Nonlinear Dyn (2022) 108:1837–1845
https://doi.org/10.1007/s11071-022-07289-7

ORIGINAL PAPER

Lévy noise promotes cooperation in the prisoner's dilemma game with reinforcement learning

Lu Wang · Danyang Jia · Long Zhang · Peican Zhu · Matjaž Perc · Lei Shi · Zhen Wang

Received: 17 September 2021 / Accepted: 7 February 2022 / Published online: 8 March 2022

Abstract Uncertainties are ubiquitous in everyday life, and it is thus important to explore their effects on the evolution of cooperation. In this paper, the prisoner's dilemma game with reinforcement learning subject to Lévy noise is studied. Specifically, diverse fluctuations mimicked by Lévy distributed noise are reflected in the payoff matrix of each player. At the same time, the self-regarding Q-learning algorithm is considered as the strategy update rule to learn the behavior that achieves the highest payoff. The results show that not only does Lévy noise promote the evolution of cooperation with reinforcement learning, it does so comparatively better than Gaussian noise. We explain this with the iterative updating pattern of the self-regarding Q-learning algorithm, which has an accumulative effect on the noise entering the payoff matrix. It turns out that under Lévy noise, the Q-value of cooperative behavior becomes significantly larger than that of defective behavior when the current strategy is defection, which ultimately leads to the prevalence of cooperation, while this is absent with Gaussian noise or without noise. This research thus unveils a particular positive role of Lévy noise in the evolutionary dynamics of social dilemmas.

Keywords Evolutionary dynamics · Prisoner's dilemma · Cooperation · Self-regarding Q-learning · Lévy noise

L. Wang · D. Jia
School of Mechanical Engineering and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China

L. Zhang · P. Zhu
School of Computer Science and School of Artificial Intelligence, Optics and Electronics (iOPEN), Northwestern Polytechnical University, Xi'an 710072, China

M. Perc
Faculty of Natural Sciences and Mathematics, University of Maribor, Maribor, Slovenia

M. Perc
Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404332, Taiwan

M. Perc
Complexity Science Hub Vienna, Vienna, Austria

M. Perc
Alma Mater Europaea, Maribor, Slovenia

L. Shi
School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming 650221, China

Z. Wang (✉)
School of Mechanical Engineering, School of Artificial Intelligence, Optics and Electronics (iOPEN), and School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, China
e-mail: nkzhenwang@163.com

1 Introduction

In practice, cooperative behaviors widely exist even though individuals are inherently selfish. This seems to contradict with Darwinian selection theory [1], where any behavior producing benefits to theirs but not directly to oneself will soon disappear. Hence, studying the emergence of cooperative behaviors among selfish individuals becomes an interesting and challenging problem which has attracted enormous attentions of various scholars. Prisoner's dilemma game, as a general metaphor for interpreting cooperation behaviors, has been well studied [2,3]. In this basic model, two players are supposed to choose cooperation (C) or defection (D) strategy simultaneously. Their payoffs are determined as follows: each player will get a reward R if both sides cooperate, or a punishment P if both sides defect. On the contrary, the sucker's payoff S and the temptation payoff T will be given to the cooperator and defector, respectively. The following prerequisites, i.e., $T > R > P > S$ and $2R > T + S$, should be met to ensure the nature of the game [4,5], which implies that defection is the best strategy irrespective of the opponent's choice. The result is that defection behavior spreads among all players, known as the social dilemma.

To aim to this dilemma, many theoretical and experimental studies have been conducted to maintain cooperation. The pioneering discovery is the effect of spatial structure on prisoner's dilemma game, which indicated that spatial structure is conductive to cooperation [5]. Along with this line, several popular structures have been considered, including small-world networks [6–8], scale-free networks [9–12], interdependent networks [13,14]. Besides, to better reflect the real-life scenarios, the strategy update rule and various mechanisms have been extensively studied in evolutionary games, such as tit-for-tat [15,16], win stay and lose shift [17,18], Fermi function [19], Bush–Mosteller [20–22], Q-learning [23,24], memory [25–27], aspiration [28–30], age structure [31–34], reward and punishment [35–39], reputation [40–42], emotions [43], coevolution [44–46], asymmetry [47], to name yet a few. In particular, since uncertainties are ubiquitous in daily life, people are inevitably effected by these uncertainties. It thus becomes interesting to introduce uncertainties into games, such as periodically oscillating payoffs [48], payoff regulation [49]. For example, Gaussian noise [50] and Lévy noise [51] are considered in the payoff matrix of games, which shows that the performance of Gaussian noise seems to be much better than that of Lévy noise in promoting cooperation. In spite of some progress, these studies are often conducted under the framework of the Fermi strategy update rule. Namely, each player chooses one of its neighbors as the object to learn its strategy with equal probability or preference, which may ignore the influence of the environment. Recently, reinforcement learning (i.e., Q-learning algorithm) has been well-studied and incorporated into the evolutionary game [23] and minority game [24] to understand the emergence of cooperative behavior. Counterintuitively, reinforcement learning fails to promote cooperation in the prisoner's dilemma game. This is because these studies are mainly conducted in a well-mixed population without considering any other mechanisms.

Inspired by all these innovations, an interesting problem puts itself forward: if we combine the payoff noise and reinforcement learning simultaneously on square lattice, will the level of cooperation be promoted or not? In this work, we consider the situation where the payoff variations are controlled by Lévy noise in the prisoner's dilemma game and the strategy update rule is self-regarding Q-learning. Numerical simulation results show that noise can promote cooperation, and compared with Gaussian noise, the Lévy noise performs better in maintaining cooperation. Consequently, the studies on the effects of Lévy noise with reinforcement learning become meaningful for further comprehension of human cooperation. Thus, the main contributions of this work consist of two aspects: (1) We propose a self-regarding Q-learning framework, where agents pursue optimal strategy by only referring to their own strategy (rather than interaction environment or neighbors' actions, which is necessary in traditional reinforcement methods). (2) As noise is introduced into reinforcement learning games, we at first explore how its accumulative effect in payoff or reward influences cooperation, while such an inherent effect is nonexistent in statistical physics or mathematics methods.

The structure of this paper is presented as follows. In Sect. 2, the Lévy noise and reinforcement learning rule are introduced into prisoner's dilemma game. In Sect. 3, simulation results are presented with extensive explanations. Finally, the conclusion and discussion are provided in Sect. 4.

2 Model

Here, the prisoner's dilemma is considered to study the emergence and maintenance of cooperation, where players are located at vertices of a square lattice with periodic boundary conditions. Without loss of generality, the rescaled payoffs matrix is $R = 1$, $P = S = 0$, $T = b$, where b $(1 < b < 2)$ characterizes the temptation to defection [5]. Given that individuals are inevitably effected by the environmental factors, such as the existence of ubiquitous uncertainties in practice, we assume that players' payoffs are influenced by uncertainties presented by random payoff variations. Specially, a Lévy noise variable θ is introduced to account for the statistical description of rare events, and the value of noise also varies due to the diversity or discrepancy among individuals. Thus, the payoff matrix M_x of player x is denoted as

$$M_x = \begin{pmatrix} 1+\theta_x & 0+\theta_x \\ b+\theta_x & 0+\theta_x \end{pmatrix}. \tag{1}$$

The Lévy noise θ_x is defined by the characteristic function [52],

$$\varphi(t) = \exp\left[-\sigma^{\alpha}|t|^{\alpha}\left(1 - i\beta \mathrm{sign}(t)\tan\frac{\pi\alpha}{2}\right) + iut\right], \tag{2}$$

where $\alpha \in (0, 2]$ is referred as the stability index, representing the jump frequency and size of the noise distribution. As α increases, the jump frequency and size will decrease. When $\alpha = 2$, Lévy noise turns to a Gaussian noise. $\sigma \geq 0$, indicates the width or standard deviation of the distribution, and $\sigma = 0$ corresponds to the case without noise. The skewness of the distribution is characterized by the parameter $\beta \in [-1, 1]$, where $\beta < 0$ skews to the left and $\beta > 0$ skews to the right. u $(u \in \mathbb{R})$ is named as the location parameter representing the mean value. Like previous research [51], we fix in our work $u = 0$ and $\beta = 0$. Thus, we mainly focus on the effects of parameters α and σ, which describe the size strength and width of the distribution of payoff variations, respectively. In addition, we certify that initial payoffs minus θ [replacing plus θ like Eq. (1)] will also guarantee the same effect as the represent payoff matrix. For simplicity, we will mainly focus on the present case in this work.

During the game of each round, a player x that is randomly select from the population with the strategy of cooperation (C) or defection (D) is depicted as s_x, i.e., $s_x = C = (1, 0)^{\mathrm{T}}$ or $s_x = D = (0, 1)^{\mathrm{T}}$. Then, player x interacts with its four neighbors and obtains its payoff P_x as

$$P_x = \sum_{y \in \Omega_x} s_x^{\mathrm{T}} M_x s_y, \tag{3}$$

where Ω_x denotes the set of neighbors for player x.

As for the strategy update process, players are supposed to adopt the self-regarding Q-learning algorithm, which is different from traditional Q-learning method involving interaction with environment [53]. Under such a novel framework, player x updates the strategy with maximal Q-value based on their experience, regardless of the neighbors' strategies (which is necessary in previous Q-learning method). Q-value is defined by Q-table to record the relative utility of different actions in different states. In the following, the state set s and action set a are supposed to be the same, i.e., $\{C, D\}$. The Q-table with states (rows) and actions (columns) is provided as follows:

$$Q(t) = \begin{bmatrix} Q_{C,C}(t) & Q_{C,D}(t) \\ Q_{D,C}(t) & Q_{D,D}(t) \end{bmatrix}, \tag{4}$$

where $Q_{s,a}(t)$ represents the Q-value of the player with state s and action a at time step t. For simplicity but without loss of generality, s indicates the current state of a player, and a indicates the action that the player may take. Subsequently, when the player interacts with the neighbors, the Q-table updates according to the following equation [23,24]:

$$Q_{s,a}(t+1) = (1-\eta)Q_{s,a}(t) + \eta[P(t) + \gamma Q^{\max}_{s',a'}(t)], \tag{5}$$

where $\eta \in (0, 1]$ means the learning rate, $P(t)$ is the calculated payoff for the current action, and $\gamma \in [0, 1)$ is the discount factor representing foresight level of players (small γ means that players pay more attention to the current payoff). $Q^{\max}_{s',a'}(t)$ is the maximum value of the Q-table in the row of next state s'. Besides, in each round, the ε-greedy exploration is used during the updating. A player either acts randomly with the probability ε (a small value) or acts with the maximum value of Q-table with the probability $1-\varepsilon$. In this way, the high payoff action will be reinforced.

Totally, the evolution of games is summarized as follows: (1) Initially, players are assigned on the vertices of a $N = L \times L$ square lattice, and their initial states are randomly assigned, i.e., players choose to cooperate and defect with equal probability. (2) Since players

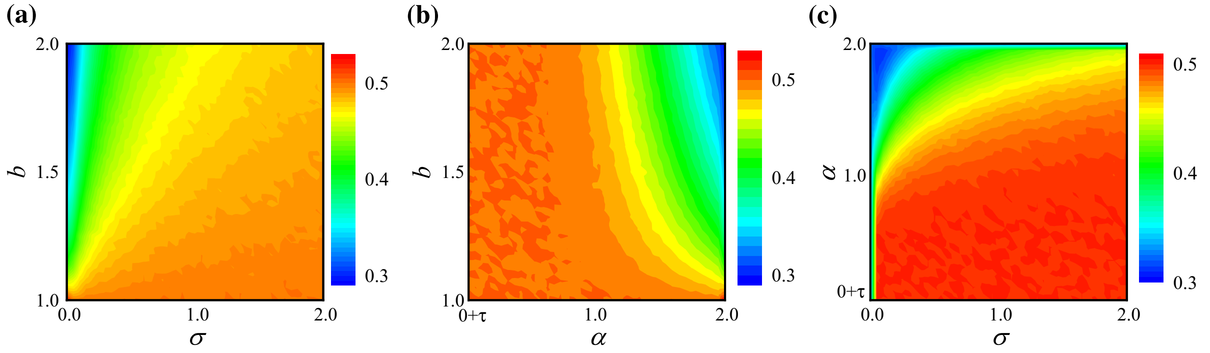

Fig. 1 Contour plots of cooperation in different panels. **a** Cooperative traits in $\sigma - b$ parameter panel with $\alpha = 1.4$, it is clear large σ (i.e., wider distribute of noise) promotes cooperation. **b** Cooperative traits in $\alpha - b$ parameter panel with $\sigma = 0.4$, it is clear that large α inhibits the level of cooperation. In addition, parameter τ is a very small value to define the limit of variable α. **c** Cooperative traits in $\sigma - \alpha$ parameter panel with $b = 1.6$, which indicates the broad diversity of noise resolving social dilemma. Other parameters are $\gamma = 0.8$, $\eta = 0.8$, $\varepsilon = 0.02$. In all cases $\tau = 0.05$

are initially unaware of the game or environment, the Q-table is initialized to zero. (3) At each round, one player x with state s and action a is randomly chosen, then interacts with its four neighbors to obtain its payoff according to Eq. (3), and updates the Q-value according to Eq. (5). (4) Next, based on the selected action a, the state of player x is updated from the current s to $s^{'}$. (5) Repeating procedures (3) and (4) for N times, one Monte Carlo step of game will be finished.

The evolutionary game is iterated forward according to the Monte Carlo simulation procedure with a 200 × 200 square lattice. The noise parameters θ_x are independently drawn for each player at each Monte Carlo step. The level of cooperation ρ_c is obtained from the average of the stationary state, namely, the last 500 steps of total 5000 steps. Meanwhile, 10 independent experiments are carried out to guarantee high accuracy.

3 Results

In this section, extensive simulations are conducted for scenarios with different parameters. Figure 1 shows how cooperation evolves under different parameter combinations. As indicated, noise plays a positive role in promoting cooperation with reinforcement learning. In Fig. 1a, when $\sigma = 0$ (corresponds to the traditional scenario of no noise), cooperation level will decrease quickly as b increases. However, when noise is considered (i.e., $\sigma > 0$), the decreasing trend of cooperation becomes much slower. Figure 1b turns to exploring the impact of parameter α on cooperation. When $\alpha \leq 1$, the level of cooperation nearly does not change, regardless of the variation of b. However, when $\alpha > 1$, increase of both parameters b and α is unbeneficial for the level of cooperation. That is, the performance of Gaussian noise ($\alpha = 2$) is inferior to that of Lévy noise. In addition, Fig. 1c depicts the contour plots of cooperation in σ–α panel. It also demonstrates the diverse effects of noise factors on the facilitation of cooperation: parameter σ promotes cooperation, yet parameter α inhibits it. Considering that cooperation level ρ_c is not significant influenced for $\alpha < 1$, we only focus on the case $\alpha \in [1, 2]$ in the following.

Then, we study the influence of Q-learning parameters on cooperative with and without noise. As depicted in Fig. 2, parameter ε plays a hybrid role in effecting the level of cooperation. In absence of noise (i.e., $\sigma = 0$), large ε promotes cooperation when b is relatively small. However, as the further increase of b, small ε performs much better in promoting cooperation, while for the case with noise, ρ_c monotonically increases with the decrease of ε. At the same time, cooperation becomes stable with the introduction of noise.

Based on the above discussions, we further investigate the impact of noise with two other parameters in Fig. 3. The top panels show the simulation results obtained for the scenario without noise. For both η and γ, ρ_c decreases with the increase of b. However, for fixed b, η and γ show different trends: cooperation level increases with the increase of η but decreases with the increase of γ. At the same time, it can be seen that

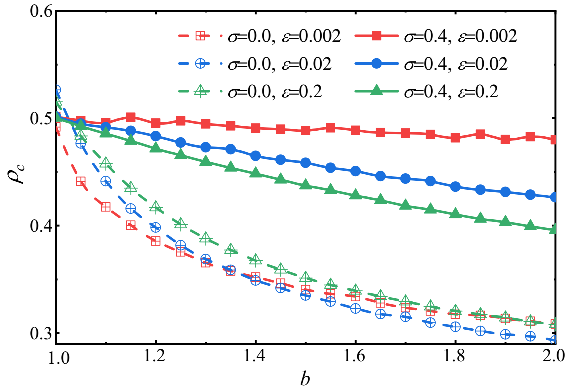

Fig. 2 Fraction of cooperation ρ_c in dependence on b for different values of ε. The results show ε plays a different role with and without noise. Other parameters are $\gamma = 0.8$, $\eta = 0.8$, $\alpha = 1.4$

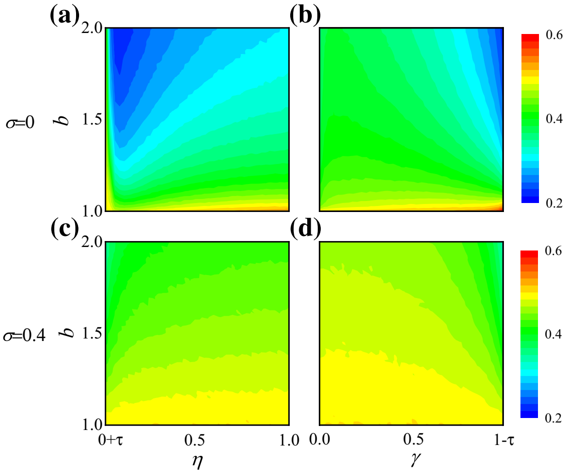

Fig. 3 Contour plots of cooperation in parameters panels with and without noise. **a** $\eta-b$ parameter panel with $\gamma = 0.8$ and **b** $\gamma-b$ parameter panel with $\eta = 0.8$ under $\sigma = 0$. The results show cooperation level increases with the increase of η but decreases with the increase of γ when there is no noise. **c** $\eta - b$ parameter panel with $\gamma = 0.8$ and **d** $\gamma - b$ parameter panel with $\eta = 0.8$ under $\sigma = 0.4$. The results show the same trends as **a** and **b**, but cooperation level is enhanced when noise is introduced. Other parameters are $\alpha = 1.4$, $\varepsilon = 0.02$. In all cases $\tau = 0.025$

there is no dramatic change in the level of cooperation when $\gamma \le 0.75$. Moreover, with a larger value of $\sigma = 0.4$ (bottom panels), where the noise is introduced, the results show the same trends as Fig. 3a, b, but the total level of cooperation is enhanced. It is worthy mentioning that the above observation is also suitable for larger σ values.

Next, in order to clearly understand the impact of noise and self-regarding Q-learning, contour plots of cooperation in $\eta - \gamma$ parameter panels are shown in Fig. 4 for different α. It is clear when noise is introduced, the level of cooperation is enhanced with the decrease of α (from Fig. 4b–d). As α is large ($\alpha = 2$, the heterogeneity of noise is small), the level of cooperation is almost the same as the case without noise (Fig. 4a). Thus, the impact of Gaussian noise on cooperation is negligible. With the increase of noise, Fig. 4c, d shows that small values of α promote cooperation reaching a higher level. Furthermore, as presented in Fig. 4d, the level of cooperation is not influenced by varying self-regarding Q-learning parameters, which indicates that Lévy noise performs better in maintaining cooperation with reinforcement learning, compared with Gaussian noise (Fig. 4b).

Considering the two noise parameters, the time evolution of ρ_c is reported in Fig. 5 for several typical values of σ and α. It can be observed that almost all evolutionary processes will arrive at the stationary state after about 1000 steps, and the results indicate that with the increase of σ and decrease of α, the cooperators' ability to resist the defectors becomes stronger. In particular, when there is no noise ($\sigma = 0$, in Fig. 5a), the level of cooperation finally stabilizes at around 32%. However, when noise is introduced, ρ_c can be significantly improved with the increase of σ (especially for $\sigma < 1$). It is also worth noting that increase of σ will change the direct decline of cooperation to a weak negative feedback effect (i.e., a normal enduring + expanding process [54]). However, when σ is sufficient large (i.e., $\sigma = 10$), there is no influence on initial cooperation level. That's why we limit σ to [0, 2] in Fig. 1. Further, Fig. 5b shows that ρ_c decreases with the increase of α. At the same time, the enduring + expanding process turns to monotonic decreasing.

The remaining problem is how to explain the above phenomenon. As described before, the adopted strategy update rule is self-regarding Q-learning, and the action that players will choose at time t relies on their current Q-table. Thus, the evolution of Q-table is a key to understand the above phenomena. Here, the evolution of average Q-table and noise from the Monte Carlo simulation of all players is recorded in Fig. 6. During simulation, the average Q-table is defined as follows:

$$\bar{Q}(t) = \begin{bmatrix} \bar{Q}_{C,C}(t) & \bar{Q}_{C,D}(t) \\ \bar{Q}_{D,C}(t) & \bar{Q}_{D,D}(t) \end{bmatrix}, \tag{6}$$

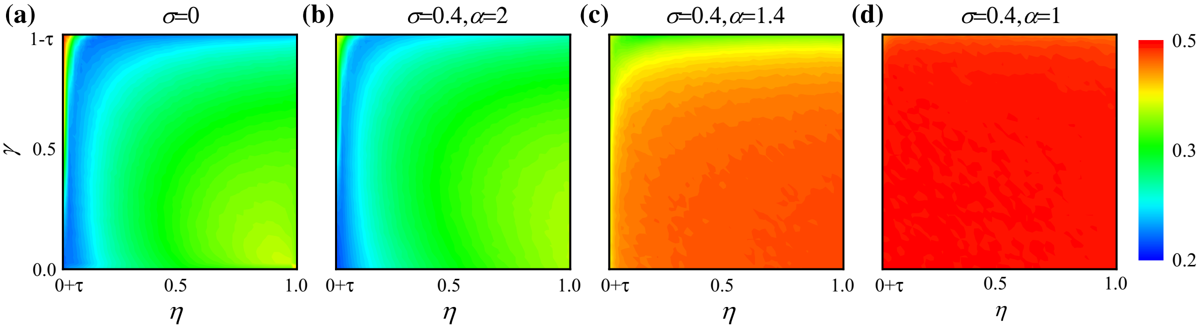

Fig. 4 Contour plots of cooperation in $\eta - \gamma$ parameter panels. From **b** to **d**, the noise parameter α is set to be 2, 1.4, and 1, respectively. Meanwhile, for comparison, panel **a** shows the cooperation level when there is no noise. The results show that compared with Gaussian noise ($\alpha = 2$), Lévy noise performs better in maintaining cooperation with reinforcement learning. Other parameters are $b = 1.6$, $\varepsilon = 0.02$. In all cases $\tau = 0.025$

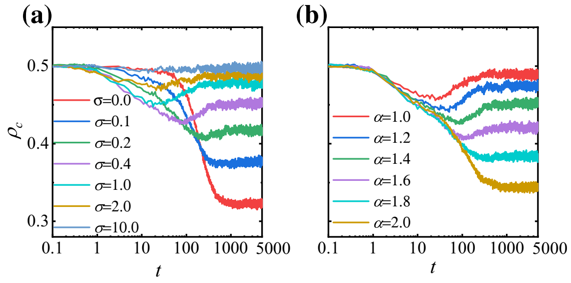

Fig. 5 The time evolution of cooperation with the noise parameters. **a** $\alpha = 1.4$ **b** $\sigma = 0.4$. The results indicate that with the increase of σ and decrease of α, the cooperators' ability to resist the defectors becomes stronger. Other parameters are $b = 1.6$, $\gamma = 0.8$, $\eta = 0.8$, $\varepsilon = 0.02$

where $\bar{Q}_{s,a}(t) = \sum_i^N Q_{s,a}(t)/N$. Figure 6 shows the evolution of Q-value is effected by the fluctuation of noise. When there is no noise (Fig. 6a) or small noise (Fig. 6b), $\bar{Q}_{s,a}$ reaches a positive stable value quickly. However, with the increase of noise, the fluctuation of $\bar{Q}_{s,a}$ will become more and more frequent, as shown in Fig. 6c, d. This is because the effect of noise is accumulated in the Q-value update formula.

We further explore the dynamic evolution processes of different $\bar{Q}_{s,a}$ in Fig. 7. As previously, absence of noise (Fig. 7a) or low noise (Fig. 7b) enables $\bar{Q}_{s,a}$ to reach a stable state but high noise (Fig. 7c, d) makes them fluctuate. It is obvious that irrespective of any case, $\bar{Q}_{C,D}$ is always large than $\bar{Q}_{C,C}$. However, as noise increases, $\bar{Q}_{D,C}$ will become greater than $\bar{Q}_{D,D}$. This interesting reversal means when the current state of a player is defection under self-regarding Q-learning rule, cooperative behavior will have an advantage over defective behavior subsequently. Thus, more defectors will turn to cooperation in next round. This can explain why Lévy noise performs better than Gaussian noise.

Finally, it is still interesting to study whether the initial setting affects the stability of cooperation (i.e., the robustness of our method). To this aim, we randomly select 10%, 30%, 50%, 70%, and 90% of overall population as cooperation in the early stage. As shown in Fig. 8, since players strongly prefer to learn the maximal-payoff strategy (avoid being exploited by others), the level of cooperation is almost unconstrained to the initial distribution, which is robust to small and larger noise values. From this perspective, cooperation can reach a steady state with the same cooperation level if players only learn the maximal-payoff action. Moreover, the case with noise (solid lines) converges to a steady state faster than the case without noise (dash lines).

4 Conclusion and discussion

To conclude, we have studied the impact of noise with reinforcement learning on the level of cooperation. Numerical simulations show that noise can enhance the survival of cooperation even when the temptation is high. In detail, the decrease of noise parameter α could support a more stable evolution of cooperation, which means that Lévy noise performs better in maintaining cooperation than Gaussian noise. Then, in order to explain the above phenomena, we further explore the time evolution of noise parameters, which indicates

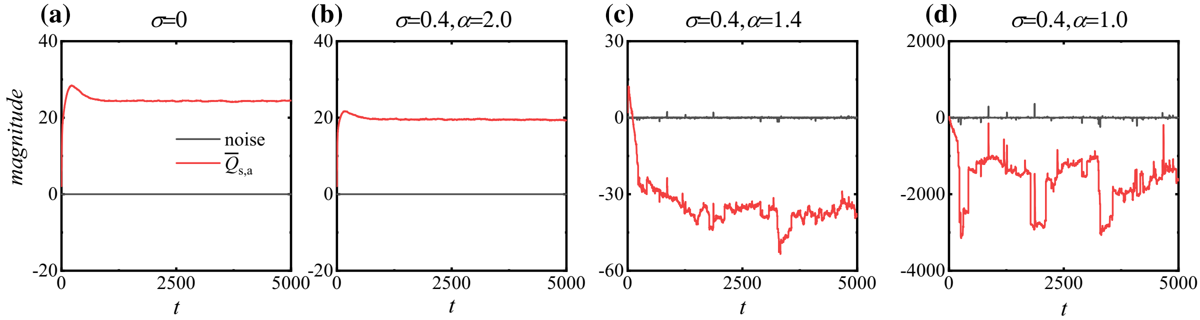

Fig. 6 Time evolution of average noise and $\bar{Q}_{s,a}$. **a** Without noise, **b**–**d**, the noise parameter α is set to be 2, 1.4, and 1, respectively. The results show that with the decrease of α, the magnitude of $\bar{Q}_{s,a}$ would change with noise fluctuation, and the effect of noise is accumulated in the Q-value update formula. Other parameters are $b = 1.6$, $\gamma = 0.8$, $\eta = 0.8$, $\varepsilon = 0.02$

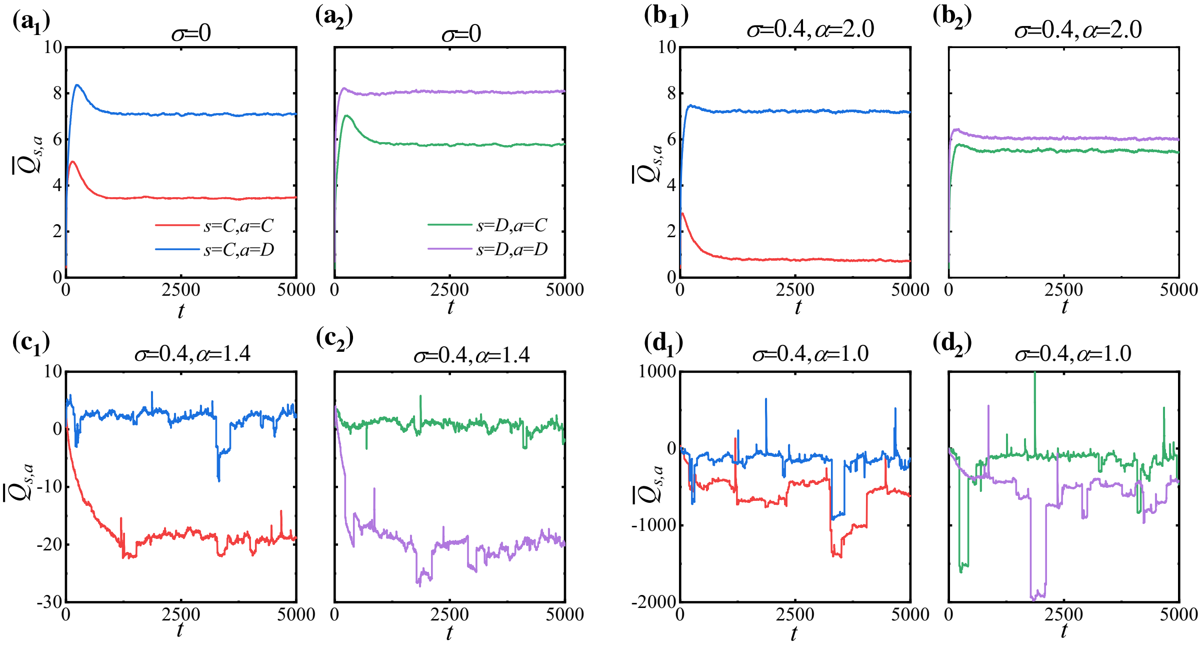

Fig. 7 Time evolution of different $\bar{Q}_{s,a}$. **a** Without noise, **b**–**d** the noise parameter α is set to be 2, 1.4, and 1, respectively. With the decrease of α, though $\bar{Q}_{C,D}$ (blue line) is always larger than $\bar{Q}_{C,C}$ (red line), $\bar{Q}_{D,C}$ (green line) and $\bar{Q}_{D,D}$ (purple line) will reverse the order, which means players will be more inclined to cooperate when the current state of a player is defection. Other parameters are $b = 1.6$, $\gamma = 0.8$, $\eta = 0.8$, $\varepsilon = 0.02$. (Color figure online)

that players can quickly converge to a stable state and the cooperators' ability to resist the defectors becomes stronger with the increase of noise parameter σ and the decrease of α. Then the time evolution of average noise and Q-value shows that the Q-value fluctuates with the increase of noise, and the effect is accumulated in the Q-value update formula. In particular, with the increase of noise, the Q-value of cooperative behavior will be larger than that of defective behavior. Thus, it leads to the improvement of cooperation. Finally, we also verify that cooperation will always converge to the same level regardless of the initial state.

In contrast to Ref. [51], where the Fermi function is utilized as the update rule, the self-regarding Q-learning algorithm unveils a particular positive role of Lévy noise in the evolutionary dynamics of social

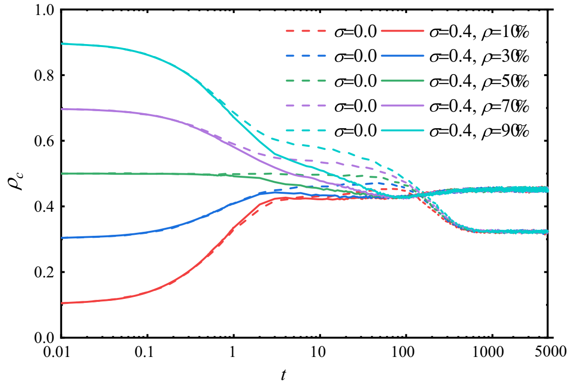

Fig. 8 Evolution of cooperation with different initial cooperation levels. The results show cooperation level will always converge to the same value regardless of the initial state. Other parameters are $b = 1.6$, $\alpha = 1.4$, $\gamma = 0.8$, $\eta = 0.8$, $\varepsilon = 0.02$

dilemmas. Furthermore, our work is also different from Refs. [23,24]. They considered Q-learning algorithm with the evolutionary game and minority game in a well-mixed population. Here, we focus on structured population with noise, which is more close to empirical observations. Thus, we believe that our study will be useful in addressing more social dilemmas that arises in real-life situations.

Acknowledgements This research was supported by National Key R&D Program of China(Grant no. 2018AAA0100905), the National Science Fund for Distinguished Young Scholars (Grants No. 62025602), the National Natural Science Foundation of China (Grant Nos. 11931015, U1803263,81961138010 and 62073263), Fok Ying-Tong Education Foundation, China (Grant No. 171105), Key Technology Research and Development Program of Science and Technology-Scientific and Technological Innovation Team of Shaanxi Province (Grant No. 2020TD-013), the Tencent Foundation and XPLORER PRIZE, the Slovenian Research Agency (Grant Nos. P1-0403 and J1-2457), and Key Research and Development Program of Shaanxi Province (Grant No 2022KW-26). Discussions with Hao Guo and Chen Chu are gratefully acknowledged.

Data availability The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

References

1. Darwin C.: The Origin of Species. Harward Univ. Press, Cambridge (1859) (Reprinted, 1964)
2. Perc, M., Marhl, M.: Evolutionary and dynamical coherence resonances in the pair approximated prisoner's dilemma game. New J. Phys. **8**(8), 142 (2006)
3. Zhang, J., Zhang, C., Chu, T., Perc, M.: Resolution of the stochastic strategy spatial prisoner's dilemma by means of particle swarm optimization. PLoS ONE **6**(7), e21787 (2011)
4. Wu, Z.X., Xu, X.J., Huang, Z.G., Wang, S.J., Wang, Y.H.: Evolutionary prisoner's dilemma game with dynamic preferential selection. Phys. Rev. E **74**, 21107 (2006)
5. Nowak, M.A., May, R.M.: Evolutionary games and spatial chaos. Nature **359**(6398), 826–829 (1992)
6. Tomassini, M., Luthi, L., Giacobini, M.: Hawks and doves on small-world networks. Phys. Rev. E **73**(1), 16132 (2006)
7. Fu, F., Liu, L.H., Wang, L.: Evolutionary prisoner's dilemma on heterogeneous Newman-Watts small-world network. Eur. Phys. J. B **56**(4), 367–372 (2007)
8. Chen, X., Wang, L.: Promotion of cooperation induced by appropriate payoff aspirations in a small-world networked game. Phys. Rev. E **77**(1), 17103 (2008)
9. Santos, F.C., Pacheco, J.M.: Scale-free networks provide a unifying framework for the emergence of cooperation. Phys. Rev. Lett. **95**(9), 98104 (2005)
10. Rong, Z., Li, X., Wang, X.: Roles of mixing patterns in cooperation on a scale-free networked game. Phys. Rev. E **76**(2), 27101 (2007)
11. Assenza, S., Gómez-Gardeñes, J., Latora, V.: Enhancement of cooperation in highly clustered scale-free networks. Phys. Rev. E **78**(1), 17101 (2008)
12. Poncela, J., Gómez-Gardenes, J., Moreno, Y.: Cooperation in scale-free networks with limited associative capacities. Phys. Rev. E **83**(5), 57101 (2011)
13. Xia, C., Li, X., Wang, Z., Perc, M.: Doubly effects of information sharing on interdependent network reciprocity. New J. Phys. **20**(7), 75005 (2018)
14. Shi, L., Shen, C., Geng, Y., Chu, C., Meng, H., Perc, M., Boccaletti, S., Wang, Z.: Winner-weaken-loser-strengthen rule leads to optimally cooperative interdependent networks. Nonlinear Dyn. **96**(1), 49–56 (2019)
15. Nowak, M.A., Sigmund, K.: Tit for tat in heterogeneous populations. Nature **355**(6357), 250–253 (1992)
16. Baek, S.K., Kim, B.J.: Intelligent tit-for-tat in the iterated prisoner's dilemma game. Phys. Rev. E **78**(1), 11125 (2008)
17. Nowak, M.A., Sigmund, K.: A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game. Nature **364**(6432), 56–58 (1993)
18. Amaral, M.A., Wardil, L., Perc, M., da Silva, J.K.L.: Stochastic win-stay-lose-shift strategy with dynamic aspirations in evolutionary social dilemmas. Phys. Rev. E **94**(3), 32317 (2016)
19. Szabó, G., Tőke, C.: Evolutionary prisoner's dilemma game on a square lattice. Phys. Rev. E **58**(1), 69–73 (1998)
20. Ezaki, T., Horita, Y., Takezawa, M., Masuda, N.: Reinforcement learning explains conditional cooperation and its moody cousin. PLoS Comput. Biol. **12**(7), e1005034 (2016)

21. Jia, D., Guo, H., Song, Z., Shi, L., Deng, X., Perc, M., Wang, Z.: Local and global stimuli in reinforcement learning. New J. Phys. **23**(8), 83020 (2021)
22. Jia, D., Li, T., Zhao, Y., Zhang, X., Wang, Z.: Empty nodes affect conditional cooperation under reinforcement learning. Appl. Math. Comput. **413**(6398), 126658 (2022)
23. Zhang, S.P., Zhang, J.Q., Chen, L., Liu, X.D.: Oscillatory evolution of collective behavior in evolutionary games played with reinforcement learning. Nonlinear Dyn. **99**, 3301–3312 (2020)
24. Zhang, S.P., Zhang, J.Q., Huang, Z.G., Guo, B.H., Wu, Z.X., Wang, J.: Collective behavior of artificial intelligence population: transition from optimization to game. Nonlinear Dyn. **95**(2), 1627–1637 (2019)
25. Wang, W.X., Ren, J., Chen, G., Wang, B.H.: Memory-based snowdrift game on networks. Phys. Rev. E **74**(5), 56113 (2006)
26. Hilbe, C., Martinez-Vaquero, L.A., Chatterjee, K., Nowak, M.A.: Memory-n strategies of direct reciprocity. Proc. Natl. Acad. Sci. USA **114**(8), 4715–4720 (2017)
27. Dong, Y., Xu, H., Fan, S.: Memory-based stag hunt game on regular lattices. Physica A **519**, 247–255 (2019)
28. Platkowski, T.: Enhanced cooperation in prisoner's dilemma with aspiration. Appl. Math. Lett. **22**(8), 1161–1165 (2009)
29. Yang, H.X., Wu, Z.X., Wang, B.H.: Role of aspiration-induced migration in cooperation. Phys. Rev. E **81**, 65101–65104 (2010)
30. Rong, Z.H., Zhao, Q., Wu, Z.X., Zhou, T., Tse, C.K.: Proper aspiration level promotes generous behavior in the spatial prisoner's dilemma game. Eur. Phys. J. B **89**(7), 1–7 (2016)
31. Szolnoki, A., Perc, M., Szabó, G., Stark, H.U.: Impact of aging on the evolution of cooperation in the spatial prisoner's dilemma game. Phys. Rev. E **80**, 21901 (2009)
32. Wang, Z., Zhu, X., Arenzon, J.J.: Cooperation and age structure in spatial games. Phys. Rev. E **85**(1), 011149 (2012)
33. Wang, Z., Wang, Z., Yang, Y.H., Yu, M.X., Liao, L.: Age-related preferential selection can promote cooperation in the prisoner's dilemma game. Int. J. Mod. Phys. C **23**(2), 1250013 (2012)
34. Han, Y., Song, Z., Sun, J., Ma, J., Guo, Y., Zhu, P.: Investing the effect of age and cooperation in spatial multigame. Physica A **541**, 123269 (2020)
35. Fowler, J.H.: Altruistic punishment and the origin of cooperation. Proc. Natl. Acad. Sci. USA **102**(19), 7047–7049 (2005)
36. Balliet, D., Mulder, L.B., Van Lange, P.A.M.: Reward, punishment, and cooperation: a meta-analysis. Psychol. Bull. **137**(4), 594–615 (2011)
37. Wu, Y., Chang, S., Zhang, Z., Deng, Z.: Impact of social reward on the evolution of the cooperation behavior in complex networks. Sci. Rep. **7**(1), 1–9 (2017)
38. Zhu, P., Guo, H., Zhang, H., Han, Y., Wang, Z., Chu, C.: The role of punishment in the spatial public goods game. Nonlinear Dyn. **102**(4), 2959–2968 (2020)
39. Song, Q., Cao, Z., Tao, R., Jiang, W., Liu, C., Liu, J.: Conditional Neutral Punishment Promotes Cooperation in the Spatial Prisoner's Dilemma Game. Appl. Math. Comput. **368**, 124798 (2020)
40. Fu, F., Hauert, C., Nowak, M.A., Wang, L.: Reputation-based partner choice promotes cooperation in social networks. Phys. Rev. E **78**(2), 26117 (2008)
41. Gallo, E., Yan, C.: The effects of reputational and social knowledge on cooperation. Proc. Natl. Acad. Sci. USA **112**(12), 3647–3652 (2015)
42. Gross, J., De Dreu, C.: The rise and fall of cooperation through reputation and group polarization. Nat. Commun. **10**(1), 1–10 (2019)
43. Wang, L., Ye, S.Q., Cheong, K.H., Bao, W., Xie, N.: The role of emotions in spatial prisoner's dilemma game with voluntary participation. Physica A **490**, 1396–1407 (2018)
44. Wang, Z., Szolnoki, A., Perc, M.: Self-organization towards optimally interdependent networks by means of coevolution. New J. Phys. **16**(3), 33041 (2014)
45. Liu, C., Guo, H., Li, Z., Gao, X., Li, S.: Coevolution of multi-game resolves social dilemma in network population. Appl. Math. Comput. **341**, 402–407 (2019)
46. Chu, C., Mu, C., Liu, J., Liu, C., Boccaletti, S., Shi, L., Wang, Z.: Aspiration-based coevolution of node weights promotes cooperation in the spatial prisoner's dilemma game. New J. Phys. **21**(6), 63024 (2019)
47. Guo, H., Li, X., Hu, K., Dai, X., Jia, D., Boccaletti, S., Perc, M., Wang, Z.: The dynamics of cooperation in asymmetric sub-populations. New J. Phys. **22**(8), 83015 (2020)
48. Babajanyan, S.G., Lin, W., Cheong, K.H.: Cooperate or not cooperate in predictable but periodically varying situations? Cooperation in fast oscillating environment. Adv. Sci. **7**(21), 2001995 (2020)
49. Jiang, L.L., Zhao, M., Yang, H.X., Wakeling, J., Wang, B.H., Zhou, T.: Reducing the heterogeneity of payoffs: an effective way to promote cooperation in the prisoner's dilemma game. Phys. Rev. E **80**(3), 031144 (2009)
50. Perc, M.: Coherence resonance in a spatial prisoner's dilemma game. New J. Phys. **8**(2), 22 (2006)
51. Perc, M.: Transition from Gaussian to Levy distributions of stochastic payoff variations in the spatial prisoner's dilemma game. Phys. Rev. E **75**(2), 22101 (2007)
52. Xu, W., Hao, M., Gu, X., Yang, G.: Stochastic resonance induced by Lévy noise in a tumor growth model with periodic treatment. Mod. Phys. Lett. B. **28**, 1450085 (2014)
53. Watkins, C.J., Dayan, P.: Q-learning. Mach. Learn. **8**(3–4), 279–292 (1992)
54. Shigaki, K., Wang, Z., Tanimoto, J., Fukuda, E.: Effect of initial fraction of cooperators on cooperative behavior in evolutionary prisoner's dilemma game. PLoS ONE **8**(11), e76942 (2013)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.