.
[16] Z. Wang, A. Szolnoki, M. Perc, Self-organization towards optimally interdependent networks by means of coevolution, New Journal of Physics 16 (3) (2014) 033041.
[17] Z. Wang, L. Wang, A. Szolnoki, M. Perc, Evolutionary games on multilayer networks: a colloquium, The European physical journal B 88 (5) (2015) 1–15.
[18] C. Xia, X. Li, Z. Wang, M. Perc, Doubly effects of information sharing on interdependent network reciprocity, New Journal of Physics 20 (7) (2018) 075005.
[19] K. Huang, X. Zheng, Z. Li, Y. Yang, Understanding cooperative behavior based on the coevolution of game strategy and link weight, Scientific reports 5 (1) (2015) 1–7.
[20] L. Wang, Z. Wang, Y. Zhang, X. Li, How human location-specific contact patterns impact spatial transmission between populations?, Scientific reports 3 (1) (2013) 1–10
[21] H. Li, L. Wang, Multi-scale asynchronous belief percolation model on multiplex networks, New Journal of Physics 21 (1) (2019) 015005.
[22] X. Chen, L. Wang, Promotion of cooperation induced by appropriate payoff aspirations in a small-world networked game, Physical Review E 77 (1) (2008) 017103.
[23] Z. Wang, L. Wang, M. Perc, Degree mixing in multilayer networks impedes the evolution of cooperation, Physical Review E 89 (5) (2014) 052813.
[24] H. Guo, D. Jia, I. Sendiña-Nadal, M. Zhang, Z. Wang, X. Li, K. Alfaro-Bittner, Y. Moreno, S. Boccaletti, Evolutionary games on simplicial complexes, arXiv preprint arXiv:2103.03498.
[25] K. Huang, Y. Cheng, X. Zheng, Y. Yang, Cooperative behavior evolution of small groups on interconnected networks, Chaos, Solitons & Fractals 80 (2015) 90–95.
[26] Z. Rong, X. Li, X. Wang, Roles of mixing patterns in cooperation on a scale-free networked game, Physical Review E 76 (2) (2007) 027101.
[27] C. Xia, Q. Miao, J. Wang, S. Ding, Evolution of cooperation in the traveler's dilemma game on two coupled lattices, Applied Mathematics and Computation 246 (2014) 389–398.
[28] H. Guo, Z. Song, S. Geček, X. Li, M. Jusup, M. Perc, Y. Moreno, S. Boccaletti, Z. Wang, A novel route to cyclic dominance in voluntary social dilemmas, Journal of the Royal Society Interface 17 (164) (2020) 20190789.
[29] A. Szolnoki, M. Perc, Second-order free-riding on antisocial punishment restores the effectiveness of prosocial punishment, Physical Review X 7 (4) (2017) 041027.
[30] C. Xia, S. Meloni, Y. Moreno, Effects of environment knowledge on agglomeration and cooperation in spatial public goods games, Advances in Complex Systems 15 (supp01) (2012) 1250056.
[31] C. Xia, S. Ding, C. Wang, J. Wang, Z. Chen, Risk analysis and enhancement of cooperation yielded by the individual reputation in the spatial public goods game, IEEE Systems Journal 11 (3) (2016) 1516–1525.
[32] K. Huang, T. Wang, Y. Cheng, X. Zheng, Effect of heterogeneous investments on the evolution of cooperation in spatial public goods game, PloS one 10 (3) (2015) e0120317.
[33] Z. Rong, Z. Wu, D. Hao, M.Z. Chen, T. Zhou, Diversity of timescale promotes the maintenance of extortioners in a spatial prisoner's dilemma game, New Journal of Physics 17 (3) (2015) 033032.
[34] Z. Rong, Z. Wu, G. Chen, Coevolution of strategy-selection time scale and cooperation in spatial prisoner's dilemma game, EPL (Europhysics Letters) 102 (6) (2013) 68005.
[35] A. Antonioni, M.P. Cacault, R. Lalive, M. Tomassini, Know thy neighbor: Costly information can hurt cooperation in dynamic networks, PLoS One 9 (10) (2014) e110788.
[36] M.A. Amaral, M.A. Javarone, Heterogeneous update mechanisms in evolutionary games: mixing innovative and imitative dynamics, Physical Review E 97 (4) (2018) 042305.
[37] Z. Song, H. Guo, D. Jia, M. Perc, X. Li, Z. Wang, Third party interventions mitigate conflicts on interdependent networks, Applied Mathematics and Computation 403 (2021) 126178.
[38] L. Wang, X. Li, Y. Zhang, Y. Zhang, K. Zhang, Evolution of scaling emergence in large-scale spatial epidemic spreading, PloS one 6 (7) (2011) e21197.
[39] A. Szolnoki, M. Perc, Seasonal payoff variations and the evolution of cooperation in social dilemmas, Scientific reports 9 (1) (2019) 1–9.
[40] X. Chen, F. Fu, L. Wang, Interaction stochasticity supports cooperation in spatial prisoner's dilemma, Physical Review E 78 (5) (2008) 051120.
[41] Z. Qin, F. Khawar, T. Wan, Collective game behavior learning with probabilistic graphical models, Neurocomputing 194 (2016) 74–86.
[42] B. Woelfing, A. Traulsen, Stochastic sampling of interaction partners versus deterministic payoff assignment, Journal of Theoretical Biology 257 (4) (2009) 689–695.
[43] A. Traulsen, M.A. Nowak, J.M. Pacheco, Stochastic payoff evaluation increases the temperature of selection, Journal of theoretical biology 244 (2) (2007) 349–356.
[44] S. Zhang, J. Dong, L. Liu, Z. Huang, L. Huang, Y. Lai, Artificial intelligence meets minority game: toward optimal resource allocation, arXiv preprint arXiv:1802.03751.
[45] L. Cui, X. Wang, Y. Zhang, Reinforcement learning-based asymptotic cooperative tracking of a class multi-agent dynamic systems using neural networks, Neurocomputing 171 (2016) 220–229.
[46] R.S. Sutton, A.G. Barto, Reinforcement learning: An introduction, MIT press, 2018.
[47] D. Michie, D.J. Spiegelhalter, C. Taylor, et al., Machine learning, Neural and Statistical Classification 13 (1994) (1994) 1–298.
[48] R. Bellman, Dynamic programming, Science 153 (3731) (1966) 34–37.
[49] R.S. Sutton, Learning to predict by the methods of temporal differences, Machine learning 3 (1) (1988) 9–44.
[50] L. Yang, Q. Sun, D. Ma, Q. Wei, Nash q-learning based equilibrium transfer for integrated energy management game with we-energy, Neurocomputing 396 (2020) 216–223.
[51] D. Jia, H. Guo, Z. Song, L. Shi, X. Deng, M. Perc, Z. Wang, Local and global stimuli in reinforcement learning, New Journal of Physics 23 (8) (2021) 083020.
[52] L. Kraemer, B. Banerjee, Multi-agent reinforcement learning as a rehearsal for decentralized planning, Neurocomputing 190 (2016) 82–94.
[53] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, nature 518 (7540) (2015) 529–533.
[54] S. Hu, C. Leung, H. Leung, Modelling the dynamics of multiagent q-learning in repeated symmetric games: a mean field theoretic approach, Advances in Neural Information Processing Systems 32 (2019) 12125–12135.
[55] T. Ezaki, Y. Horita, M. Takezawa, N. Masuda, Reinforcement learning explains conditional cooperation and its moody cousin, PLoS computational biology 12 (7) (2016) e1005034.
[56] Y. Horita, M. Takezawa, K. Inukai, T. Kita, N. Masuda, Reinforcement learning accounts for moody conditional cooperation behavior: experimental results, Scientific reports 7 (1) (2017) 1–10.
[57] J. Li, C. Zhang, Q. Sun, Z. Chen, J. Zhang, Changing the intensity of interaction based on individual behavior in the iterated prisoner's dilemma game, IEEE Transactions on Evolutionary Computation 21 (4) (2016) 506–517.
[58] R. Kümmerli, C. Colliard, N. Fiechter, B. Petitpierre, F. Russier, L. Keller, Human cooperation in social dilemmas: comparing the snowdrift game with the prisoner's dilemma, Proceedings of the Royal Society B: Biological Sciences 274 (1628) (2007) 2965–2970.
[59] Y. Heller, E. Mohlin, Observations on cooperation, The Review of Economic Studies 85 (4) (2018) 2253–2282.
[60] Z. Wang, S. Kokubo, J. Tanimoto, E. Fukuda, K. Shigaki, Insight into the so-called spatial reciprocity, Physical Review E 88 (4) (2013) 042145.

Zhao Song received the B.S. degree in automation from Northwestern Polytechnical University, China, in 2018. She is currently working toward the Ph.D. degree with the School of Mechanical Engineering and Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, China. Her current research interests include game theory and reinforcement learning.

Hao Guo received the B.S. degree from Changchun University, China, in 2015, and M.S. degree from Yunnan University of Finance and Economics, China, in 2018. He received the Ph.D. degree from Northwestern Polytechnical University, China, in 2022. His research interests include game theory and reinforcement learning.

Danyang Jia received the B.S. degree from Ningbo University, China, in 2015, and M.S. degree from Yunnan University of Finance and Economics, China, in 2018. She received the Ph.D. degree from Northwestern Polytechnical University, China, in 2022. Her research interests include evolutionary game theory, behavioral science, and reinforcement learning.



Matjaž Perc is currently a Professor of physics with the University of Maribor, Maribor, Slovenia. He is a Member of Academia Europaea and the European Academy of Sciences and Arts, and among top 1% most cited physicists according to 2020 Clarivate Analytics data. He was also the 2015 recipient of the Young Scientist Award for Socio and Econophysics from the German Physical Society, and the 2017 USERN Laureate. In 2018, he was also the recipient of Zois Award, which is the highest national research Award in Slovenia. In 2019, he became the Fellow of the American Physical Society.

Xuelong Li is currently a Full Professor with the School of Computer Science and Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China.

Zhen Wang received the Ph.D. degree from Hong Kong Baptist University, Hong Kong, China, in 2014. He is currently a Professor with the school of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an, China. He has authored/coauthored over 100 research papers and four review papers with 16000+ citations. His current research interests include complex networks, evolutionary game, and data science.