SCIENCE CHINA
Information Sciences

• LETTER •

December 2023, Vol. 66 229201:1–229201:2
https://doi.org/10.1007/s11432-022-3741-3

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

Received 8 July 2022/Revised 14 December 2022/Accepted 10 February 2023/Published online 20 November 2023

Citation Liu J Z, Fan C Y, Peng Y C, et al. Emergent leader-follower relationship in networked multiagent systems. Sci China Inf Sci, 2023, 66(12): 229201, https://doi.org/10.1007/s11432-022-3741-3

Understanding the collective dynamics of leader-follower systems is a crucial and challenging issue. Previous studies have a priori designated leaders. However, the leader-follower relationship is self-organized in human society over time and social interactions [1]. Evolutionary game theory [2] provides an effective framework for understanding the emergence of cooperation. Some mechanisms are proposed to promote cooperation, such as reputation [3]. Recently, several studies have focused on the emergence of cooperation through reinforcement learning game dynamics [4]. Additionally, the dynamic process in other learning algorithms has also been studied, such as regret minimization [5].

In this study, we consider the settings with these two kinds of game players and describe them by proposing a new population game. To study this new game framework, we used the prisoner's dilemma game with voluntary participation. Furthermore, we demonstrated that the proposed mechanism produces a leader-follower relationship using the Monte Carlo (MC) simulation. The reinforcement learning players can perform like "leaders" in encouraging cooperation due to the ability to learn reciprocity from the environment, while the evolutionary game theory players perform like "followers" and become more cooperative by imitating cooperative leaders.

Model. There are two agents and three actions in the prisoner's dilemma game with voluntary participation: overcooperation (C), defection (D), and loner (L). In each round, the agent chooses an action and obtains a timely profit based on the joint actions of two agents. Cooperation yields a reward R, whereas mutual defection results in a punishment P. The cooperator obtains the sucker's payoff S, and the defector obtains the temptation T in the mixed-population case. For simplicity, we take $T = b$, $R = 1$, $P = 0$, and $S = 0$. Therefore, the payoff can be described by the payoff matrix with only one parameter b $(1 \leqslant b \leqslant 2)$. $\delta \in (0, 1)$ denotes the payoff of the risk-averse loner and its opponent. Appendix A shows the payoff matrix.

The RLP is the agent in this game, who makes decisions based on Q-learning. The game ends when agent i chooses an action a_j and receives an immediate payoff r_t^i at each time step t. Therefore, there is no state transition. The RLP updating equation based on Q-learning is as follows:

$$Q_{t+1}^i(a_j) = (1-\eta)\, Q_t^i(a_j) + \eta r_t^i, \tag{1}$$

where we use ε-greedy as the RLP's exploration policy. The policy $\pi_i(s)$ for agent i given current state s is defined as

$$\pi_i(s) = \begin{cases} \arg\max_{a \in \mathrm{A}_i} Q_i(s,a), & \text{with probability } 1-\varepsilon, \\ U(\mathrm{A}_i), & \text{with probability } \varepsilon, \end{cases} \tag{2}$$

where $U(\mathrm{A}_i)$ denotes a sample from the uniform distribution over the action space.

The agent who uses the "imitation rule" is the evolutionary game player (EGP). The probability that EGP i imitates the strategy of the other agent j depends on the following:

$$W = \frac{1}{1+\exp((r_i - r_j)/K)}, \tag{3}$$

where the parameter K measures the extent of noise. r_i and r_j are the profits of the EGP and the neighbor.

Each agent is initialized on an $L \times L$ square lattice with periodic boundaries and neighbors. At each time step of the MC simulation, each agent plays the game with its neighbors and updates its strategy based on its role (i.e., RLP or EGP). We calculated the frequency of every strategy (C, D, and L) at the end of each MC step. The detailed concurrent evolution framework is presented in Appendix B.

Results. Figure 1 shows that the frequency of cooperation (fc) is higher in the mixed population than in the pure population. Although defection is a nonexclusive strategy

---

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

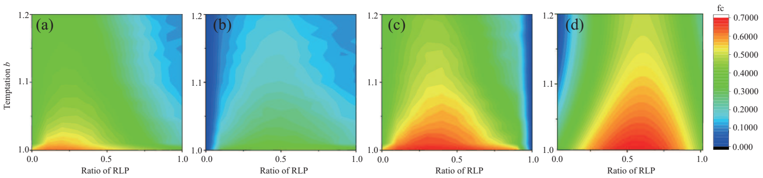

Figure 1 Frequency of cooperation (fc) is represented as a function of temptation b and the ratio of RLPs in four situations: (a) mixed-population, (b) subpopulation of RLPs, (c) subpopulation of EGPs, and (d) mixed-population in the BA scale-free network. The BA network consists of 1000 nodes with an average degree of 8.

in a pure population of EGPs, introducing a slight number of RLPs into the environment is sufficient to stimulate cooperation. Further introduction of RLPs makes cooperation a dominant behavior, but only while their density remains moderate. Specifically, cooperation peaks at a density of RLPs of approximately 25% in the total population and gradually declines afterward but it never returns to zero (Figure 1(a)). Similar phenomena can also be observed in both RLPs and EGPs subpopulations (Figures 1(b) and (c)). This indicates that a combination of two dynamics makes individuals more cooperative, regardless of whether they are RLPs or EGPs. Similar results were obtained on the Barabási-Albert (BA) scale-free network, which is illustrated in Figure 1(d). The frequency of cooperation also tends to increase first and then decrease. To further clarify the results above, some snapshots during the evolution process are presented in Appendix C. We separately analyzed the dynamics of RLPs and EGPs to study the underlying causes.

RLPs as leaders for cooperation. For RLPs, the dominant strategy is loner after training, and the frequency of cooperation in the mixed-population (36%) is significantly higher than in the pure RLP population (20%). By summarizing the RLPs' cooperative feedback strategy across all strategy combinations of neighbors, we found that reciprocity has emerged among RLPs, meaning they can feedback more cooperative behaviors when facing more cooperation. The probability of cooperation gradually decreases from 31.179% to 0.014% (see Table D1) as the surrounding cooperative neighbors decrease. The loner strategy is essential for the maintenance of cooperation in the population because the proportion of loner-cooperation (LC) pairs is the highest among the pairs that include cooperation (LC, CC, and CD). A detailed analysis of RLPs can be found in Appendix D.

EGPs as followers for cooperation. For EGPs, cooperation is also significantly promoted in mixed-population. However, the frequency of cooperation of EGPs is 58%, which is significantly higher than the frequency of RLPs (36%), implying that there are more cooperative EGPs in the population. By examining the strategy pair between different subpopulations, we found that raising the ratio of RLPs brings a high frequency of CC pairs in the mixed-population but only up to an exact threshold, after which the LL pairs increased, which means the loner strategy again dominates. Moreover, the frequency of CC pairs is highest among individuals belonging to the same game dynamic, while LC and CL pairs are the highest among individuals belonging to different game dynamics. Therefore, the RLP plays a twofold role which serves as an interspecific loner and a leader for EGP cooperation while EGPs serve as followers. We further analyzed the effect of the loner strategy and found that the loner strategy can safeguard a small area to suppress the emergence of defection, and then the cooperative EGPs around the loner strategy are stimulated by the loner strategy and form clusters to protect themselves from the invasion of defection. Appendix E contains a detailed analysis.

Overall, an EGP adopts an action by imitating its neighbors, while an RLP updates its Q-value by evaluating its reward obtained by interacting with the environment and all opponents. An RLP adopts an action just because the Q-values of the action adopted are higher than those of actions not adopted. Thus, an RLP explores itself and thus performs like the leader, while an EGP imitates and thus performs like the follower.

Dynamics of the mixed-population. To show how cooperation evolves in the population, we theoretically model the dynamics of EGPs and RLPs using pair approximation. The result of the approximate calculation is consistent with the previous MC simulation. Detailed dynamics equation and the result are presented in Appendix F.

Conclusion. A new mixed-population game dynamics was studied in the prisoner's dilemma game with voluntary participation. We found that the leader-follower relationship emerges among the population, and EGPs are the main participant in cooperation, while RLPs perform like leaders who encourage EGPs to cooperate. There is an optimal value of the ratio of the two-game dynamics that promotes cooperation.

Acknowledgements This work was supported by National Science Fund for Distinguished Young Scholarship of China (Grant No. 62025602), National Natural Science Foundation of China (Grant Nos. 11931015, 11971421, 12271471, 62066045, 62366058, U22B2036), Excellent Youths Project for Basic Research of Yunnan Province (Grant No. 202101AW070015), Science Foundation of Yunnan Province (No. 202001BB050063), Open Foundation of Key Laboratory in Software Engineering of Yunnan Province (Grant No. 2020SE407), Fok Ying-Tong Education Foundation, China (Grant No. 171105), and the Tencent Foundation and XPLORER PRIZE.

Supporting information Appendixes A–F. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Côté S, Lopes P N, Salovey P, et al. Emotional intelligence and leadership emergence in small groups. Leadership Q, 2010, 21: 496–508
2 Weibull J W. Evolutionary Game Theory. Cambridge: MIT Press, 1997
3 Pfeiffer T, Tran L, Krumme C, et al. The value of reputation. J Royal Soc Interface, 2012, 9: 2791–2797
4 Chu C, Li Y, Liu J, et al. A formal model for multiagent Q-learning dynamics on regular graphs. In: Proceedings of the 31st International Joint Conference on Artificial Intelligence, 2022. 194–200
5 Wang Z, Mu C J, Hu S Y, et al. Modelling the dynamics of regret minimization in large agent populations: a master equation approach. In: Proceedings of the 31st International Joint Conference on Artificial Intelligence, 2022. 534–540