

RESEARCH ARTICLE

Strategies for updating rules driven by reinforcement learning to solve social dilemmas

Yang Wang^{1,2*}, Xingchen Yu¹, Shounan Lu³

1 School of Information and Control Engineering, North China Institute of Science and Technology, Langfang, Hebei, China, **2** Key Laboratory of Brain-Computer Interface Technology Application of the Ministry of Emergency Management, Langfang, Hebei, China, **3** School of Mathematical Sciences, Laboratory of Mathematics and Complex Systems, MOE, Beijing Normal University, Beijing, China

* wy0926go@163.com



Abstract

This study incorporates historical performance into traditional imitation rules and proposes a moderated strategy update rule. In this framework, an individual's temporal historical performance is calculated using the BM model. By adjusting the parameter δ , the influence of historical performance on strategy learning is determined, and the evolution of cooperation is subsequently observed. Results show that the proposed strategy update rule promotes cooperation more effectively than the traditional version, and systemic cooperation is further enhanced as δ increases. The reason why the proposed rule enhances cooperation is that it amplifies the evaluation of cooperative behavior while compressing the evaluation of defective behavior. Although establishing system objectives may hinder the diffusion of cooperative behavior, appropriate performance evaluation mechanisms can mitigate this adverse effect. Our results indicate that multidimensional evaluation can provide a theoretical basis for explaining cooperative behavior in complex environments.

OPEN ACCESS

Citation: Wang Y, Yu X, Lu S (2026) Strategies for updating rules driven by reinforcement learning to solve social dilemmas. PLoS One 21(3): e0341925. <https://doi.org/10.1371/journal.pone.0341925>

Editor: Xiaojie Chen, University of Electronic Science and Technology of China, CHINA

Received: June 28, 2025

Accepted: January 14, 2026

Published: March 10, 2026

Copyright: © 2026 Wang et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data availability statement: All relevant data are within the paper and its [Supporting information](#) files.

Funding: 1. Fundamental Research Funds for the Central Universities (Grant No 3142024036) 2. Self-funded Project of Langfang Science and Technology Research and Development Program (Grant No 2024011066).

1. Introduction

Cooperation is fundamental to the stability of biological ecosystems and human social systems. Yet, reconciling cooperative behaviors with Darwinian principles of natural selection [1] presents a persistent theoretical challenge. This has prompted most researchers to explore the principles underlying cooperation. Nowak's seminal work summarizes five key mechanisms facilitating cooperation: kin selection, direct reciprocity, indirect reciprocity, network reciprocity, and group selection [2]. Crucially, network reciprocity exploits underlying population structure, enabling cooperators to cluster and thereby sustain cooperation [3]. Subsequent research has extended this foundational framework, incorporating diverse mechanisms such as punishment and reward [4–7], environmental feedback [8,9], social diversity [10], teaching activities [11,12], reputation [13], among others [14].

Competing interests: The authors have declared that no competing interests exist.

While numerous studies posit strategy learning in cooperative evolution primarily through imitation rules [3]—where benefits serve as the key driver—the reality involves complex, multidimensional influences. Strategy learning behavior results from the interplay of factors like self-learning and social learning [15]. Scholars have explored the impact of diverse strategy updating rules on the emergence and maintenance of cooperation. For instance, Yan and Hui demonstrated that integrating reputation mechanisms significantly enhances cooperation [16]. Similar conclusions were reached by Zhang et al. and He et al. [17,18]. Lu and Wang, incorporating past-performance into learning rules, found that increasing the weight of historical outcomes progressively strengthens system-wide cooperation [19]. Other investigated rules include popularity-driven [20] and experience-driven updates [21]. Collectively, these findings indicate that multi-factor dependent learning rules generally foster cooperative evolution.

Despite this extensive exploration, the predominant focus remains on extrinsic social attributes, such as reputation. Consequently, a critical gap persists: the evolutionary patterns and outcomes of system cooperation under composite strategy learning rules driven primarily by intrinsic individual attributes remain unexplored.

In addition, most existing studies focus on the immediate benefits of individual behavior within a single interaction round, overlooking the accumulated experience from prior games. This approach fails to capture the natural phenomenon in which organisms adapt their social strategies based on environmental cues, including feedback from past experiences. Reinforcement learning (RL) rules, however, effectively incorporate the cumulative influence of such memory effects [22–24]. Consequently, researchers have increasingly explored RL in evolutionary cooperation studies. For instance, Jia et al. demonstrated that incorporating RL enhances system-wide cooperation [25]. However, the research conducted by Lu et al [19], focused on the impact of reinforcement learning based relationship strength adjustment on cooperation, neglecting the dual effects of internal and external factors in strategy learning. The studies of Jia et al. [25] and Geng et al. [26] also overlooked the role of individual intrinsic factors in strategy learning. In addition, although Zhang et al. combined reinforcement learning with consensus learning rules to study how different policy update mechanisms affect cooperative evolution, the model assumes that individuals are rational [27,28], which does not reflect the actual cooperative evolution situation. Notably, recent findings suggest RL not only accounts for conditional cooperation but also explains patterns of emotional reciprocity [29,30].

Accordingly, we conceptualize the system's consistency goal [31–35] as an intrinsic driver of individual behavior. Achieving this goal serves as one criterion for evaluating behavioral performance: success in the preceding round raises the current performance score, while failure lowers it. Drawing on reinforcement learning principles, we accumulate behavioral information across successive rounds to assess historical performance. From a global perspective, we evaluate individual performance and use the resulting assessment as a measure of social evaluation. By taking the interactive payoffs among individuals as the basis for mutual assessment, the strategy learning process is systematically guided through the integration of both

social and individual evaluations. On this basis, we examine the evolution of cooperation within the system. This update rule incorporates both real-time game payoffs and historical behavior to govern strategy revisions. Our results show that this modified update mechanism significantly promotes the emergence of prosocial behaviors in the system.

2. Model

In this work, the weak Prisoner's Dilemma is used [3], and without loss of generality, set the game payoff $T=b$ ($b > 1$), $R=1$, and $P=S=0$, and follows: $T>R>P>S$ and $2R>T+S$. The corresponding payoff matrix M in Eq. (1).

$$M = \begin{pmatrix} R = 1 & S = 0 \\ T = b & P = 0 \end{pmatrix} \tag{1}$$

Then, we construct a two-dimensional spatial network with periodic boundary to depict the relationships between individuals in the system. Initially, each individual randomly adopts either cooperation ($S_i=C$) or defection ($S_i=D$) with equal probability p , as specified in Eq. (2), interacting with its four nearest neighbors, accumulating income P_i based on Eq. (3), where Ω_i is the set of individual i 's neighbors.

$$S_i = C = (1, 0)^T, S_j = D = (0, 1)^T \tag{2}$$

$$P_i = \sum_{j \in \Omega_i} S_i^T M S_j \tag{3}$$

Subsequently, the BM model [22–24] is employed within a reinforcement learning framework to calculate and assess an individual's historical performance. In performance evaluation, the system's aspiration level for consistency serves as the evaluation benchmark. If an individual's cumulative performance during the evaluation period reaches or exceeds this benchmark, their score increases; otherwise, it decreases. And this adjustment mechanism operates persistently. Within BM reinforcement learning, the performance evaluation process occurs in two distinct steps. First, performance is scored according to the degree of deviation between the cumulative revenue and the expected system consistency target, in Eq. (4), where β ($\beta \geq 0$) is the stimulus sensitivity to the reinforcement signal of (P_i-A). Subsequently, evaluating based on individual strategies, in Eq. (5), where g_i represents the satisfaction of players with the difference between P_i and A , then, the global evaluation results E_i of individual historical behavior performance can be quantified, with the calculation formula specified as follows.

$$g_i = \tanh[\beta(P_i - A)] \tag{4}$$

$$E_i(t+1) = \begin{cases} E_i(t) + [1 - E_i(t)] \times g_i, & \text{if } S_i(t) = C \text{ and } g_i \geq 0 \\ E_i(t) + E_i(t) \times g_i, & \text{if } S_i(t) = C \text{ and } g_i < 0 \\ E_i(t) - E_i(t) \times g_i, & \text{if } S_i(t) = D \text{ and } g_i \geq 0 \\ E_i(t) - [1 - E_i(t)] \times g_i, & \text{if } S_i(t) = D \text{ and } g_i < 0 \end{cases} \tag{5}$$

Where, parameter A represents the consistency goal or expected level of the system, and define $A = k_i \alpha$, where $k_i = 4$ denotes player i 's degree, representing the number of their four nearest neighbors [36,37], and α signifies the system's consistency aspiration or goal level.

Finally, to refine learning strategies, this study proposes a moderated update rule. This rule integrates individual historical performance assessment with imitation dynamics, combining game payoff and historical performance to guide strategy

updates, refers to the previous strategy learning rule setting and linearly adds them together [38,39]. The parameter δ ($\delta \in [0, 1]$) modulates the weighting of historical performance in learning. When $\delta=0$, the strategy learn rule returns to its traditional version [3]. When $\delta>0$, both game payoffs and historical performance jointly shape strategy adaptation dynamics. Specifically, during the strategy update or learning process, the focal individual i will randomly select a nearest neighbor j and decide whether to learn the strategy of neighbor j based on probability W that based on Eq. (6). Here, parameter K quantifies the stochastic noise level that enables irrational decisions. As $K \rightarrow 0$, agent i deterministically adopts the strategy of adjacent agent j , whereas when $K \rightarrow \infty$, the strategy imitation occurs randomly. Following Ref. [40], we set $K=0.5$.

$$W_{(S_i(t) \leftarrow S_j(t))} = \frac{1}{1 + \exp\left[-\frac{(P_i - P_j) + \delta * (E_i(t) - E_j(t))}{K}\right]} \tag{6}$$

To evaluate the effectiveness of the proposed cooperation-enhancing mechanism, Monte Carlo simulations (MCS, which stands for Monte Carlo step) were performed on a 200×200 lattice network. Initially setting $p=0.5$ and $E_i(0) = 0.5$, each individual updated their strategy once per full MCS cycle on average. The equilibrium cooperation frequency f_c was measured at 1×10^4 MCS, with data averaged over 3×10^3 MCS to minimize fluctuations. Results reflect 20 independent trials.

3. Result

Fig 1 illustrates the evolution of cooperation level f_c across defect temptation values b for varying δ . When $\delta=0$, strategy updates revert to conventional imitation rules, causing cooperators to rapidly disappear even at low b . For $\delta>0$, however, agents incorporate historical performance into strategy evaluation. This modification substantially elevates cooperation levels, with higher δ values further amplifying cooperative behavior. In particular, when δ is set to 4, the payoff dimensions become consistent with historical performance, promoting a greater degree of system cooperation. Consequently, the moderated update rule extends the critical threshold b for cooperation extinction beyond conventional imitation, thereby promoting both the emergence and sustainability of cooperation.

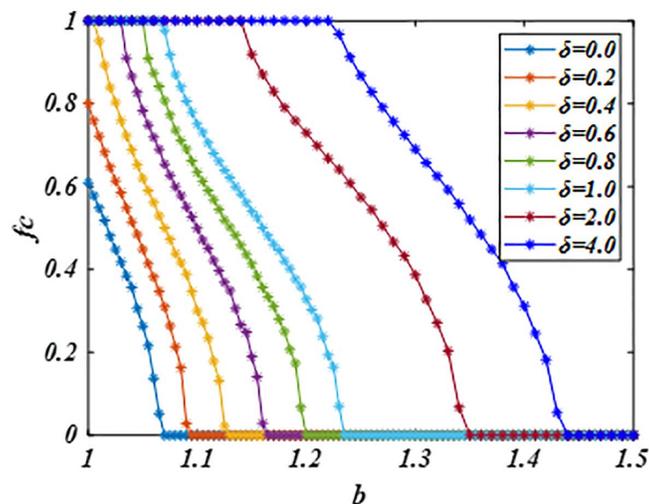


Fig 1. Illustrates f_c versus b under different δ values. Vertically, with increasing δ , the system evolves to a stable state with a greater degree of cooperation. Showing that the proposed mechanism promotes cooperation and broadens its disappearance threshold b . Parameters are set to $\beta=2$, $\alpha=0.5$ based on Ref. [24] and $L=200$.

<https://doi.org/10.1371/journal.pone.0341925.g001>

We further investigated the impact of the parameter δ on strategic evolution from the perspective of population dynamics. Fig 2 shows the temporal evolution of cooperation for $b = 1.1$ under different δ values. The process exhibits two distinct stages: an initial decline followed by a rise. After reaching a minimum, the cooperation level increases steadily until it stabilizes at an evolutionary equilibrium. This characteristic dip-and-rise pattern reflects the intense competition between cooperators and defectors, which is typical of network reciprocity [41–44]. The proposed moderated strategy-update rule significantly enhances cooperation. Notably, higher δ values drive the system toward a more cooperative equilibrium, accelerate the evolution of cooperation, and shorten the duration of the second phase. These results confirm that the modified rule effectively promotes cooperation.

Moving forward, we analyze the evolutionary dynamics by examining the spatial distribution of individual strategies to observe the competition between cooperative and competitive behaviors. As shown in Fig 3, when cooperators are initially placed in distinct positions, they quickly disappear under traditional strategy update rules. However, when moderated strategy update rules are applied, cooperators gradually spread within clusters of defectors. They gain a competitive advantage over defectors, which strengthens as parameter δ increases. Ultimately, this advantage allows cooperators to dominate, leading to a higher level of system cooperation at equilibrium. Notably, in the stable state, cooperators survive within numerous small, compact clusters.

Then, we further analyze the impact of sensitivity parameter β on cooperative evolution given parameter $\delta = 1.0$ and $\alpha = 0.5$. And the impact of the system consistency aspiration or goal value α on cooperative evolution under the given parameter $\beta = 2$ and $\delta = 1.0$.

In Fig 4(a), at $\delta = 1.0$ and $\alpha = 0.5$, the system cooperation level increases with the sensitivity parameter β , demonstrating its impact on cooperative evolution. This may be because the increase in sensitivity parameters amplifies the gap between game benefits and consistency aspiration or goals. Under this evaluation criterion, the results of cooperative payoff evaluation are amplified while the results of defective payoff evaluation are compressed, which accelerates the propagation and diffusion of cooperative behavior in the system, which also explains the strengthening effect of the proposed mechanism on cooperative behavior. However, as the sensitivity parameter β increases further, it suppresses cooperation in systems characterized by strong temptation. Under high-intensity social dilemma (b is large), defection obtain greater payoff that

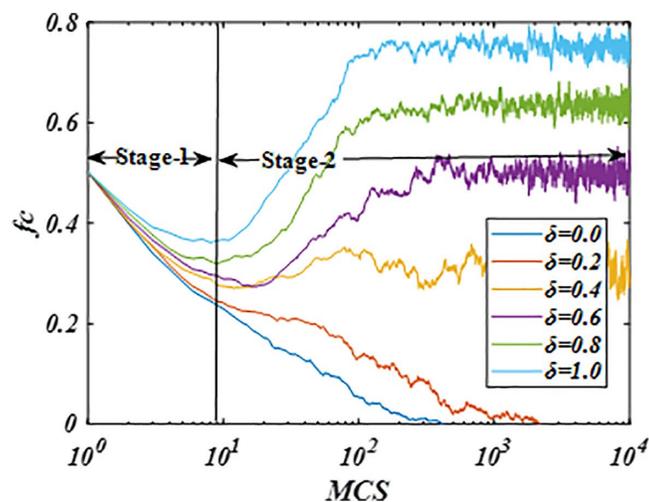


Fig 2. Illustrates temporal evolution of cooperative for some δ values, showing that the proposed mechanism can promote cooperation. With increasing δ , the system evolves to a stable state with a greater degree of cooperation. The results indicate that higher parameter δ can promote cooperation. Parameters are set to $\beta = 2$, $\alpha = 0.5$ and $L = 200$.

<https://doi.org/10.1371/journal.pone.0341925.g002>

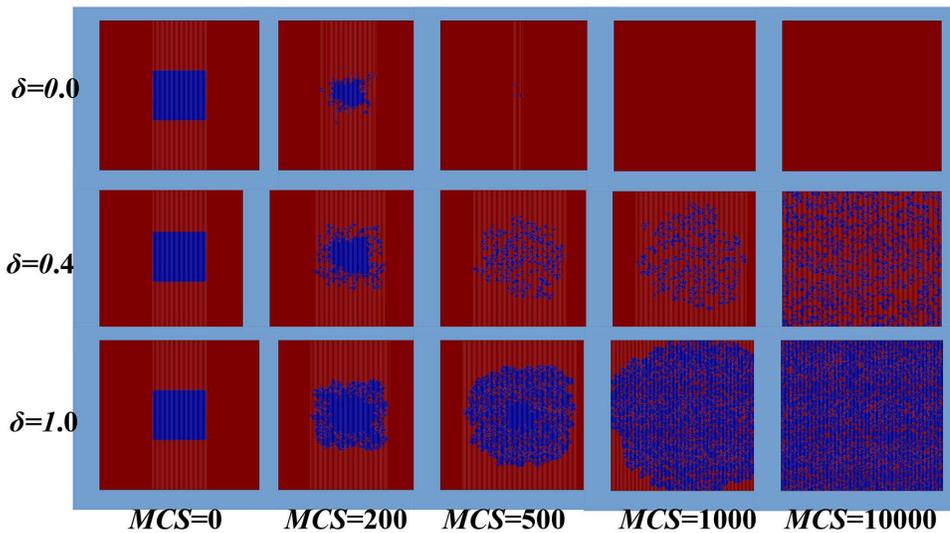


Fig 3. Shows snapshots of spatial distributions for blue cooperators and purple defectors at different MCS for various δ value. Vertically, the diffusion of cooperative behavior accelerates with increasing δ . Simulations indicate that larger δ significantly promotes cooperation. Parameters are set to $b = 1.1$, $\alpha = 0.5$, $\beta = 2$ and $L = 200$.

<https://doi.org/10.1371/journal.pone.0341925.g003>

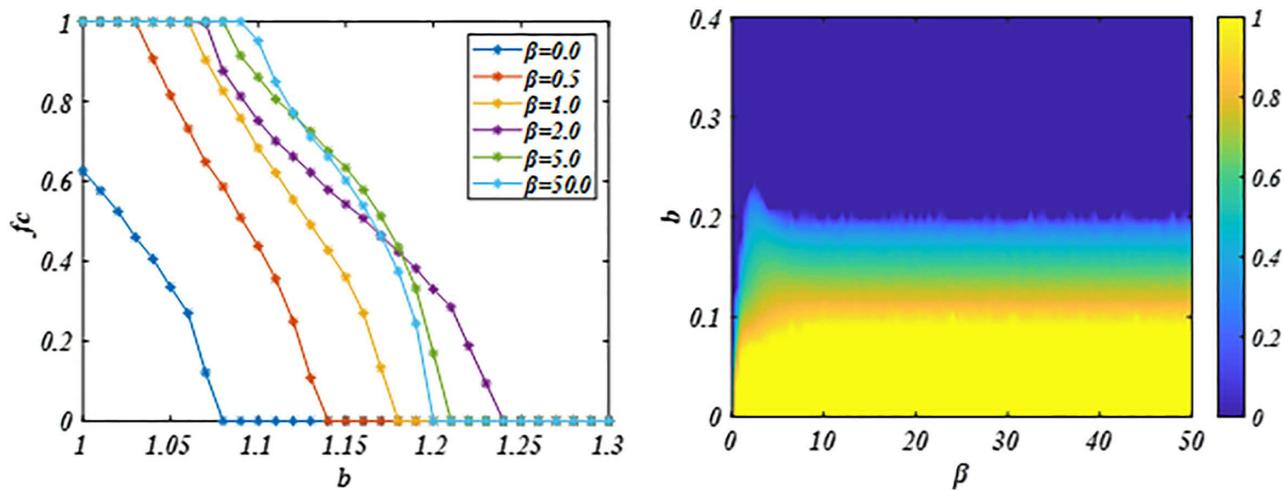


Fig 4. Panel (a) plots the fc against the b for different β values. Panel (b) demonstrates the dynamic impact of the β on cooperation evolution, revealing that while higher β values promote cooperation, the effect is only significant up to a certain threshold. The results indicate that a modestly sensitivity parameter β can promote cooperation. Parameters are set to $\alpha = 0.5$, $\delta = 1.0$ and $L = 200$.

<https://doi.org/10.1371/journal.pone.0341925.g004>

can achieving system goals. This overestimation of defection’s benefits facilitates its spread, thereby reducing overall cooperation. Fig 4(b) reveals the nonlinear dynamic influence of β on the evolution of cooperation. Thus, we can know that a modestly sensitivity parameter β can promote cooperation.

Fig 5(a) further examines the influence of the aspiration level α on the evolution of cooperation. The results show that lower system goals improve goal fulfillment among cooperators, thus promoting cooperation. However, an increase in the temptation b triggers an invasion of defectors, leading to a monotonic decline in system-wide cooperation. Despite this

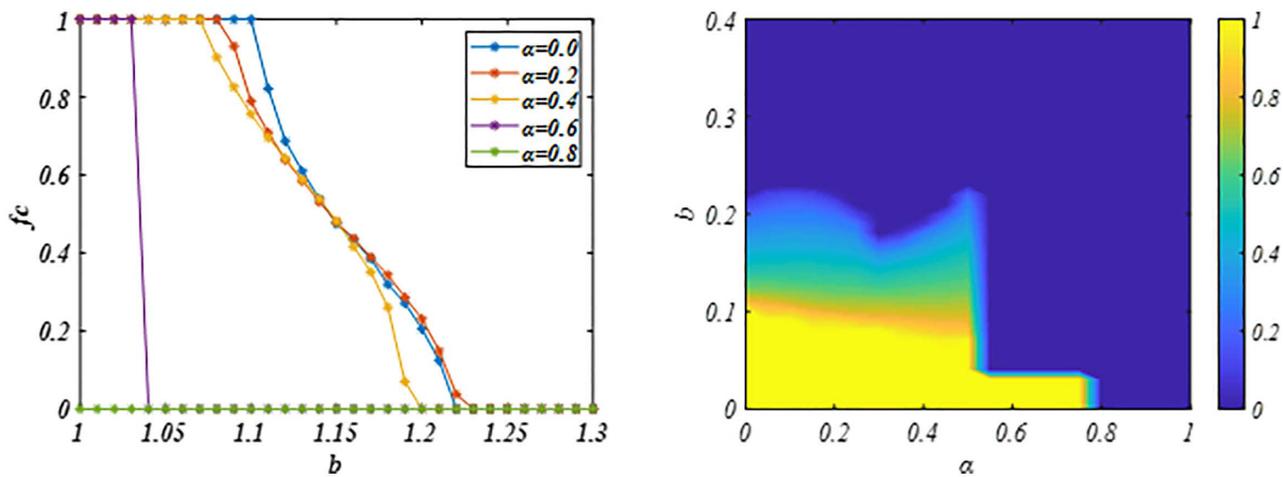


Fig 5. Panel (a) plots the f_c against the b under different α values, Panel (b) shows the dynamic impact of α on the cooperative evolution, demonstrating that high consistency aspiration or goal suppress the propagation of cooperative behavior in the system. The results indicate that lower system consistency desires or goals can promote cooperation. Parameters are set to $\beta=2$, $\delta=1.0$ and $L=200$.

<https://doi.org/10.1371/journal.pone.0341925.g005>

decrease, cooperation remains substantially higher than in traditional scenarios. Furthermore, cooperation diminishes as α rises. This is likely because higher α values make it more difficult for cooperative individuals to meet expectations. Under such evaluation criteria, the performance of individuals who fail to achieve α is accentuated, which hinders the propagation of cooperation. Therefore, as illustrated in Fig 5(b), setting lower system aspiration levels (or goals) can facilitate the spread of cooperation within this framework.

Further analysis in Fig 6 explores the dynamic influence of parameters β and δ on the evolution of cooperation. Horizontally, increasing the sensitivity parameter β alleviates the negative effect of aspiration level setting on cooperation

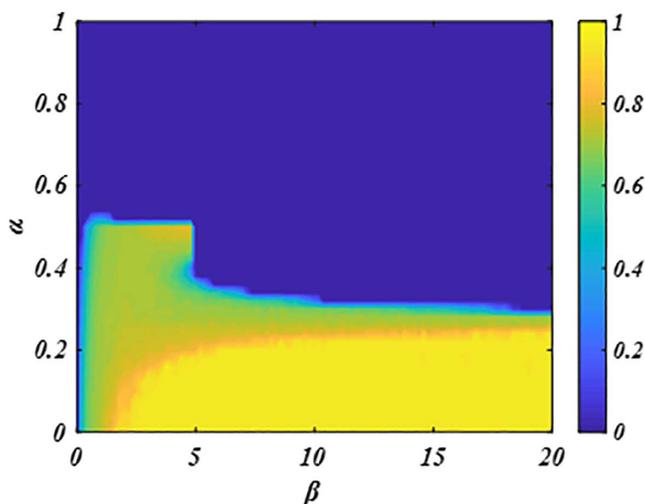


Fig 6. Illustrates the dynamic evolution of f_c with parameter β - α . Horizontally, system cooperation increases as β increases, whereas vertically, it decreases as α increases. The results indicate that parameters β and δ jointly determine the cooperative evolution in the system. Parameters are set to $\delta=1.0$ and $L=200$.

<https://doi.org/10.1371/journal.pone.0341925.g006>

within the system. Although this enhancement promotes collaborative behavior, its extent remains limited. Vertically, establishing appropriate aspiration levels is crucial for both initiating and maintaining cooperation. Thus, these parameters jointly shape the evolution of cooperation in the system.

To better illustrate the impact of parameter K on cooperative evolution, we calculated the evolutionary dynamics of the system for several values of K . As shown in Fig 7, the results indicate that the system maintains robustness under small noise fluctuations. However, when the noise amplitude becomes large, cooperative evolution is significantly disrupted. This finding is consistent with the conclusion reported by [40]. Moreover, analysis of the evolution of system cooperation across different network sizes ($L = 50$ – 300) shows consistent cooperation evolution, indicating that the system is robust.

4. Conclusion

Cooperation is recognised as foundation for sustaining socio-economic development. Building upon prior research and empirical observations of individual behavior within real social systems, this study introduces a novel strategy update rule. This rule incorporates a system-wide consensus objective and employs an individual's historical attainment of this objective as a key performance metric, calculated using the Bush-Mosteller (BM) model within a reinforcement learning framework.

Results demonstrate that, compared to conventional update mechanisms, the proposed rule significantly amplifies cooperative behavior within the system. This enhancement stems from the rule's systematic devaluation of defection payoffs while concurrently amplifying the perceived benefits of cooperation. This direct mechanism effectively suppresses the proliferation of defection strategies and accelerates the diffusion of cooperative ones. Furthermore, the inherent multidimensionality of the composite performance metric constrains opportunities for defectors by selectively filtering potential imitators. This aligns with empirical evidence indicating that social evaluations rarely hinge on a singular criterion. Critically, establishing an appropriate evaluative benchmark directly fosters cooperative outcomes within groups, and the methodology for behavioral assessment relative to this benchmark is paramount for optimizing group management. Collectively, our findings suggest that multidimensional evaluation creates more favorable conditions for the emergence and persistence of cooperation within complex environmental systems. This research offers theoretical insights into the mechanisms underpinning cooperative behavior in collective settings.

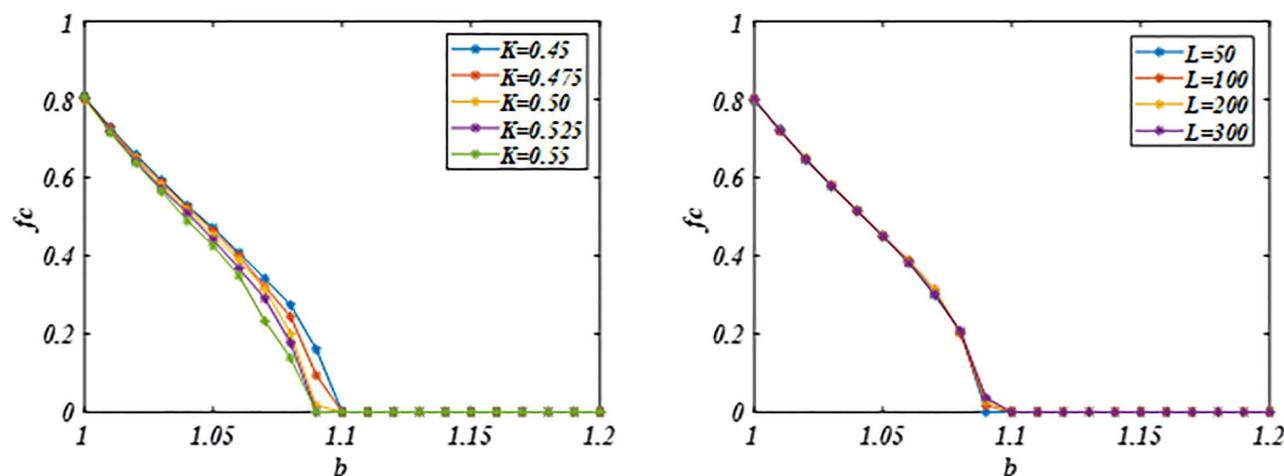


Fig 7. Panel (a) plots the f_c against the b under different K values. Panel (b) plots the f_c against the b under different L values. The result is consistent. The system exhibits robustness against minor noise interference and differences in network scale. Which indicating that the system is robust. Parameters are set to $\beta=2$, $\alpha=0.5$, $\delta=0.2$ and $L=200$ in Panel (a), and $\beta=2$, $\alpha=0.5$, $\delta=0.2$ and $K=0.5$ in Panel (b).

<https://doi.org/10.1371/journal.pone.0341925.g007>

Supporting information

S1 File. This code is used to calculate [Fig 1](#).
(TXT)

S2 File. This code is used to calculate [Fig 2](#).
(TXT)

S3 File. This code is used to calculate [Fig 3](#).
(TXT)

S4 File. This code is used to calculate [Fig 4\(a\)](#).
(TXT)

S5 File. This code is used to calculate [Fig 4\(b\)](#).
(TXT)

S6 File. This code is used to calculate [Fig 5\(a\)](#).
(TXT)

S7 File. This code is used to calculate [Fig 5\(b\)](#).
(TXT)

S8 File. This code is used to calculate [Fig 6](#).
(TXT)

S9 File. This code is used to calculate [Fig 7\(a\)](#).
(TXT)

S10 File. This code is used to calculate [Fig 7\(b\)](#).
(TXT)

S11 File. Data of all figure.
(ZIP)

Author contributions

Conceptualization: Xingchen Yu.

Data curation: Yang Wang.

Funding acquisition: Yang Wang.

Methodology: Yang Wang, Shounan Lu.

Project administration: Yang Wang, Shounan Lu.

Resources: Xingchen Yu, Shounan Lu.

Software: Yang Wang, Xingchen Yu, Shounan Lu.

Supervision: Shounan Lu.

Validation: Yang Wang.

Writing – original draft: Yang Wang, Shounan Lu.

Writing – review & editing: Shounan Lu.

References

1. Hauert C, Szabó G. Game theory and physics. *Am J Phys*. 2005;73(5):405–14. <https://doi.org/10.1119/1.1848514>
2. Nowak MA. Five rules for the evolution of cooperation. *Science*. 2006;314(5805):1560–3. <https://doi.org/10.1126/science.1133755> PMID: [17158317](https://pubmed.ncbi.nlm.nih.gov/17158317/)
3. Nowak MA, May RM. Evolutionary games and spatial chaos. *Nature*. 1992;359(6398):826–9. <https://doi.org/10.1038/359826a0>
4. Cressman R, Wu J-J, Li C, Tao Y. Game experiments on cooperation through reward and punishment. *Biol Theory*. 2013;8(2):158–66. <https://doi.org/10.1007/s13752-013-0106-2>
5. Sigmund K, Hauert C, Nowak MA. Reward and punishment. *Proc Natl Acad Sci*. 2001;98(19):10757–62.
6. Liu L, Wang L, Niu W, Hua S. Dynamic sanctioning mechanism for cooperative multi-agent systems. *Exp Syst Appl*. 2026;296:128873. <https://doi.org/10.1016/j.eswa.2025.128873>
7. Hua S, Liu L. Coevolutionary dynamics of population and institutional rewards in public goods games. *Exp Syst Appl*. 2024;237:121579. <https://doi.org/10.1016/j.eswa.2023.121579>
8. Ding R, Wang X, Liu Y, Zhao J, Gu C. Evolutionary games with environmental feedbacks under an external incentive mechanism. *Chaos Solitons Fractals*. 2023;169:113318. <https://doi.org/10.1016/j.chaos.2023.113318>
9. Chen Y-D, Guan J-Y, Wu Z-X. Coevolutionary game dynamics with localized environmental resource feedback. *Phys Rev E*. 2025;111(2–1):024305. <https://doi.org/10.1103/PhysRevE.111.024305> PMID: [40103166](https://pubmed.ncbi.nlm.nih.gov/40103166/)
10. Perc M, Szolnoki A. Social diversity and promotion of cooperation in the spatial prisoner's dilemma game. *Phys Rev E Stat Nonlin Soft Matter Phys*. 2008;77(1 Pt 1):011904. <https://doi.org/10.1103/PhysRevE.77.011904> PMID: [18351873](https://pubmed.ncbi.nlm.nih.gov/18351873/)
11. Szolnoki A, Perc M. Coevolution of teaching activity promotes cooperation. *New J Phys*. 2008;10(4):043036. <https://doi.org/10.1088/1367-2630/10/4/043036>
12. Szolnoki A, Szabó G. Cooperation enhanced by inhomogeneous activity of teaching for evolutionary Prisoner's Dilemma games. *Europhys Lett*. 2007;77(3):30004. <https://doi.org/10.1209/0295-5075/77/30004>
13. Wang Z, Wang L, Yin Z-Y, Xia C-Y. Inferring reputation promotes the evolution of cooperation in spatial social dilemma games. *PLoS One*. 2012;7(7):e40218. <https://doi.org/10.1371/journal.pone.0040218> PMID: [22808120](https://pubmed.ncbi.nlm.nih.gov/22808120/)
14. Liu L, Chen X, Szolnoki A. Coevolutionary dynamics via adaptive feedback in collective-risk social dilemma game. *Elife*. 2023;12:e82954. <https://doi.org/10.7554/eLife.82954> PMID: [37204305](https://pubmed.ncbi.nlm.nih.gov/37204305/)
15. Han X, Zhao X, Xia H. Hybrid learning promotes cooperation in the spatial prisoner's dilemma game. *Chaos Solitons Fractals*. 2022;164:112684. <https://doi.org/10.1016/j.chaos.2022.112684>
16. Bi Y, Yang H. Based on reputation consistent strategy times promotes cooperation in spatial prisoner's dilemma game. *Appl Math Comput*. 2023;444:127818. <https://doi.org/10.1016/j.amc.2022.127818>
17. Zhang H, An T, Wang J, Wang L, An J, Zhao J, et al. Reputation-based adaptive strategy persistence can promote cooperation considering the actual influence of individual behavior. *Phys Lett A*. 2024;508:129495. <https://doi.org/10.1016/j.physleta.2024.129495>
18. He J, Wang J, Yu F, Zheng L. Reputation-based strategy persistence promotes cooperation in spatial social dilemma. *Phys Lett A*. 2020;384(27):126703. <https://doi.org/10.1016/j.physleta.2020.126703>
19. Lu S, Wang Y. Past-performance-driven strategy updating promote cooperation in the spatial prisoner's dilemma game. *Appl Math Comput*. 2025;491:129220. <https://doi.org/10.1016/j.amc.2024.129220>
20. Xu J, Deng Z, Gao B, Song Q, Tian Z, Wang Q, et al. Popularity-driven strategy updating rule promotes cooperation in the spatial prisoner's dilemma game. *Appl Math Comput*. 2019;353:82–7. <https://doi.org/10.1016/j.amc.2019.01.060>
21. Lu S, Wang Y. Experience-driven learning and interactive rules under link weight adjustment promote cooperation in spatial prisoner's dilemma game. *Appl Math Comput*. 2025;497:129381. <https://doi.org/10.1016/j.amc.2025.129381>
22. Jia D, Guo H, Song Z, Shi L, Deng X, Perc M, et al. Local and global stimuli in reinforcement learning. *New J Phys*. 2021;23(8):083020. <https://doi.org/10.1088/1367-2630/ac170a>
23. Masuda N, Nakamura M. Numerical analysis of a reinforcement learning model with the dynamic aspiration level in the iterated Prisoner's dilemma. *J Theor Biol*. 2011;278(1):55–62. <https://doi.org/10.1016/j.jtbi.2011.03.005> PMID: [21397610](https://pubmed.ncbi.nlm.nih.gov/21397610/)
24. Ezaki T, Horita Y, Takezawa M, Masuda N. Reinforcement learning explains conditional cooperation and its moody cousin. *PLoS Comput Biol*. 2016;12(7):e1005034. <https://doi.org/10.1371/journal.pcbi.1005034> PMID: [27438888](https://pubmed.ncbi.nlm.nih.gov/27438888/)
25. Jia D, Li T, Zhao Y, Zhang X, Wang Z. Empty nodes affect conditional cooperation under reinforcement learning. *Appl Math Comput*. 2022;413:126658. <https://doi.org/10.1016/j.amc.2021.126658>
26. Geng Y, Liu Y, Lu Y, Shen C, Shi L. Reinforcement learning explains various conditional cooperation. *Appl Math Comput*. 2022;427:127182. <https://doi.org/10.1016/j.amc.2022.127182>
27. Zhang L, Li Y, Xie Y, Feng Y, Huang C. The combined effects of conformity and reinforcement learning on the evolution of cooperation in public goods games. *Chaos Solitons Fractals*. 2025;193:116071. <https://doi.org/10.1016/j.chaos.2025.116071>

28. Horita Y, Takezawa M, Inukai K, Kita T, Masuda N. Reinforcement learning accounts for moody conditional cooperation behavior: experimental results. *Sci Rep.* 2017;7:39275. <https://doi.org/10.1038/srep39275> PMID: [28071646](https://pubmed.ncbi.nlm.nih.gov/28071646/)
29. Tampuu A, Matiisen T, Kodelja D, Kuzovkin I, Korjus K, Aru J, et al. Multiagent cooperation and competition with deep reinforcement learning. *PLoS One.* 2017;12(4):e0172395. <https://doi.org/10.1371/journal.pone.0172395> PMID: [28380078](https://pubmed.ncbi.nlm.nih.gov/28380078/)
30. Ding Z-W, Zheng G-Z, Cai C-R, Cai W-R, Chen L, Zhang J-Q, et al. Emergence of cooperation in two-agent repeated games with reinforcement learning. *Chaos Solitons Fractals.* 2023;175:114032. <https://doi.org/10.1016/j.chaos.2023.114032>
31. Bendor J, Mookherjee D, Ray D. Aspiration-based reinforcement learning in repeated interaction games: an overview. *Int Game Theory Rev.* 2001;03(02n03):159–74. <https://doi.org/10.1142/s0219198901000348>
32. Zhang L, Huang C, Li H, Dai Q, Yang J. Cooperation guided by imitation, aspiration and conformity-driven dynamics in evolutionary games. *Phys A: Stat Mech Appl.* 2021;561:125260. <https://doi.org/10.1016/j.physa.2020.125260>
33. Perc M, Wang Z. Heterogeneous aspirations promote cooperation in the prisoner's dilemma game. *PLoS One.* 2010;5(12):e15117. <https://doi.org/10.1371/journal.pone.0015117> PMID: [21151898](https://pubmed.ncbi.nlm.nih.gov/21151898/)
34. Li Z, Yang Z, Wu T, Wang L. Aspiration-based partner switching boosts cooperation in social dilemmas. *PLoS One.* 2014;9(6):e97866. <https://doi.org/10.1371/journal.pone.0097866> PMID: [24896269](https://pubmed.ncbi.nlm.nih.gov/24896269/)
35. Liu X, He M, Kang Y, Pan Q. Aspiration promotes cooperation in the prisoner's dilemma game with the imitation rule. *Phys Rev E.* 2016;94(1–1):012124. <https://doi.org/10.1103/PhysRevE.94.012124> PMID: [27575094](https://pubmed.ncbi.nlm.nih.gov/27575094/)
36. You T, Shi L, Wang X, Mengibaev M, Zhang Y, Zhang P. The effects of aspiration under multiple strategy updating rules on cooperation in prisoner's dilemma game. *Appl Math Comput.* 2021;394:125770. <https://doi.org/10.1016/j.amc.2020.125770>
37. Chen Y-S, Yang H-X, Guo W-Z, Liu G-G. Promotion of cooperation based on swarm intelligence in spatial public goods games. *Appl Math Comput.* 2018;320:614–20. <https://doi.org/10.1016/j.amc.2017.10.022>
38. Lu S, Dai J, Zhu G, Guo L. Investigating the effectiveness of interaction-efficiency-driven strategy updating under progressive-interaction for the evolution of the prisoner's dilemma game. *Chaos Solitons Fractals.* 2023;172:113493. <https://doi.org/10.1016/j.chaos.2023.113493>
39. Wang J, He J, Yu F. Heterogeneity of reputation increment driven by individual influence promotes cooperation in spatial social dilemma. *Chaos Solitons Fractals.* 2021;146:110887. <https://doi.org/10.1016/j.chaos.2021.110887>
40. Szabó G, Vukov J, Szolnoki A. Phase diagrams for an evolutionary prisoner's dilemma game on two-dimensional lattices. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2005;72(4 Pt 2):047107. <https://doi.org/10.1103/PhysRevE.72.047107> PMID: [16383580](https://pubmed.ncbi.nlm.nih.gov/16383580/)
41. Perc M, Szolnoki A, Szabó G. Restricted connections among distinguished players support cooperation. *Phys Rev E Stat Nonlin Soft Matter Phys.* 2008;78(6 Pt 2):066101. <https://doi.org/10.1103/PhysRevE.78.066101> PMID: [19256899](https://pubmed.ncbi.nlm.nih.gov/19256899/)
42. Wang Z, Szolnoki A, Perc M. Interdependent network reciprocity in evolutionary games. *Sci Rep.* 2013;3:1183. <https://doi.org/10.1038/srep01183> PMID: [23378915](https://pubmed.ncbi.nlm.nih.gov/23378915/)
43. Szolnoki A, Perc M. Promoting cooperation in social dilemmas via simple coevolutionary rules. *Eur Phys J B.* 2008;67(3):337–44. <https://doi.org/10.1140/epjb/e2008-00470-8>
44. Perc M, Szolnoki A. Social diversity and promotion of cooperation in the spatial prisoner's dilemma game[J]. *Physical Review E-Statistical, Nonlinear, and Soft Matter Physics*, 2008;77(1):01190.