SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Emergent Leader-follower Relationship in Networked Multiagent Systems

Jinzhuo Liu[1,3,5], Chenyou Fan[1], Yunchen Peng[1], Jinze Du[1], Zhen Wang[3,4*] & Chen Chu[2,3*]

[1]*School of Software, Yunnan University, Kunming* 650504*, Yunnan, China;*
[2]*School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming* 650221*, Yunnan, China;*
[3]*School of Artificial Intelligence, Optics and Electronics(iOPEN), Northwestern Polytechnical University, Xi'an* 710072*, Shaanxi, China;*
[4]*School of Cybersecurity, Northwestern Polytechnical University, Xi'an* 710072*, Shaanxi, China;*
[5]*Engineering Research Center of Cyberspace, Yunnan University, Kunming* 650504*, Yunnan, China*

Appendix A The prisoner's dilemma game with voluntary participation

The payoff matrix of the prisoner's dilemma game with voluntary participation is shown in Table A1.

Table A1 A general form of prisoner's dilemma games with voluntary participation

strategy	cooperation	defection	loner
cooperation	R, R	S, T	δ, δ
defection	T, S	P, P	δ, δ
loner	δ, δ	δ, δ	δ, δ

In our model, without loss of generality, δ is fixed at 0.3 [1].

Appendix B The mixed-population concurrent evolution framework

Each agent becomes a RLP or EGP with probability ρ and $1-\rho$, respectively. The RLP explores the environment to find the best strategy, and the EGP only imitates the neighbour's strategy; this neighbour can be a RLP or EGP. The calculation of both benefits is the same, and the difference between the two agents lies in the decision model. The RLP decides strategy according to the Q vector regarding what it learned, and the EGP only imitates the neighbour's strategy. At each time step of the MC simulation, r_i denotes the payoff of agent i, referring to the following equation:

$$r_i = \sum_{j=1}^{k} P_{ij}, \tag{B1}$$

where k is the number of neighbors and P_{ij} denotes the payoff that agent i obtained from a neighbor j. To ensure that the system has reached steady state, our iteration step is $t = 3 * 10^4$. Then, we use the lattice $L = 300$, learning rate $\eta = 0.9$, greedy factor $\epsilon = 0.02$ and noise $K = 0.1$ [2]. In addition, for each group of parameter values, the last $5 * 10^3$ steps of each dataset are averaged to ensure proper accuracy.

The evolutionary process is presented in Algorithm B1. At the beginning of the game, initialize a square lattice of size $L * L$ with periodic boundaries and neighbours. The probability of any agent becoming RLP of EGP is ρ, $1-\rho$, respectively (line 1); this distribution will not change during the evolutionary process. Each player selects cooperation, defection, and loner strategies with the same probability (line 2). At the start of MC simulation (line 3), the decision-making process of RLPs and EGPs are in lines (6-10) and lines (12-14). We calculate the frequency of every strategy (C, D, L) in real time at the end of each MC step (line 17).

Appendix C Snapshots for RLPs

Some typical evolutionary snapshots of clusters are shown in Figure C1. RLPs are usually myopic [3], which means that their best strategy tends to converge to the loner strategy with a fixed benefit. Snapshots for $\rho = 0$ show that the system freezes in the traditional situation [1], the inherent cyclic dominance of the strategies (C→D→L→C), resulting in self-organizing polydomain structures on square lattices. However, in the mixed-population, cooperation becomes the favourite strategy. This means that a combination of myopic RLPs and strategy-imitation based EGPs produces fascinating dynamics that significantly promote cooperation.

* Corresponding author (email: zhenwang0@gmail.com, chuchenynufe@hotmail.com)

Algorithm B1 A mixed-population concurrent evolution framework

Require: a set of $L * L$ agents, a set of A available strategies, the MC step T, the ratio of RLP ρ, a matrix game G, the number k of neighbours per agent

Ensure: frequency of cooperation, defection and loner FC, FD, FL, respectively

initialize agents' role (RLP or EGP) $\leftarrow$according to ρ
the strategy of the agent is assigned cooperation, defection, and loner with the same probability
while $t < T$ **do**
  for each agent $i = 0 \to L * L$ **do**
    if agent i is an RLP **then**
      agent i selects a strategy $a \in A$ according to its policy;
      plays the game G with k neighbours using the selected strategy;
      receives an immediate payoff r_t^i and updates its Q-value:
      $Q_{t+1}^i(a) = (1-\eta) Q_t^i(a) + \eta r_t^i$
      updates its policy using Equation (2)
    else
      agent i plays the game G with k neighbours;
      receives an immediate payoff r_t^i;
      randomly selects a neighbour agent j and imitates the strategy of j with the probability from Equation (3)
    end if
  end for
  calculate (FC, FD, FL)
  $t \leftarrow$t +1
end while

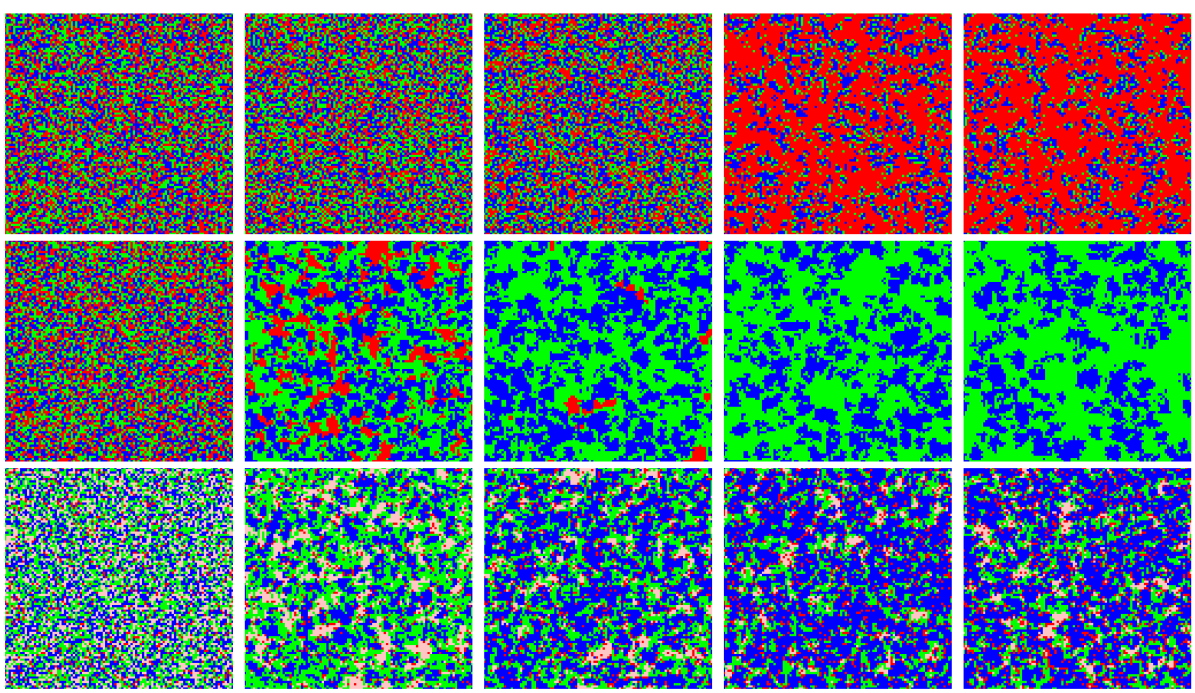

Figure C1 Typical snapshots of the distribution of strategies in MC steps 0, 10, 100, 1000, and 10000. All results are obtained for $b = 1.02$. From the top to bottom, ρ is equal to 1, 0 and 0.25, respectively. Cooperation, defection and loner are coloured blue, green and red. In particular, in the snapshots of $\rho = 0.25$, we distinguish the populations of the loner strategy; RLP loners and EGP loners are red and flesh-coloured, respectively.

Appendix D Game dynamics of RLPs

The dominant strategy of RLPs is loner after training, regardless of whether the RLPs are in the pure leader population or the mixed-population (Figure D1).

We have studied the interspecific interaction of RLP strategy pairs (Figure D2a). In the entire population, the loner strategy is dominant (frequency of loner>50%) and unaffected by the frequency of RLPs. The loner strategy is essential for the maintenance of cooperation in the population. This is because the proportion of loner-cooperation (LC) pairs is the highest among the pairs that include cooperation (LC, CC, CD). There is a weak peak about the ratio of loner-cooperation (LC and CL) pairs changing with the ratio of RLPs, which occurs similarly among cooperation-cooperation (CC) pairs. We also find that RLPs' strategy converges much faster than EGPs' strategy. Specifically, most RLPs maintain their strategies after 5 iterations, while 20% of EGPs always change their strategies after even 400 iterations. This can be explained as follows: EGPs select their strategies using the imitation rule, which means that they will change their strategies with a significant probability when they interact with individuals with different strategies, while RLPs adopt their strategies based solely on their own Q-values, which are much more stable.

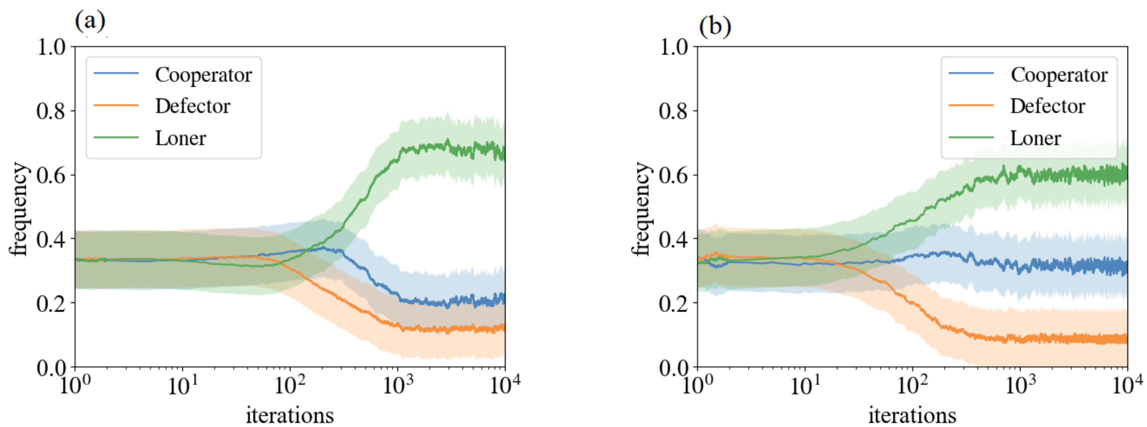

Figure D1 Time series of the frequency of the cooperation, defection and loner strategies of RLPs; panel (a) and panel (b) depict parameter $\rho = 0$ and $\rho = 0.25$. All results are obtained for $b = 1.02, K = 0.1$.

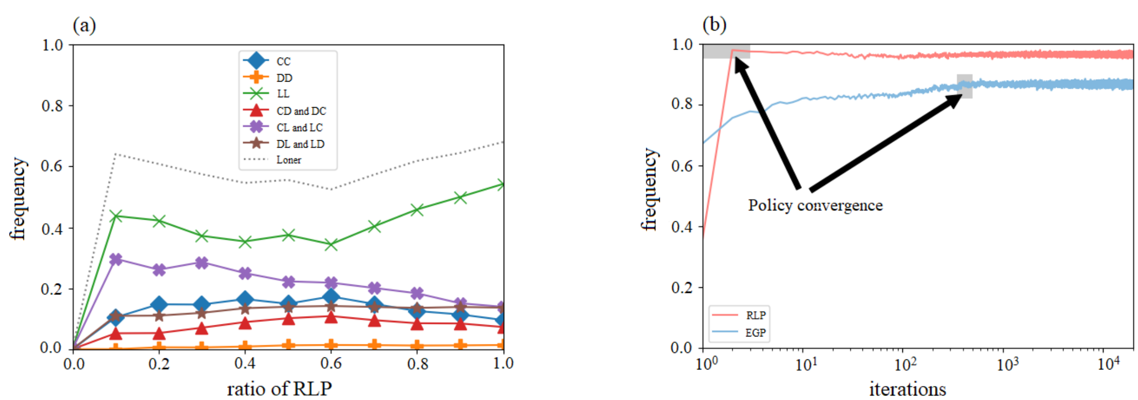

Figure D2 Panel (a) shows the time series of the frequency of the cooperation, defection and loner strategies for RLPs. Panel (b) shows the frequency of individuals that fix their strategies in the two populations when $\rho = 0.25$. All results are obtained for $b = 1.02, K = 0.1$.

Appendix E Game dynamics of EGPs

Compared with the results of the population completely filled with EGPs (Figure E1b), cooperation is significantly promoted in the mixed-population ($\rho = 0.25$), while defection is suppressed. As previously discussed, the frequency of cooperation of RLPs in the mixed-population is 36% (Figure D1b), which is much lower than the frequency of EGPs (58%). Then, it is obvious that there should be more cooperative EGPs in the mixed-population, which is consistent with the result shown in Figure E1c that the frequency of cooperation of EGPs is significantly higher (68%) than the total frequency of cooperation in the population (58%).

To better reveal the processes behind the results shown in (Figure E2), we analyse the representative cooperation scenarios. Examining the pairwise interplay reveals that raising the ratio of RLPs brings a high frequency of cooperation in the mixed-population but only up to an exact threshold, after which the loner strategy again dominates (Figure E2a). The introduction of RLPs completely changes the dominance cyclic of strategies, and defection and loner become more cooperative. Finally, the combination of RLPs and EGPs results mainly in loner-cooperation pairs, the frequency of which is hump-shaped (Figure E2c). When we combine Figure D2a and Figure E2, the results demonstrate a twofold role of RLPs.

Table D1 We summarize the probability distribution of RLPs' cooperative feedback strategy in the case of all permutations and combinations of neighbour strategies. We can use this metric to evaluate the degree of reciprocity displayed by RLPs. All results are obtained for $b = 1.02, K = 0.1$.

C	D	L	f_c($\rho = 1$)	f_c($\rho = 0.25$)
4	0	0	4.561%	31.179%
3	1	0	11.634%	20.598%
3	0	1	9.774%	16.745%
2	1	1	17.684%	10.788%
2	2	0	7.914%	8.046%
2	0	2	12.958%	4.327%
1	2	1	7.088%	4.297%
1	1	2	12.296%	2.222%
1	3	0	0.043%	0.829%
1	0	3	11.459%	0.518%
0	3	1	0.038%	0.177%
0	2	2	0.153%	0.133%
0	1	3	0.557%	0.059%
0	0	4	3.834%	0.059%
0	4	0	0%	0.014%

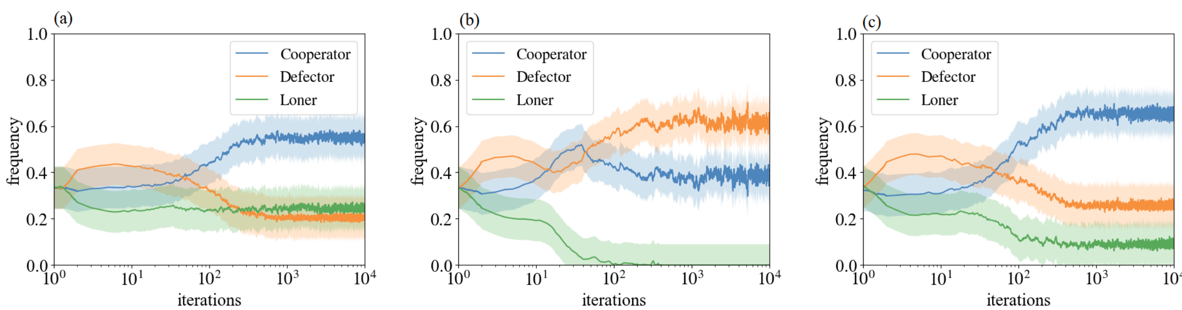

Figure E1 Time series of the frequency of different strategies. Panel (a) shows the time series of three strategies in the mixed-population ($\rho = 0.25$), while panel (b) shows the results in the population for $\rho = 0$. Furthermore, panel (c) shows the results for EGPs in the mixed-population ($\rho = 0.25$). All results are obtained for $b = 1.02, K = 0.1$.

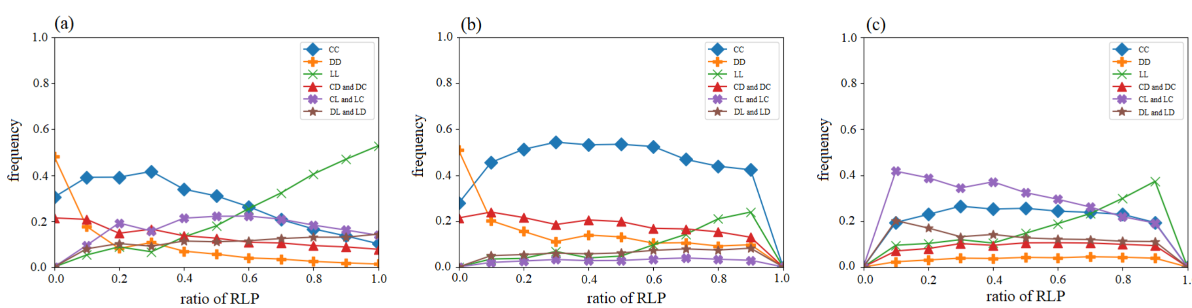

Figure E2 Analysis of pairwise interactions between (a) any two individuals, (b) individuals belonging to the same game dynamic, and (c) individuals belonging to the different game dynamics. Here, CC (DD, LL) denotes pairs of cooperation (defection, loner). In addition, CD (CL,DL) denotes pairs of cooperation and defection (cooperation and defection, defection and loner) and vice versa (DC, LC, LD). All results are obtained for $b = 1.02, K = 0.1$.

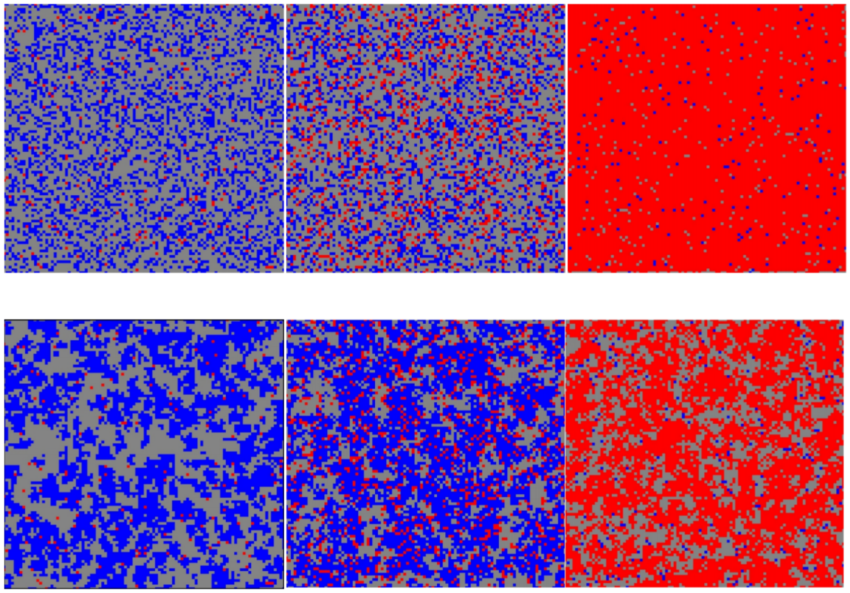

Figure E3 Typical snapshots of the distribution of strategies in MC steps 1 and 30,000. For all ρ parameters, we replace RLPs by zealots (players which always choose the loner strategy), from the left to right, for parameter $\rho = 0.02$, $\rho = 0.25$, and $\rho = 0.98$, respectively. The difference between the snapshots in the row is the MC step t, which is 1 and 30,000. The cooperation of the EGPs is marked in blue, zealots are marked in red, and the remaining strategies are marked in grey.

Regarding the snapshot in Figure C1 where the MC step equals 10,000 for $\rho = 0.25$, we find that the clusters of cooperation have irregular shapes, exhibit scattered relative positions, and are surrounded by sporadic loners. To investigate the underlying cause of the emergent cooperation in the mixed-population, we reconstruct an environment in which individuals that always choose loner (zealots) replace RLPs. As shown in Figure E3, under the condition of introducing a small number of RLPs ($\rho = 0.02$), cooperation is not fully promoted, and other strategies still occupy the dominant position in the population. For the case of $\rho = 0.25$, sporadic zealots safeguard a small area to suppress the emergence of defection, and then the cooperative EGPs around the zealots are stimulated by the zealots and form clusters to protect themselves from the invasion of defection. Finally, the mixed-population ultimately converges to the steady state in which the cooperation of EGPs reaches a high frequency and cooperation spontaneously dominates the population. However, we can see that a high ratio of zealots ($\rho = 0.98$) that always choose loner sabotages the emergence of cooperation.

Appendix F Dynamics of the mixed-population

We first consider the dynamics of EGPs. Although the traditional replication dynamic equation can solve the evolutionary dynamics of the well-mixed structure, it cannot accurately predict the evolution of a structured population because it ignores the network structure. Here, we introduce pair approximation to model the dynamics of EGPs. Instead of considering the frequency of strategies as in well-mixed populations, pair approximation tracks the frequencies of strategy pairs. The given probability $p_{s,s'}$ is to find a player with strategy s and an adjacent player with strategy s', where $s, s' \in \{c, d, l\}$. To track the changes in the frequencies of all pairs in the system, we need to determine $p_{cc}, p_{dd}, p_{ll}, p_{cd}, p_{cl}, p_{dl}$ (considering the symmetry condition $p_{s,s'} = p_{s',s}$). Under restrictions $p_{cc} + p_{dd} + p_{ll} + 2p_{cd} + 2p_{cl} + 2p_{dl} = 1$, we can approximate the dynamics of the three-strategy game by two differential equations:

$$
\begin{aligned}
\overset{\bullet}{\rho}_{i,i} = {} & \frac{2p_{i,j}}{{F^i_{EGP}}^3 {F^j_{EGP}}^3} \left\{ \begin{array}{l} \sum\limits_{x,y,z} [n_i(x,y,z)+1]\, p_{j,x}p_{j,y}p_{j,z} \sum\limits_{u,v,w} p_{i,u}p_{i,v}p_{i,w} W\left[P_j(x,y,z) \leftarrow P_i(u,v,w)\right] \\ -\sum\limits_{x,y,z} n_i(x,y,z) p_{i,x}p_{i,y}p_{i,z} \sum\limits_{u,v,w} p_{j,u}p_{j,v}p_{j,w} W\left[P_i(x,y,z) \leftarrow P_j(u,v,w)\right] \end{array} \right\} \\
& + \frac{2p_{i,k}}{{F^i_{EGP}}^3 {F^k_{EGP}}^3} \left\{ \begin{array}{l} \sum\limits_{x,y,z} [n_i(x,y,z)+1] p_{k,x}p_{k,y}p_{k,z} \sum\limits_{u,v,w} p_{i,u}p_{i,v}p_{i,w} W\left[P_k(x,y,z) \leftarrow P_i(u,v,w)\right] \\ -\sum\limits_{x,y,z} n_i(x,y,z) p_{i,x}p_{i,y}p_{i,z} \sum\limits_{u,v,w} p_{k,u}p_{k,v}p_{k,w} W\left[P_i(x,y,z) \leftarrow P_k(u,v,w)\right] \end{array} \right\},
\end{aligned} \tag{F1}
$$

$$
\begin{aligned}
\overset{\bullet}{\rho}_{i,j} = {} & \frac{2p_{i,j}}{{F^i_{EGP}}^3 {F^j_{EGP}}^3} \left\{ \begin{array}{l} \sum\limits_{x,y,z} \left[\frac{2-2n_i(x,y,z)-n_k(x,y,z)}{2}\right] p_{j,x}p_{j,y}p_{j,z} \sum\limits_{u,v,w} p_{i,u}p_{i,v}p_{i,w} W\left[P_j(x,y,z) \leftarrow P_i(u,v,w)\right] \\ +\sum\limits_{x,y,z} \left[\frac{2n_i(x,y,z)-4+n_k(x,y,z)}{2}\right] p_{i,x}p_{i,y}p_{i,z} \sum\limits_{u,v,w} p_{j,u}p_{j,v}p_{j,w} W\left[P_i(x,y,z) \leftarrow P_j(u,v,w)\right] \end{array} \right\} \\
& + \frac{2p_{i,k}}{{F^i_{EGP}}^3 {F^k_{EGP}}^3} \left\{ \begin{array}{l} \sum\limits_{x,y,z} [n_j(x,y,z)/2] p_{k,x}p_{k,y}p_{k,z} \sum\limits_{u,v,w} p_{i,u}p_{i,v}p_{i,w} W\left[P_k(x,y,z) \leftarrow P_i(u,v,w)\right] \\ -\sum\limits_{x,y,z} [n_j(x,y,z)/2]\, p_{i,x}p_{i,y}p_{i,z} \sum\limits_{u,v,w} p_{k,u}p_{k,v}p_{k,w} W\left[P_i(x,y,z) \leftarrow P_k(u,v,w)\right] \end{array} \right\} \\
& + \frac{2p_{j,k}}{{F^j_{EGP}}^3 {F^k_{EGP}}^3} \left\{ \begin{array}{l} \sum\limits_{x,y,z} [n_i(x,y,z)/2] p_{k,x}p_{k,y}p_{k,z} \sum\limits_{u,v,w} p_{j,u}p_{j,v}p_{j,w} W\left[P_k(x,y,z) \leftarrow P_j(u,v,w)\right] \\ -\sum\limits_{x,y,z} [n_i(x,y,z)/2]\, p_{j,x}p_{j,y}p_{j,z} \sum\limits_{u,v,w} p_{k,u}p_{k,v}p_{k,w} W\left[P_j(x,y,z) \leftarrow P_k(u,v,w)\right] \end{array} \right\}
\end{aligned} \tag{F2}
$$

In equations (F1) and (F2), $n_s(x, y, z)$ is the number of players using strategy s in neighbouring players' set (x, y, z). Here, $F_s = \sum_{s'} p_{s,s'}$ is the frequency of each particular strategy s, where s' again runs over the set of all possible strategies. Note that

F_s establishes the formal connection between the mean-field theory and the pair approximation by converting pair configuration probabilities $p_{s,s'}$ to approximations of configuration probabilities of large clusters, thus warranting at least qualitatively identical results of both approaches. $W\left[P_i\left[x,y,z\right] \leftarrow P_j\left[u,v,w\right]\right]$ is the strategy adoption function to be specified above by Equation (3). Moreover, $P_s\left(x,y,z\right)$ denotes the player adopting strategy s interacting with neighbours adopting strategies (x,y,z) plus a player adopting strategy s'. For details regarding the derivation of each differential equation, we refer the reader to [4] and [5], where the pair approximation method is accurately described. Moreover, by considering the interaction between RLPs and EGPs, we have further improved the calculation of payoff as a weighted sum $\bar{P}_s\left(x,y,z\right)$ as follows:

$$\bar{P}_s(x,y,z) = (1-\rho)P_s(x,y,z) + \rho P^{i}_{RLP}, \tag{F3}$$

where ρ is the ratio of RLPs. Then, we consider the dynamics of RLPs. For simplicity, we assume that all RLPs share a Q-vector. Then, the dynamics of RLPs can be modelled by the following three equations:

$$\begin{cases} r_t(a_j) = \rho P^{a^t_j}_{RLP} + (1-\rho)P^{a^t_j}_{EGP} \\ Q_{t+1}(a_j) = (1-\eta)Q_t(a_j) + \eta r_t(a_j) \\ F^j_{RLP} = \frac{e^{\tau Q_t(a_j)}}{\sum_{\forall a \in A} e^{\tau Q_t(a)}} \end{cases}. \tag{F4}$$

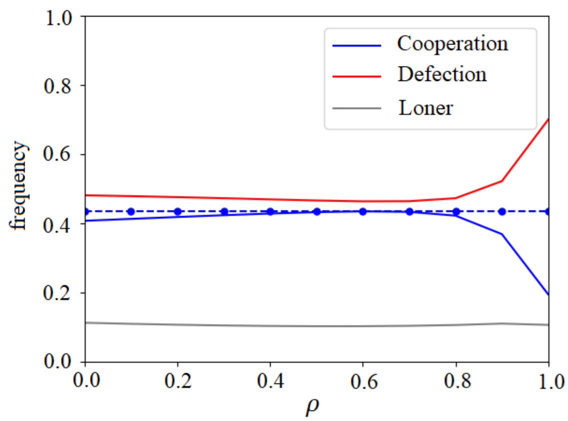

Figure F1 Result from the pair approximation numerical solutions. Frequency of cooperation with different ratios of RLP ρ using the modeled dynamics. As ρ increases, the frequency rises then falls and thus has a peak value (which is highlighted by the dashed line) when ρ is equal to 0.6. The result is obtained for $b = 1.02$, $K = 0.1$.

The result of the approximate calculation is consistent with that using MC simulation. Although the result of the approximate calculation has some quantitative differences from that of MC simulation because our framework for the dynamics ignores the diversity of Q values of each player, the result is similar to the MC simulation where there is a peak in the frequency of cooperation.

References

1 Szabó G, Hauert C. Evolutionary prisoner's dilemma games with voluntary participation. Physical Review E, 2002, 66(6): 062903.
2 Tanimoto J. Impact of deterministic and stochastic updates on network reciprocity in the prisoner's dilemma game. Physical Review E, 2014, 90(2): 022105.
3 Hu S, Leung C-W, Leung H-F. Modelling the dynamics of multiagent q-learning in repeated symmetric games: a mean field theoretic approach. Advances in Neural Information Processing Systems, 2019, 32: 12125-12135.
4 Hauert C, Szabó G. Game theory and physics. American Journal of Physics, 2005, 73(5): 405-414.
5 Guo H, Song Z, Geček S, et al. A novel route to cyclic dominance in voluntary social dilemmas. Journal of the Royal Society Interface, 2020, 17(164): 20190789.