Applied Mathematics and Computation 504 (2025) 129505

Contents lists available at ScienceDirect

Applied Mathematics and Computation

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

Full Length Article

Environmental information perception enhances cooperation in stochastic public goods games via Q-learning

Yipeng Li [a], Xiangyue Hu [b], Xing Jin [a,c,*], Huizhen Zhang [a], Jiajia Yang [a], Zhen Wang [a,c,*]

[a] School of Cyberspace, Hangzhou Dianzi University, Hangzhou, 310018, Zhejiang, China
[b] Zhuoyue Honors College, Hangzhou Dianzi University, Hangzhou, 310018, Zhejiang, China
[c] Experimental Center of Data Science and Intelligent Decision-Making, Hangzhou Dianzi University, Hangzhou, 310018, Zhejiang, China

ARTICLE INFO

Keywords:
Stochastic public goods games
Environmental information perception
Q-learning
Cooperation

ABSTRACT

Cooperation is the foundation of social progress, but due to rational individuals often prioritize personal interests, reciprocal cooperation is undermined. The Public Goods Game (PGG) is a classic model for studying group interactions. Traditional PGG assumes a static environment, but in reality, the environment is dynamically changing, and there is an interaction between individual behavior and the environment. Therefore, the stochastic game framework is proposed and applied to study the feedback mechanisms between behavior and the environment. This paper takes the two-state environmental transition mechanism as an example to explore the impact of environmental information perception ability on individual decision-making in the stochastic PGG. Specifically, we use the Q-learning algorithm to depict individual decision-making behavior and consider two types of individuals with different perception abilities: individuals with environmental perception ability select the best action based on the current environmental state, while individuals without environmental perception ability make decisions based solely on historical experience. The experimental results show that environmental information perception significantly lowers the cooperation threshold in the stochastic PGG. By analyzing the microscopic interaction modes of individuals, we find that there is an isolation zone effect between different strategy populations, which effectively prevents the erosion of defection behaviors and ensures the internal stability of cooperation. The extended experiments further validate the robustness of the results. This study shows that environmental information is beneficial for promoting the evolution of cooperation. These findings provide new insights into the cooperation mechanisms in stochastic PGG and offer valuable guidance for promoting cooperation in real-world societies.

1. Introduction

Cooperation runs through both macro- and micro-life systems, serving as a cornerstone for the progress of human society. It is an interdisciplinary issue spanning sociobiology, physics, and economics [1–4]. Individuals who only pursue personal goals may lead

* Corresponding authors at: School of Cyberspace, Hangzhou Dianzi University, Hangzhou, 310018, Zhejiang, China.
E-mail addresses: liyipeng@hdu.edu.cn (Y. Li), huxiangyue@hdu.edu.cn (X. Hu), jinxing@hdu.edu.cn (X. Jin), zhanghz9436@gmail.com (H. Zhang), jiajiayang@hdu.edu.cn (J. Yang), wangzhen@hdu.edu.cn (Z. Wang).

https://doi.org/10.1016/j.amc.2025.129505
Received 18 February 2025; Received in revised form 25 April 2025; Accepted 1 May 2025

Available online 16 May 2025

to the loss of mutually beneficial cooperation [5]. Humanity has long faced this dilemma, particularly in socio-economic relations. Therefore, extensive research has been conducted to explore the emergence mechanisms of cooperation in social dilemma games [6].

The Public Goods Game (PGG) [7–9] is one of the classical models used to study group interactions. In this model, the total cost of the contributions of the cooperators, multiplied by a synergy factor r, constitutes the overall benefit, which is then equally distributed among all participants. However, this game model is prone to the "free-rider" problem [10], where some individuals benefit from the cooperation of others without contributing themselves.

To address this issue, numerous studies have explored various incentive mechanisms to promote cooperation among individuals, such as punishment [11–15], rewards [16,17], reputation [18,19], and conditional strategies [20]. Traditional PGG models assume that each individual can interact with all other participants. However, in real-world scenarios, individuals typically interact only with specific neighbors, such as friends, colleagues, or family members within a social network. To account for this, the Spatial Public Goods Game (SPGG) [21–24] was introduced to investigate the evolution of cooperation within structured groups.

In previous studies, stochasticity has been introduced into the public goods game, including factors such as probabilistic strategy exploration under decision-making uncertainty [25], heterogeneous payoff structures with dynamic adjustments [26], and randomized partner matching mechanisms [27]. However, these studies did not account for the stochastic changes in the environmental state, such as the impact of individual behavior on the environment and the feedback loop where the environment, in turn, influences individual decision-making. To explore this, the concept of the stochastic PGG was introduced [28–30]. Quan et al. [31] incorporated a scale return coefficient with reputation-based environmental feedback into the SPGG. Their simulations demonstrated that this innovation promotes cooperation, reduces the critical value of the enhancement factor, and influences cluster formation. Wang et al. [32] integrated the PGG with dynamic environmental feedback mechanisms, proposing a generalized ecological evolutionary game model. By introducing environmental factors that induce environmental transitions, Lyu et al. [33] observed that as the initial environmental factors and the cooperation-enhancement defection-degradation ratio increase, the steady cooperation level of the social network improves significantly, ultimately leading to a high-return environment. By employing a mimicry mechanism combined with the Win-Stay Lose-Shift (WSLS) strategy, Hilbe et al. [34] found that the dependency of the public resource on previous interactions can greatly enhance the propensity for cooperation. By adjusting the synergy factor r, Yang et al. [35] simulated the impact of environmental changes on individual behavior, thereby influencing the dynamics of cooperation. Ma et al. [36] emphasized the critical role of nonlinear environmental feedback, demonstrating that adjusting environmental feedback parameters can effectively incentivize cooperative behavior and foster both the stability and evolution of cooperation across varying environmental conditions.

Existing research on the SPGG has not fully considered the impact of environmental information perception on individual decision-making. For example, the common WSLS strategy [37] and the reputation-based scale return coefficient mechanism [31] rely solely on payoff judgments and do not incorporate environmental perception into decision-making considerations. Recent studies [38] have explored how the availability of information shapes the evolution of cooperation in the repeated prisoner's dilemma, revealing that cooperation sometimes occurs only when the environmental state information is precise, while sometimes it thrives when there is a lack of environmental state information. However, in the work on the PGG, the influence of information on decision-making remains an unclear issue.

Therefore, this study explores the impact of environmental information perception ability on the emergence of cooperation in the SPGG. We consider a stochastic PGG model, where the system enters the more valuable game in the next round when an individual has three or more cooperative neighbors, and enters to the less valuable game when the number of cooperative neighbors is fewer than three. Specifically, we use the Q-learning algorithm [39,40] to simulate individual strategy updates in different environments, and considering two types of individuals: one type has environmental perception ability and selects the optimal action based on the current environmental state, with the Q-table containing both current environmental state information and historical experience; the other type lacks environmental perception ability and makes decisions solely based on historical experience, with the Q-table containing only historical action choices. Based on this, we compare three systems: the Static Environment (SE), the environment Transitions but is Not Perceived by individuals (TNP), and another where the environment Transitions and is Perceived by individuals (TP).

The experimental results show that: (1) The cooperation threshold in TNP is later than in SE. This is because, in TNP, by studying the evolutionary relationship of the populations, it is found that the upper limit of the cooperation Q-value is lower than that of the defection Q-value, causing individuals to tend to choose defection. As a result, each individual makes strategy choices only in the less valuable game; (2) The cooperation threshold in TP is earlier than in SE. This is because, in TP, through microscopic analysis, it is found that in cluster-to-cluster contact, the isolation zone phenomenon emerges at the cluster boundary during the strategy evolution process, which allows individuals within the isolation zone to effectively resist the erosion of defection strategies. In the case of individual-to-individual contact, since each individual is uniformly distributed on the grid, there are at least two cooperative individuals in each group participating in the PGG, thereby stabilizing cooperative behavior in the system; (3) Further extended experiments are conducted from six dimensions: synchronous update interactions; different r-value differences between the more valuable game and the less valuable game; a more stringent environmental transition mechanism; different Q-learning state paradigms; a larger von Neumann neighborhood range; and different network topologies. The results consistently confirm the important role of environmental information in cooperation evolution.

2. Model

Consider a public goods game (PGG), where individuals are distributed on an $L \times L$ square lattice network with periodic boundary conditions. Each individual interacts only with its neighbors in the von Neumann neighborhood [41,42]. In the model, individuals

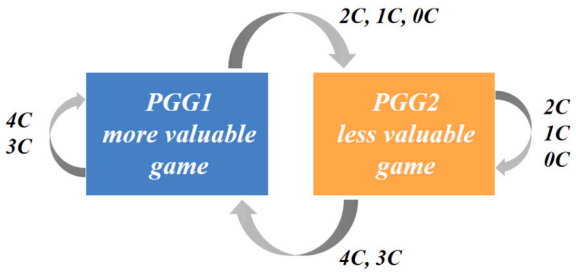

Fig. 1. A two-state stochastic public goods game model. PGG1 (left, in blue) represents the more valuable game, and **PGG2** (right, in orange) represents the less valuable game.

make a choice between cooperation (C) or defection (D) based on their environment and the behavior of their neighbors, i.e., the action set $A = \{C, D\}$.

In a PGG initiated by player i, five participants (player i and its four von Neumann neighbors), contribute to a common pool. The total contribution is multiplied by a synergy factor r, and the resulting amount is equally distributed among all participants. Therefore, the payoff formulas for cooperation and defection are as follows:

$$\pi_C = \frac{r \cdot n_C}{n_C + n_D} - c \tag{1}$$

$$\pi_D = \frac{r \cdot n_C}{n_C + n_D} \tag{2}$$

where n_C and n_D denote the number of cooperators and defectors in the group, and c represents the cost associated with cooperation.

In real-world dynamic systems, such as the abundance of public resources, there is often a characteristic where cooperation increases the resource level, creating a positive feedback loop that enriches the environment and enhances future payoffs [16,17]. In contrast, defection reduces the resource pool, leading to a negative feedback loop that depletes the shared environment and diminishes future benefits. To model this, we adopt a two-state stochastic public goods game (PGG), following prior work [34,38]. Through this dynamic, cooperation allows players to engage in increasingly valuable interactions, while defection leads to a gradual decline in game value. In each round, players can be in one of the two possible games, $S = \{s_1, s_2\}$, where s_1 represents the more valuable game, with a synergy factor of r_1, s_2 represents the less valuable game, with a synergy factor of r_2 ($r_1 > r_2$). The level of cooperation in the environment can be partially characterized by the number of cooperators in the environment. Therefore, in our model, the environmental switching is driven by the number of cooperative neighbors [34,38]. In summary, a two-state Markov chain [43,44] can be constructed to formally describe this stochastic system. Specifically, as shown in the Fig. 1, when an individual has three or more cooperative neighbors, the system transitions to the more valuable game in the next round. The transition process for this condition can be mathematically expressed as:

$$P(S_{t+1} = s_1 | n_C(t) \geq 3, S_t = s_1 \text{ or } S_t = s_2) = 1, \tag{3}$$

where S_{t+1} represents the state of the system at the next time step, and $n_C(t)$ denotes the number of cooperative neighbors that an individual has at time t.

If the number of cooperative neighbors is fewer than three, the system enters the less valuable game. The transition process for this condition can be expressed as:

$$P(S_{t+1} = s_2 | n_C(t) < 3, S_t = s_1 \text{ or } S_t = s_2) = 1. \tag{4}$$

To investigate the impact of environmental information perception on cooperation, two settings are compared. In the perceiving setting, players are able to obtain the current environmental state before making decisions. Therefore, their strategies are influenced by the current state. In the non-perceiving setting, individuals are unable to access the current environmental information.

The goal of individuals is to maximize their payoffs, and the Q-learning algorithm is applied for strategy updates [45,46], with each individual using a Q-table to guide their choices. In the perceiving setting, the Q-table is a table with two dimensions, consisting of states and actions:

$$Q_{s,a}(t) = \begin{bmatrix} Q_{s_1,C}(t) & Q_{s_1,D}(t) \\ Q_{s_2,C}(t) & Q_{s_2,D}(t) \end{bmatrix} \tag{5}$$

where $Q_{s,a}(t)$, stored in the Q-table, denotes the expected reward for choosing action a in the current state s at time step t. Individuals then choose the action with the highest Q-value in the corresponding state based on the Q-table.

For the non-perceiving setting, the Q-table is a one-dimensional table that only includes actions:

$$Q_a(t) = \begin{bmatrix} Q_C(t) & Q_D(t) \end{bmatrix} \tag{6}$$

where $Q_a(t)$ represents the expected reward of the individual's choice of action a at time step t. In this setting, agents select either cooperation (C) or defection (D) based on which action has the higher Q-value.

Without loss of generality, each element in the Q-table is initialized to 0, and each individual is initially assigned a random action from the action set. In the Monte Carlo simulation, the evolution proceeds through an asynchronous updating process [47–49]. In each iteration, $L \times L$ individual interactions are performed, with one central individual i randomly selected at each time. Based on the current environmental state, the individual makes action decisions with its four neighbors and obtains payoffs through interaction with them. Subsequently, all participants update their Q-tables based on their respective immediate payoffs.

The Q-value update formula for the perceiving setting is as follows:

$$Q^i_{s,a}(t+1) = (1-\alpha)Q^i_{s,a}(t) + \alpha\left(\pi_i(t) + \gamma \max_{a'} Q^i_{s',a'}(t)\right) \tag{7}$$

where $\max_{a'} Q^i_{s',a'}(t)$ denotes the maximum Q-value in the next state s', α $(0 < \alpha \leq 1)$ denotes the learning rate, γ $(0 \leq \gamma < 1)$ denotes the discount factor, which determines the proportion of future rewards for updating the strategy, and $\pi_i(t)$ represents the reward obtained by individual i at time step t when taking action a in state s.

In order to isolate and examine the role of environmental information in shaping cooperative behavior and to minimize the influence of other variables, we have implemented a simplified Q-learning approach specifically for non-perceiving setting, focusing on the core question of how environmental perception affects cooperation dynamics. The formula is as follows:

$$Q^i_a(t+1) = (1-\alpha)Q^i_a(t) + \alpha\pi_i(t) \tag{8}$$

To balance the exploration of unknown states with the exploitation of existing experience, individuals can utilize both their experience and explore unknown states during the decision-making process. Therefore, to prevent local equilibrium, the ϵ-greedy algorithm is employed [50]. Each individual has two options when selecting an action: with probability ϵ, the action is chosen randomly, or with probability $1-\epsilon$, the action corresponding to the maximum Q-value in the current state as indicated by the Q-table is selected. It is important to note that there may be multiple actions with the highest Q-value, in which case the individual randomly selects one for learning. The detailed evolution process is provided in Algorithm 1.

Algorithm 1 Pseudocode of PGG with Q-learning and Environmental Information Perception.

Input: α, γ, ϵ, r_1, r_2, c
// Initialization
Initialize MC simulation step n;
for each node $i \in L \times L$ **do**
    Initialize each agent i's strategy C or D randomly;
    Initialize Q-table elements to 0;
end for
for each node $i \in L \times L$ **do**
    Initialize each agent i's environment based on the strategies of its neighbors;
end for
// Strategy Interaction
for each time step $t \in [1, n]$ **do**
    for $L \times L$ interactions **do**
        select an agent i randomly;
        get i's current environment state s_i;
        if prob $< \epsilon$ **then**
            select an action randomly from A;
        else
            select the action with the highest Q-value in the current state;
        end if
        calculate i's payoff π according to Eq. (1) and Eq. (2);
        update Q-table according to Eq. (7) and Eq. (8);
        update i's neighbors' Q-values similarly;
    end for
end for

3. Results

To investigate the impact of environmental information perception on the emergence of cooperation, we illustrate in the Fig. 2 the trends in cooperation rates as the synergy factor r varies across two types of systems: one where the environment Transitions but is Not Perceived by individuals (TNP), and another where the environment Transitions and is Perceived by individuals (TP). Additionally, for baseline comparison, we include the trends in cooperation rates as r varies in a Static Environment system(SE). To ensure the system reaches a stable state, the cooperation rates are averaged over the final 5000 time steps out of a total of 50000 time steps. To minimize the impact of uncertainty, the results are averaged across five independent experimental runs. For TNP and TP, the default setting assumes a difference of 0.2 between the synergy factors of the two environments, with the larger r value used as the horizontal axis in the Fig. 2.

The Fig. 2 demonstrates that in SE, individuals begin to cooperate when the synergy factor exceeds the critical threshold ($r_c^{\mathrm{SE}} \approx 4.50$). Two noteworthy phenomena are observed:

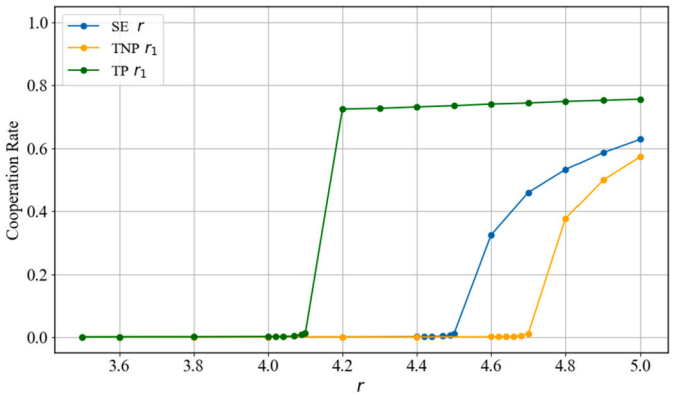

Fig. 2. Cooperation rate at convergence versus the synergy factor r in three systems. The cooperation emergence threshold is approximately $r_c = 4.50$ for SE, $r_c = 4.70$ for TNP, and $r_c = 4.10$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

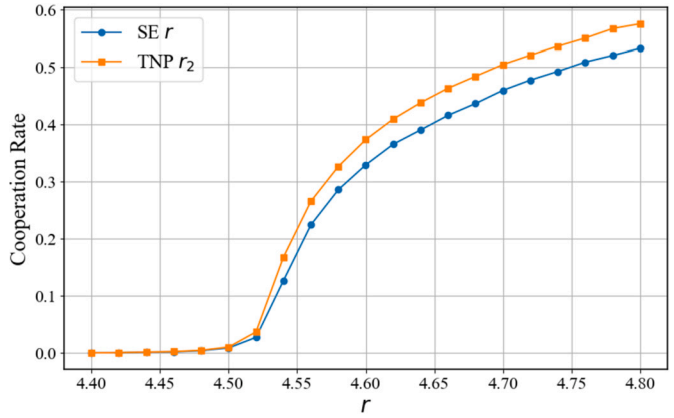

Fig. 3. Comparison of cooperation rates in SE and TNP. The **blue curve** represents the cooperation rate in SE. The **orange curve** represents the cooperation rates in TNP, corresponding to r_2 values. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

Phenomena 1: In TNP, the cooperation threshold is observed at $r_c^{\mathrm{TNP}} \approx 4.70$, which is higher than that in SE.

Phenomena 2: In TP, the cooperation threshold is lowered to $r_c^{\mathrm{TP}} \approx 4.10$, which is earlier than that in SE.

Therefore, under environmental transitions, if individuals cannot perceive environmental information (Phenomenon 1), the emergence of cooperation is delayed. In contrast, the ability of individuals to perceive environmental information (Phenomenon 2) effectively promotes the emergence of cooperative behavior within the system.

3.1. Explanation of Phenomenon 1

To further explore the differences between TNP and SE, the Fig. 3 illustrates the cooperation rate dynamics as a function of r_2 in TNP. This approach provides a focused perspective for observing and analyzing the distinctions between the two systems. As clearly shown in the Fig. 3, before reaching the threshold of 4.5, the blue (SE) and orange (TNP) curves almost completely overlap. After surpassing the threshold, the cooperation rate of the orange curve begins to slightly exceed that of the blue curve, with the difference gradually widening over time.

Based on the observed phenomenon of overlapping curves, we hypothesize that in systems with environmental transitions, although both more valuable and less valuable games exist, individuals predominantly remain in the less valuable game before r reaches the threshold. Once r surpasses the threshold, the system transitions into a state where the more valuable and less valuable games coexist. To validate this hypothesis, the dynamic changes in the number of individuals in the more valuable and less valuable games under conditions of TNP are shown in the Fig. 4. Experiments are conducted based on the following three parameter sets: $r_1 = 4.5, r_2 = 4.3$ (before reaching the cooperation threshold, Fig. 4(a)), $r_1 = 4.7, r_2 = 4.5$ (at the cooperation threshold, Fig. 4(b)), and $r_1 = 4.8, r_2 = 4.6$ (after exceeding the cooperation threshold, Fig. 4(c)).

From the Fig. 4, it can be observed that before the cooperation threshold $r = 4.5$ is reached (Fig. 4(a) and (b)), individuals in the system predominantly evolve into the less valuable game under stable conditions. This scenario closely resembles the behavior observed in SE ($r = 4.3$). This observation supports the hypothesis regarding the cooperation threshold in TNP: before r reaches the threshold, individuals almost entirely reside in the less valuable game. When r exceeds the threshold (Fig. 4(c)), the number of individuals in the more valuable game increases significantly. Since individual behavior is influenced by the actions of their neighbors,

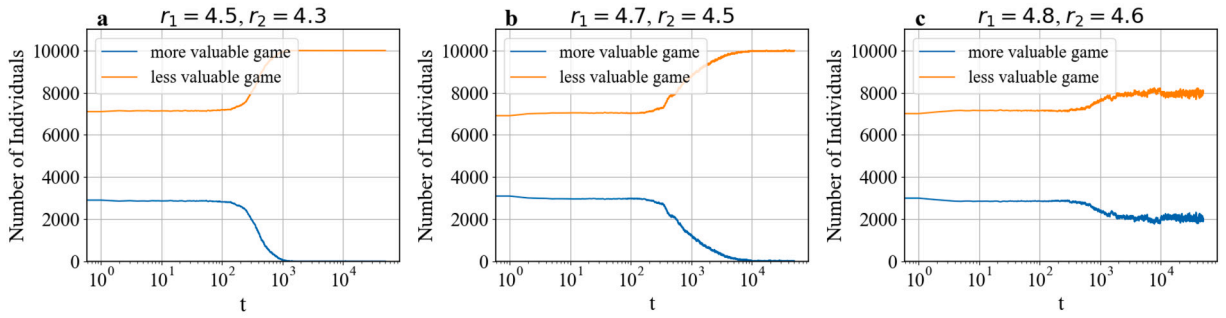

Fig. 4. Evolution of individual distributions in more valuable and less valuable games under different r_1 and r_2 values. The left, middle, and right panels illustrate the number of individuals in the **more valuable game** (blue line) and the **less valuable game** (orange line) over time steps. Each panel corresponds to different synergy factor pairs: **(a)** $r_1 = 4.5$, $r_2 = 4.3$, **(b)** $r_1 = 4.7$, $r_2 = 4.5$, and **(c)** $r_1 = 4.8$, $r_2 = 4.6$. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

statistical analysis based solely on quantity changes cannot fully explain this phenomenon. Therefore, it is necessary to further analyze the dynamic evolution of individual strategies from a spatial perspective to explore the underlying patterns. To achieve this, we focus on the strategy evolution under a two-dimensional grid, analyzing the dynamic changes in individual strategies from the perspective of spatial distribution. For individuals updating their decisions using the Q-learning algorithm, two categories can be identified based on their Q-table: cooperators ($Q_c > Q_d$) and defectors ($Q_c < Q_d$). To better understand the evolution and interaction of these two types of individuals, the grid is divided into two halves. The left half is initialized with cooperators ($Q_c = 0.5, Q_d = 0$, marked in blue), while the right half is initialized with defectors ($Q_c = 0, Q_d = 0.5$, marked in red).

The Fig. 5 demonstrates significant evolution in the spatial distribution of strategies over time. At $t = 500$, the blue cooperative regions are gradually invaded by the red defection strategy. By $t = 2000$, the cooperative regions become further fragmented, remaining only as isolated “islands”. Eventually, at $t = 45000$, the cooperative strategy nearly disappears, and the system transitions into a stable state dominated by the defection strategy.

From this, the following two findings can be summarized:

Finding 1: Defection clusters gradually invade cooperation clusters;

Finding 2: Cooperative individuals sporadically appear within defection clusters.

To explain Finding 1, we record the changes in the Q-values of individual 6 on the boundary during each step in the Fig. 6(b), and divide the process into three phases: Phase 1 (first 200 steps), Phase 2 (from 200 to 3000 steps), and Phase 3 (after 3000 steps).

For Individual 6, it participates in public goods games centered on Individuals 2, 5, 10, 7, and itself.

In the Phase 1, when Individual 6 participates in public goods games centered on Individuals 2, 5, 10, and itself, at least four individuals choose to cooperate, leading to a steady increase in its Q_c. However, when Individual 6 participates in strategy updates centered on Individual 7, where only itself chooses cooperation while all others choose defection, this causes a decrease in its Q_c. Consequently, in the early stages, Individual 6 typically experiences four increases and one decrease in Q_c per round. Additionally, due to the presence of the exploration rate, Individual 6 occasionally chooses defection. Since its neighbors are all cooperators, this leads to a staircase-like increase in its Q_d during the early stages.

In the Phase 2, oscillations in Q_c and Q_d emerge, primarily due to the upper limits on their growth. Specifically, the Q-value updates follow the formula:

$$Q_{\text{new}} = 0.1 \times \pi + 0.9 \times Q_{\text{old}} \tag{9}$$

Taking Individual 6 as an example, when it participates in public goods games centered on Individuals 2, 5, 10, and itself, the presence of four cooperative individuals in the group results in a maximum immediate payoff of:

$$\frac{n_C \times r_1}{n} - 1 = \frac{4 \times 4.7}{5} - 1 = 2.76 \tag{10}$$

When Individual 6 participates in a public goods game centered on Individual 5, if there are five cooperative individuals in the group, its payoff can reach:

$$\frac{n_C \times r_1}{n} - 1 = \frac{5 \times 4.7}{5} - 1 = 3.7 \tag{11}$$

In the five public goods game interactions within the spatial structure (one centered on itself and four centered on its neighbors), Individual 6 participates in a public goods game centered on Individual 5 only once, which leads to an increase in its Q_c. In the other three interactions, centered on Individuals 2, 10, and itself, Q_c increases if it is below 2.76 at the time; however, if Q_c exceeds 2.76, it decreases. Participation in a public goods game centered on Individual 7 always results in a decrease in Q_c. As a result, the upper limit of Q_c stabilizes around 2.76. Consequently, as shown in the Fig. 6(b), Individual 6, located at the boundary, exhibits noticeable oscillations in Q_c around 2.76.

Regarding Q_d, when Individual 6 selects defection and there are three cooperative individuals in the group, the immediate payoff is calculated as follows:

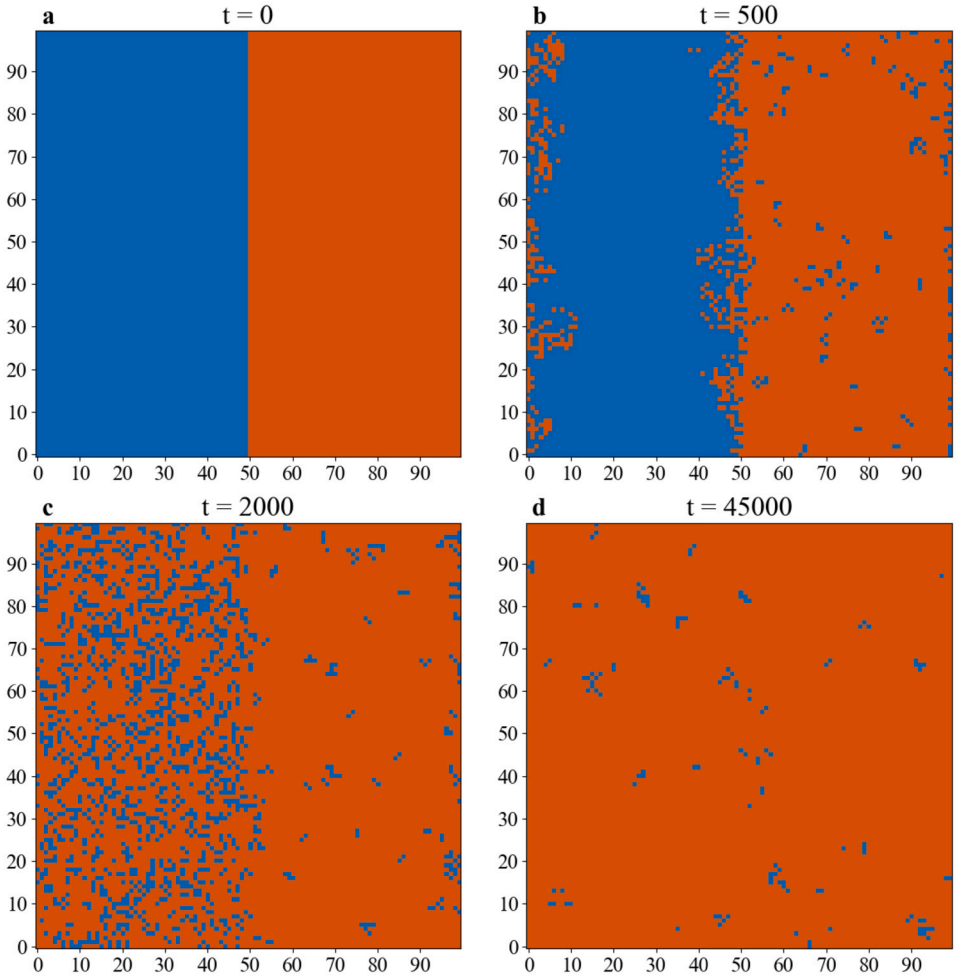

Fig. 5. Spatial evolution of cooperation and defection over time in TNP. The blue regions represent cooperators, and the red regions represent defectors. The four panels correspond to different time steps: **(a)** $t = 0$, **(b)** $t = 500$, **(c)** $t = 2000$, and **(d)** $t = 45000$. The cooperative region (blue) gradually shrinks over time, while defection (red) dominates the system. The results shown were obtained with the following values: $r_1 = 4.7$, $r_2 = 4.5$, $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

$$\frac{n_C \times r_1}{n} = \frac{3 \times 4.7}{5} = 2.82. \tag{12}$$

When Individual 6 participates in the strategy update centered on Individual 5, where there are four cooperative individuals in the group, the defection payoff is:

$$\frac{n_C \times r_1}{n} = \frac{4 \times 4.7}{5} = 3.76. \tag{13}$$

Thus, the upper limit of Q_d for Individual 6 lies between 2.82 and 3.76. For boundary individuals, there will always be a moment when Q_d exceeds Q_c, prompting a transition into a red (defector) individual.

In Phase 3, as the number of cooperative individuals decreases, both Q_c and Q_d decline. However, in a population dominated by defectors, Q_c may drop below zero, while Q_d consistently remains positive, leading to Q_c being lower than Q_d in most cases during the later stages of the evolutionary process. Consequently, in the final phase of the system, the advantage of Q_d leads to red individuals gradually becoming the prevailing population.

Another observed finding arises due to the widespread presence of defectors, which causes Q_d of red individuals to gradually decrease. Meanwhile, the introduction of the exploration rate ensures that there is always a certain probability of random strategy selection, such as when individual 6 and individual 7 randomly both adopt cooperative behaviors. As shown in the Fig. 7(a), when individual 6 participates in a public goods game centered around individual 7 (or individual 7 participates in a public goods game centered around individual 6), both individuals 6 and 7 receive positive payoffs, which promotes an increase in Q_c. In the population dominated by defectors, Q_d approaches 0, causing individual 6 and individual 7 to become cooperators, forming a scattered distribution of blue individuals in a red-dominated system, as shown in the Fig. 7(b). However, in subsequent strategy interactions, when individual 6 participates in a public goods game with defectors (such as individual 5) as the center, individual 6's cooperative

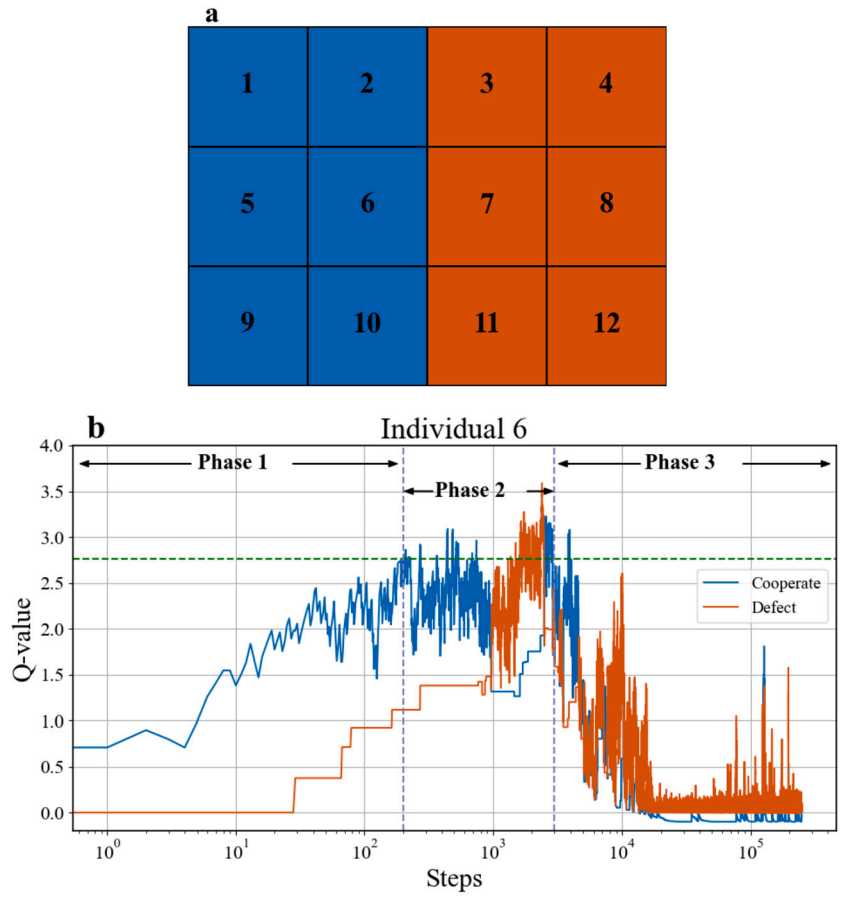

Fig. 6. Analysis of the boundary region and strategy dynamics in TNP. (a) The local spatial boundary configuration between cooperators (blue) and defectors (red). (b) The Q-value dynamics for individual 6 at the boundary. To more intuitively depict the update process of Q-values, we record the data immediately after each Q-value update. Therefore, we use "Steps" as the x-axis label. The results shown were obtained with the following values: $r_1 = 4.7$, $r_2 = 4.5$, $c = 1$, $\alpha = 0.1$, $\gamma = 0.9$, $\epsilon = 0.01$, and $L = 100$.

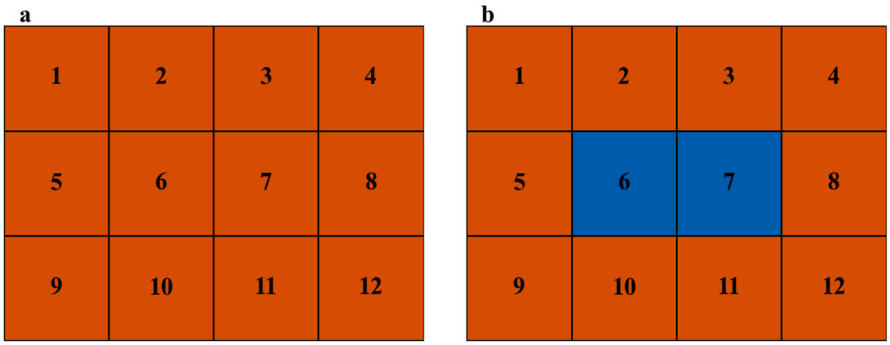

Fig. 7. Analysis of strategy finding 2. (a) The local spatial configuration with defectors (red) before the transition, where all individuals are defectors. (b) The evolution after the transition, where individuals 6 and 7 become cooperators (blue), forming scattered blue clusters in the red group.

behavior is exploited by other defectors, leading to a decrease in Q_c, causing them to revert to red individuals. Consequently, the system exhibits a cycle where small clusters of blue individuals appear and disappear.

By observing the behavioral changes of individual agents in the above two-dimensional grid, we can conclude that as the number of defectors in the population gradually increases, the overall cooperation rate continuously declines, ultimately leading to the majority of individuals in the system entering the less valuable game. This state effectively transforms the system into a scenario resembling the "environment-static" case, where individuals only make strategy choices in the less valuable game. This result provides a valid explanation for Phenomenon 1, where in TNP, the cooperation threshold is later than in SE.

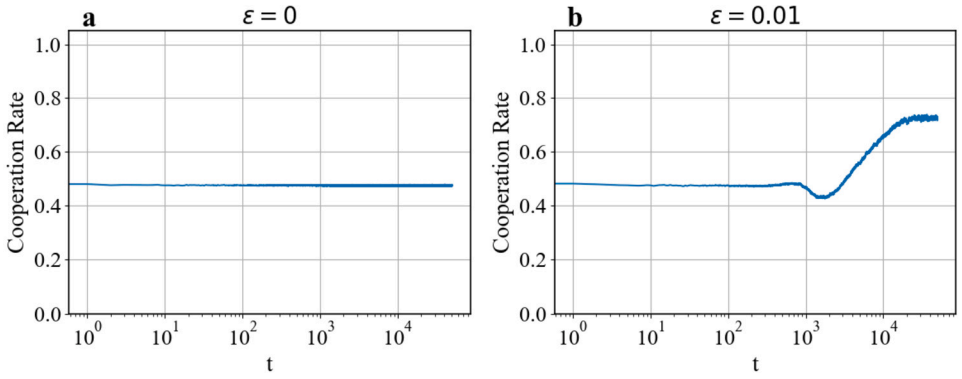

Fig. 8. Comparison of cooperation rate under different ϵ. The left panel shows the cooperation rate with $\epsilon = 0$, while the right panel corresponds to $\epsilon = 0.01$. The cooperation rate remains constant when $\epsilon = 0$ but increases significantly over time when $\epsilon = 0.01$. The results shown were obtained with the following values: $r_1 = 4.2$, $r_2 = 4.0$, $c = 1$, $\alpha = 0.1$, $\gamma = 0.9$, and $L = 100$.

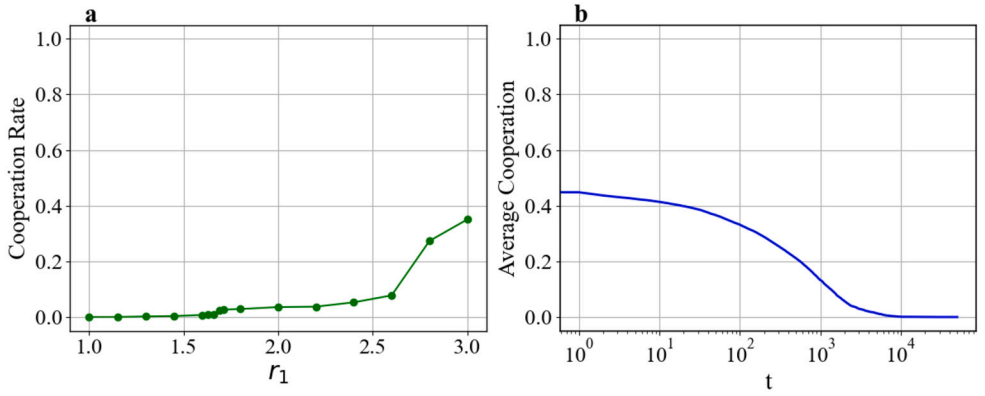

Fig. 9. Cooperation rate versus synergy factor r_1 for TP and time steps for SE ($\epsilon = 0$). (a) The cooperation rate in TP as the r_1 value changes. The cooperation emergence threshold is around $r_c = 1.67$. (b) The cooperation rate in SE at $r = 3$ as a function of time steps. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, and $L = 100$.

3.2. Explanation of Phenomenon 2

From Phenomenon 2, we are pleased to observe that cooperation is more likely to emerge when environmental transitioning occurs and individuals are able to perceive environmental transitions. Therefore, we focus on investigating the mechanisms behind this phenomenon. In the public goods game, defection is the dominant strategy [51]. However, the introduction of the exploration rate (ϵ) [52] triggers cooperative behavior in local spaces, leading to the formation of reciprocal clusters that challenge the dominance of defection. To explore the effect of ϵ on strategy evolution, we present the evolution of cooperation rates under different ϵ settings. As shown in the Fig. 8, it is surprised to find that even in the absence of exploration ($\epsilon = 0$) (Fig. 8(a)), the cooperation rate, although lower than when the exploration rate is 0.01, still stabilizes around 0.5. This result suggests that while ϵ has a significant effect on the system, it is not the fundamental cause of cooperation.

On the other hand, to verify whether Phenomenon 2 still holds, the above experiment is repeated in the absence of randomness, with $\epsilon = 0$. As shown in the Fig. 9(a), after adjusting ϵ to 0, in TP, the cooperation threshold is approximately 1.67. In contrast, for SE, during the threshold experiment, we find that at higher r values, the system requires a longer period of interaction to converge. Therefore, to save experimental time, $r = 3$ is chosen as an example to demonstrate the evolution of the cooperation rate in the SE case, as shown in the Fig. 9(b), where the cooperation rate ultimately converges to 0. It can be reasonably inferred that when $r < 3$, cooperation cannot emerge, meaning the threshold is more than 3. In conclusion, when $\epsilon = 0$, the threshold for TP is lower than the threshold for SE, phenomenon 2 still holds. This further validates that the exploration rate is not the fundamental cause of cooperation emergence under environmental transitioning.

To explore the underlying causes of cooperation emergence, we further analyze the evolution of individual behavior at the micro level. Specifically, by examining the environmental states (more valuable game, less valuable game) and the corresponding actions (cooperation, defection), the aim is to reveal how individuals adapt and adjust their behavior patterns in dynamic environments. For generality, six individuals from positions (7,7) to (7,12) on the grid are selected to demonstrate their respective behavior patterns and dynamic changes. To enhance clarity and intuition, we visualize the strategy choices of each individual for the first 1000 steps.

The results are shown in the Fig. 10, where the vertical coordinates represent the individual's behavior patterns, with 0, 1, 2, and 3 indicating "cooperation in the more valuable game", "defection in the more valuable game", "cooperation in the less valuable game",

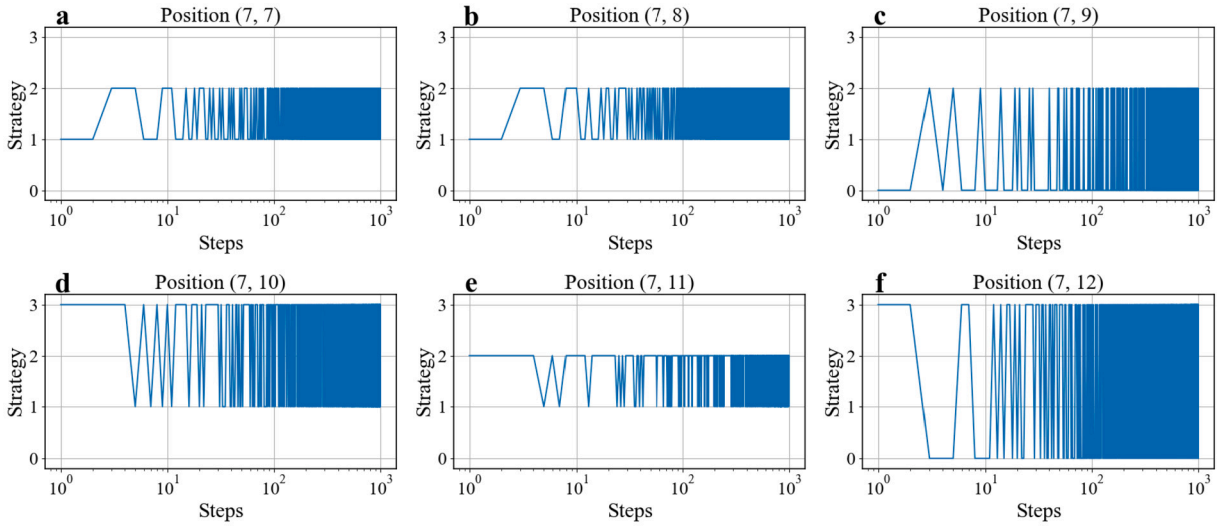

Fig. 10. The strategy evolution from positions (7,7) to (7,12). The vertical axis represents four strategies: 0 (cooperate in the more valuable game), 1 (defect in the more valuable game), 2 (cooperate in the less valuable game), and 3 (defect in the less valuable game). Each panel corresponds to a specific position, illustrating the strategy changes over time. To more intuitively depict each individual's strategy selection, we record the data immediately after each individual makes a strategy update. Therefore, we use "Steps" as the x-axis label. The results shown were obtained with the following values: $r_1 = 3.0$, $r_2 = 2.8$, $c = 1$, $\gamma = 0.9$, $\alpha = 0.1$, $\epsilon = 0$, and $L = 100$.

Table 1
Average percentage of four strategies. The results shown were obtained with the following values: $r_1 = 3.0$, $r_2 = 2.8$, $c = 1$, $\gamma = 0.9$, $\alpha = 0.1$, $\epsilon = 0$, and $L = 100$.

Strategy	Description	Percentage (%)
Strategy 1	Cooperate in both more valuable and less valuable games	19.45
Strategy 2	Cooperate in the more valuable game but defect in the less valuable game	28.24
Strategy 3	Defect in the more valuable game but cooperate in the less valuable game	23.59
Strategy 4	Defect in both more valuable and less valuable games	28.72

and "defection in the less valuable game", respectively. In the Fig. 10(a), the individual at position (7,7) adopts behavior pattern 1 during the first two steps, which represents being in the more valuable game and choosing defection when facing its neighbors. Subsequently, this defection leads to environmental degradation, and the system transitions to the less valuable game. After entering the less valuable game, the individual adopts behavior pattern 2 and chooses cooperation, which restores the environment to the more valuable game by the 6th step. Therefore, the individual's behavior pattern can ultimately be characterized as "defection in the more valuable game + cooperation in the less valuable game".

Similarly, individuals at position (7,8) (Fig. 10(b)) and position (7,11) (Fig. 10(e)) adopt the pattern "defection in the more valuable game + cooperation in the less valuable game", individuals at position (7,9) (Fig. 10(c)) follow the pattern "cooperation in both games", individuals at position (7,10) (Fig. 10(d)) adopt the pattern "defection in both games", and individuals at position (7,12) (Fig. 10(f)) follow the pattern "cooperation in the more valuable game + defection in the less valuable game". Based on the above behavior patterns, four strategies present in the system can be summarized: Strategy 1: "cooperation in the more valuable game + cooperation in the less valuable game", Strategy 2: "cooperation in the more valuable game + defection in the less valuable game", Strategy 3: "defection in the more valuable game + cooperation in the less valuable game", Strategy 4: "defection in the more valuable game + defection in the less valuable game".

It is worth noting that the Fig. 10 only shows the first 1000 steps of the evolutionary process, with subsequent strategies remaining consistent. For brevity, the later steps are not displayed. This stability in strategy can be attributed to the adaptive decision-making of individuals under deterministic conditions ($\epsilon = 0$): when an individual adopts a strategy in a particular environment that yields non-negative payoffs, they tend to maintain that strategy, leading to the formation of a stable behavior pattern.

To gain a comprehensive understanding of the system's strategy distribution, we perform a statistical analysis of the strategies of individuals in the stable state of the system, as shown in the Table 1. It can be observed that the proportion of each strategy is approximately 20%, with strategy 1 having the lowest proportion and strategy 4 having the highest proportion.

Considering that in public goods games, individual behavior is influenced by neighbors, to explore why the four strategies can coexist in the system, it is necessary to further analyze the behavior patterns under spatial distribution. For generality, the four strategies are randomly distributed in space, and the Q-values are initialized according to the characteristics of each strategy. For example, for individuals adopting strategy 1, the cooperation Q_c values in both the more valuable and less valuable games are initialized to 0.5, while the defection Q_d values are set to 0. As shown in Fig. 11, the spatial distribution and evolutionary trajectories

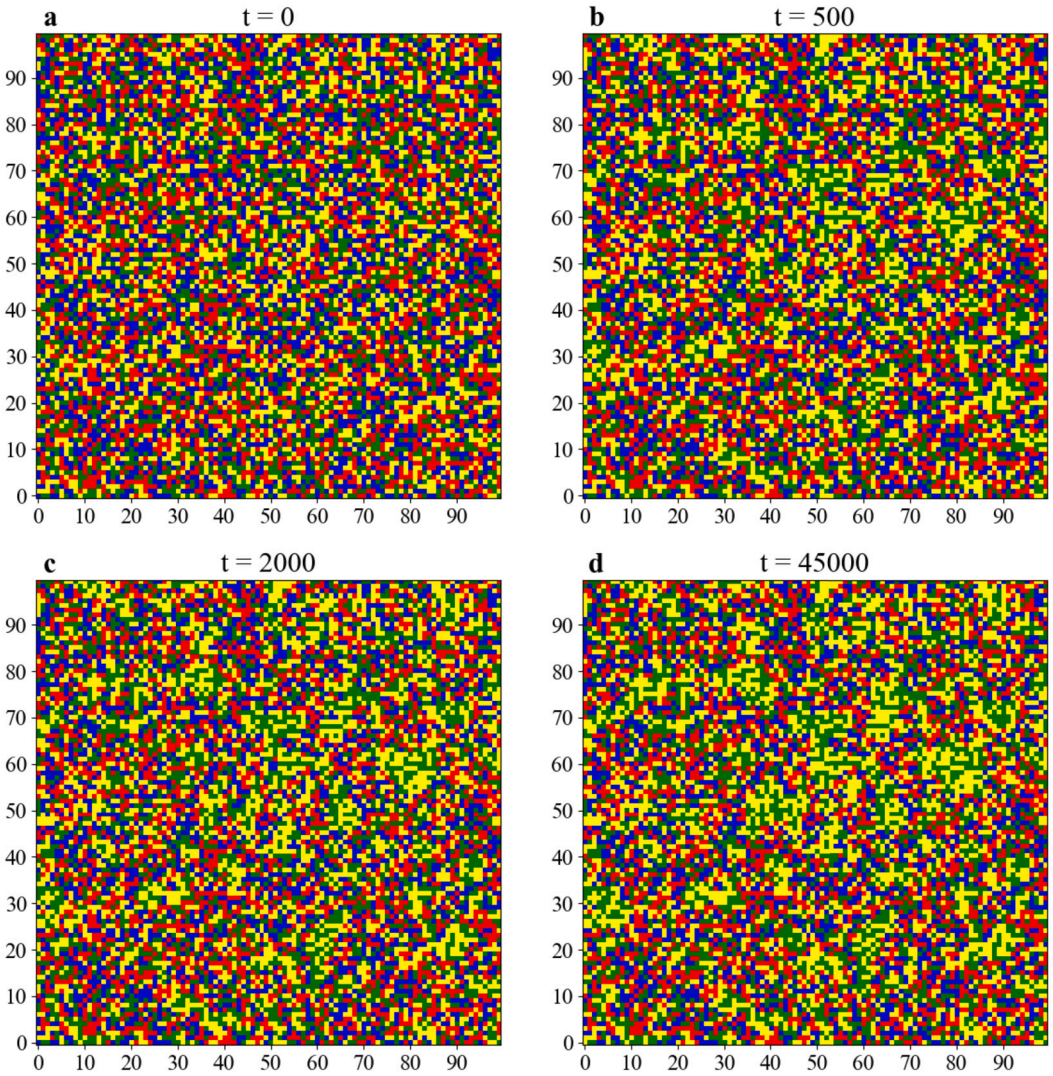

Fig. 11. Spatial distribution of four strategies over time. The colors represent the following strategies: **Blue** for Strategy 1 (cooperate in both the more valuable and less valuable games), **Yellow** for Strategy 2 (cooperate in the more valuable game but defect in the less valuable game), **Red** for Strategy 3 (defect in the more valuable game but cooperate in the less valuable game), and **Green** for Strategy 4 (defect in both the more valuable and less valuable games). The four panels correspond to different time steps: **(a)** $t = 0$, **(b)** $t = 500$, **(c)** $t = 2000$, and **(d)** $t = 45000$. The results shown were obtained with the following values: $r_1 = 3.0$, $r_2 = 2.8$, $c = 1$, $\gamma = 0.9$, $\alpha = 0.1$, $\epsilon = 0$, and $L = 100$.

of these four strategies are displayed, where blue, yellow, red, and green represent individuals adopting strategy 1, strategy 2, strategy 3, and strategy 4, respectively.

As observed in the Fig. 11, the spatial distribution of the four strategies at different time steps (0, 500, 2000, 45000) shows almost no significant changes. This indicates that, despite the system evolving through multiple time steps, the distribution of strategies adopted by individuals in space remains relatively stable. Since it is difficult to directly observe the evolution of strategies by eye, further statistical analysis of the dynamic transitions between strategies is necessary, with the results shown in the Fig. 12, which are the average data obtained from five independent experiments.

The results show that the evolutionary changes of various strategies are mainly concentrated in the initial stages, with transitions occurring primarily from blue individuals (strategy 1) to yellow individuals (strategy 2) and from red individuals (strategy 3) to green individuals (strategy 4). The number of changes remains within single digits. As the time steps increase, the number of transitions between strategies significantly decreases. This indicates that, in the later stages of the system's evolution, the strategies gradually stabilize, with strategy transitioning between individuals effectively ceasing, and the system gradually reaching a stable state. This also validates the finding of spatial distribution stability in the Fig. 11.

To understand the reasons behind the occurrence of only two types of transitions, from blue individuals to red individuals and from red individuals to green individuals, we begin by analyzing the microscopic mechanisms of strategy transitions based on spatial contact modes. Considering that the contact modes between different strategy populations in the spatial structure can be categorized into

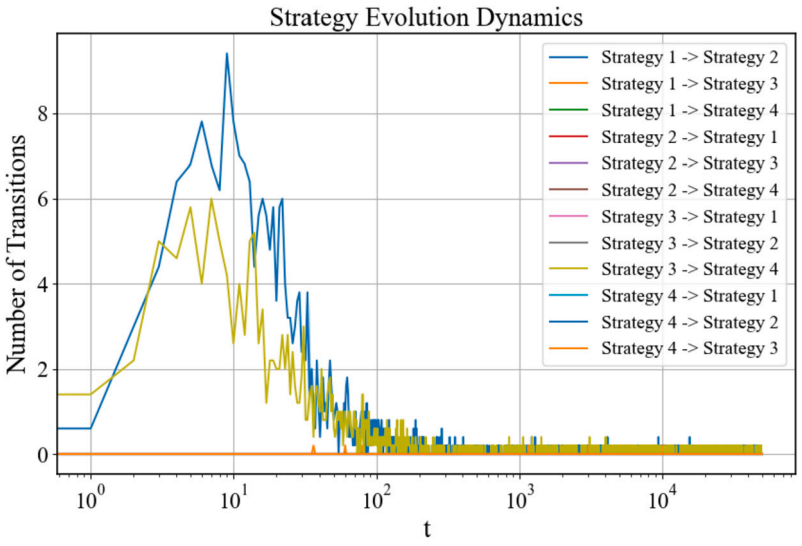

Fig. 12. Number of transitions between strategies at each time step in a perceiving environment. The results shown were obtained with the following values: $r_1 = 3.0$, $r_2 = 2.8$, $c = 1$, $\gamma = 0.9$, $\alpha = 0.1$, $\epsilon = 0$, and $L = 100$.

three types: cluster-to-cluster contact, individual-to-cluster contact, and individual-to-individual contact. In the following analysis, we will examine the co-evolutionary relationships between strategies based on these three contact modes.

Cluster-to-cluster contact. For ease of analysis, the four strategies are uniformly distributed on the spatial grid, as shown in the Fig. 13. By time step 50, the yellow region gradually erodes part of the boundary of the blue region, while a yellow isolation zone also appears at the interface between the blue and green regions. However, in the subsequent evolutionary steps, the strategy distribution stabilizes, with the amplitude of changes significantly reduces, demonstrating the high stability of the system. This result further validates the finding of dynamic stability of strategies in the spatial distribution, as shown in the Fig. 11.

To clearly understand the interaction mechanisms and dynamic changes between strategies, we focus on the boundary regions between blue and green, as well as between blue and yellow, with the boundaries shown in the Fig. 14 and Fig. 15. The Fig. 14(a) shows the initial stage of the blue and green boundary region, while the Fig. 14(b) shows the situation after the formation of the yellow isolation zone in the stable state. The Fig. 15(a) shows the initial stage of the blue and yellow boundary region, and the Fig. 15(b) illustrates the situation after the formation of a yellow serrated boundary in the stable state.

The boundary between the blue and green regions. In the Fig. 14(a), when individual 6 participates in the strategy update centered around individual 7, it chooses cooperation. However, individual 7 and its three neighbors choose defection, which gradually lowers Q_c in the less valuable game of individual 6. Once this value becomes negative, individual 6 will adopt defection when faced with the less valuable game, thus transitioning from blue to yellow. Similarly, individuals 2 and 10 also turn yellow, resulting in the formation of a yellow isolation zone at the boundary between the blue and green regions, as shown in the Fig. 14(b). So why does the yellow isolation zone not further invade the interior of the blue region? In the Fig. 14(b), although individual 6 has turned yellow, the green individuals can only alter individual 6's actions in the less valuable game, making it choose defection, but they cannot influence individual 6's environmental state. That is, when individual 6 updates its strategy as the central individual, individuals 2, 5, and 10 all cooperate, which ensures that individual 6 remains in the more valuable game. Therefore, individual 5 never interacts with the less valuable game, and the interior of the blue region remains stable.

The boundary between the blue and yellow regions. In the Fig. 15(a), due to initialization randomness, individual 7, as the central individual, may start in either the more valuable game or the less valuable game. If initialized in the more valuable game, individual 7 and its four neighbors will all choose cooperation, and when individual 7 updates its strategy as the central individual in the next step, it remains in the more valuable game. As a result, individual 6 will not be exploited and will maintain its cooperative behavior in the more valuable game, thus not transitioning to yellow. If individual 7 starts in the less valuable game, only individual 6 chooses cooperation, which results in individual 6's payoff becoming negative, leading to the transition to defection in the less valuable game, ultimately turning yellow. The uncertainty of the initial environment of individual 7 causes individual 6 to be invaded by individual 7, resulting in the two types shown in the Fig. 15(b) and (c). Similar to the mechanism of the yellow isolation zone, individual 7 can only influence individual 6's strategy in the less valuable game. However, when individual 6 is the central individual, its environment remains in the more valuable game. This prevents the interior individuals of the blue region from interacting with the less valuable game, thus preventing further erosion. Consequently, the boundary of the blue region is eroded by the yellow region into a serrated shape, as shown in the Fig. 13.

The boundary between the blue and red regions. For boundary individual in the red region, if its initial state is in the more valuable game, it will choose defection, with only one of its four neighbors choosing cooperation. This not only causes it to transition to the less valuable game in the next step but also keeps Q_d in the more valuable game positive. After entering the less valuable game, since both itself and its four neighbors all choose cooperation, it will return to the more valuable game, which also increases Q_c in

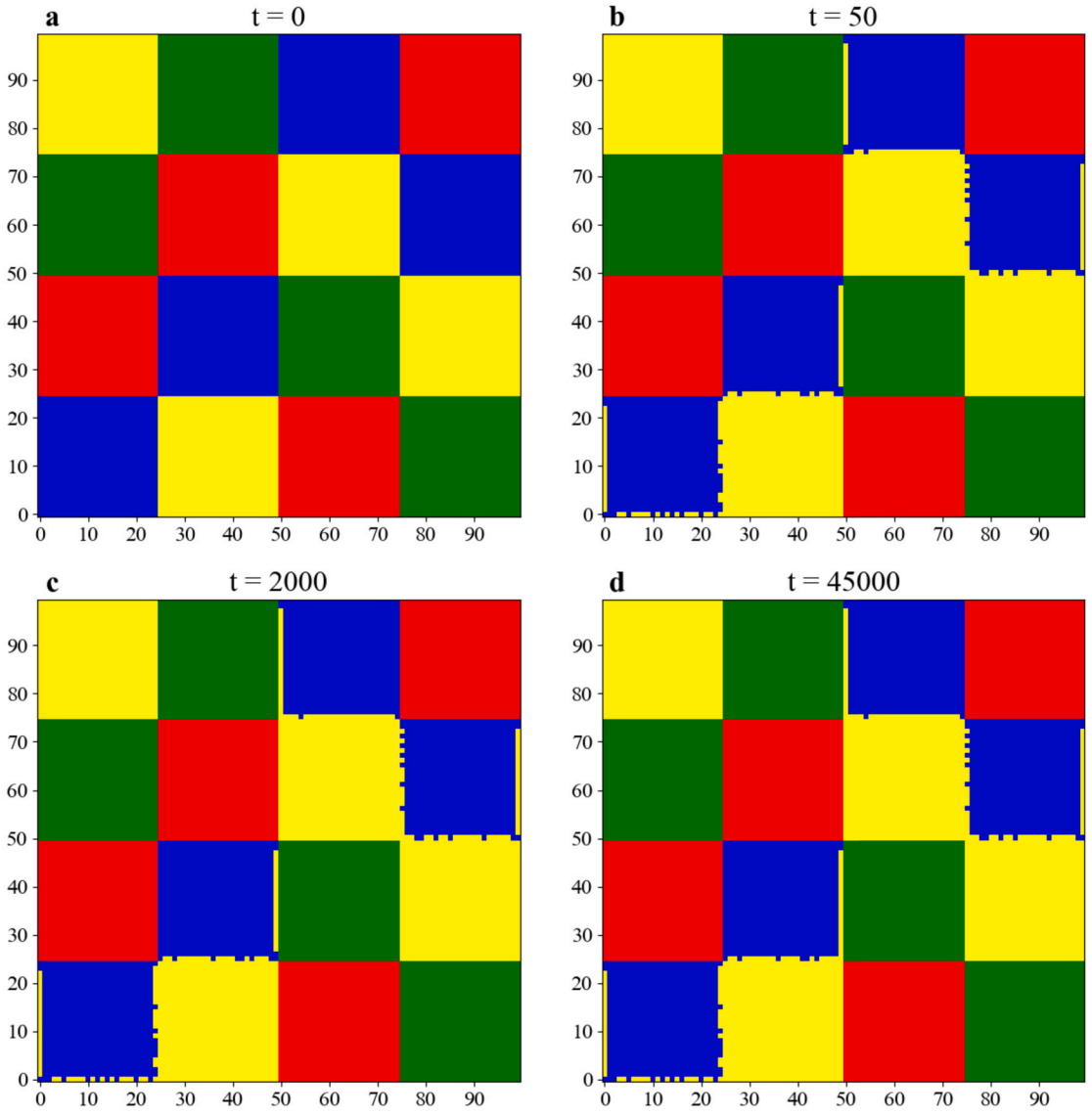

Fig. 13. The cluster-to-cluster contact mode. The four strategies are represented as follows: **Blue** for Strategy 1 (cooperate in both the more valuable and less valuable games), **Yellow** for Strategy 2 (cooperate in the more valuable game but defect in the less valuable game), **Red** for Strategy 3 (defect in the more valuable game but cooperate in the less valuable game), and **Green** for Strategy 4 (defect in both the more valuable and less valuable games). The panels correspond to different time steps: **(a)** $t = 0$, **(b)** $t = 50$, **(c)** $t = 2000$, and **(d)** $t = 45000$. This analysis focuses on the interactions at the boundaries where clusters of different strategies come into contact, emphasizing the evolution of cluster dynamics over time. The results shown were obtained with the following values: $r_1 = 3.0$, $r_2 = 2.8$, $c = 1$, $\gamma = 0.9$, $\alpha = 0.1$, $\epsilon = 0$, and $L = 100$.

the less valuable game. Through this cycle between the two environments, Q_d in the more valuable game and Q_c in the less valuable game of the red individual remain positive, preventing a strategy change. Similarly, for boundary individuals in the blue region, Q_c in the more valuable game and less valuable game also maintain a relative advantage, ensuring that their strategy remains unchanged. In conclusion, when the blue and red regions meet at the boundary, the strategy remains stable, which validates the findings in the Fig. 13.

The analysis of the boundaries between the **yellow and red regions**, the **yellow and green regions**, and the **red and green regions** is similar to that of the boundary between the **blue and red regions**, with boundary stability being maintained in all cases.

In conclusion, through the analysis of cluster-to-cluster contact, we observe that the transition process from blue individuals to yellow individuals is more pronounced.

Individual-to-cluster contact. Without loss of generality, in order to study the evolutionary behavior of isolated individuals in other strategy populations, we place them in populations with different strategies and observe the following three transitions:

• **The red individual transitions into the green individual.** As shown in the Fig. 16(a), in the environment where green individuals form clusters, it can be observed that, due to the defection strategy chosen by surrounding individuals, the cooperative

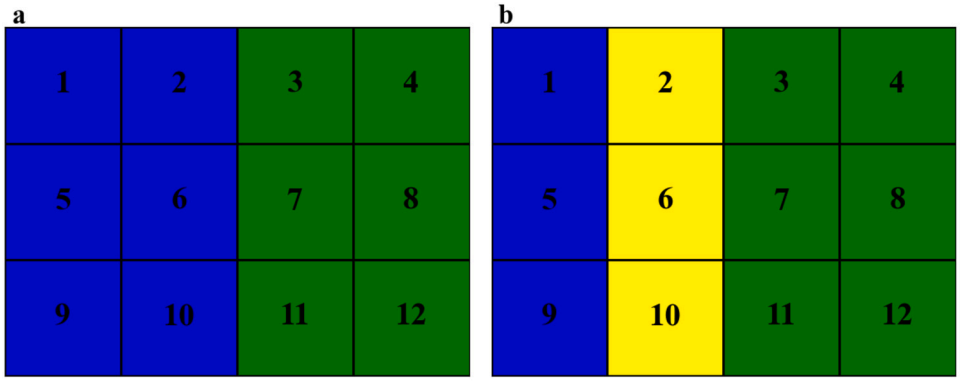

Fig. 14. Boundary dynamics of strategy interactions of cluster-to-cluster contact. The two subfigures represent the evolution of interactions at the boundaries between blue and green regions: **(a)** shows the initial stage of the boundary between blue individuals (Strategy 1: cooperation in both the more valuable and less valuable games) and green individuals (Strategy 4: defection in both the more valuable and less valuable games). **(b)** shows the situation after the formation of the yellow isolation zone in the stable state, where yellow individuals (Strategy 2: cooperation in the more valuable game and defection in the less valuable game) invade the boundary between the blue and green regions, converting some blue individuals into yellow.

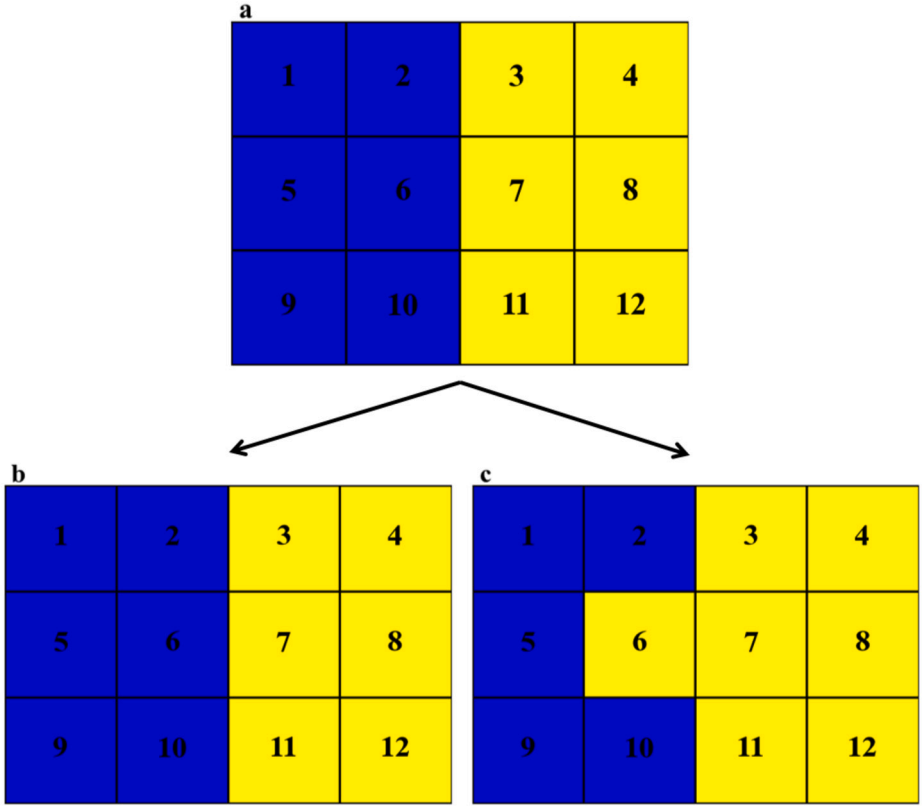

Fig. 15. Boundary dynamics of strategy interactions of cluster-to-cluster contact. The three subfigures represent the evolution of interactions at the boundaries between blue and yellow regions: **(a)** shows the initial stage of the boundary between blue individuals and yellow individual. **(b)** shows the erosion when individual 7 starts in the more valuable game, leading to the formation of a yellow serrated boundary. **(c)** shows the erosion when individual 7 starts in the less valuable game, leading to a similar boundary formation.

behavior of the red individual in the less valuable game continuously results in negative payoffs, causing Q_c in the less valuable game to decrease steadily, eventually falling below Q_d. This environmental pressure drives the red individual to gradually transition into the green individual, adopting the strategy of defection in both environments. Similarly, the red individual is found to transition into green in the yellow cluster (Fig. 16(k)).

• **The blue individual transitions into the yellow individual.** In the green cluster shown in the Fig. 16(b), due to the fact that all neighbors of the blue individual choose defection, its Q_c in the less valuable game will continuously decrease. At the same time, due to the lack of cooperative neighbors, the more valuable game is nearly impossible to occur, which results in no updates to

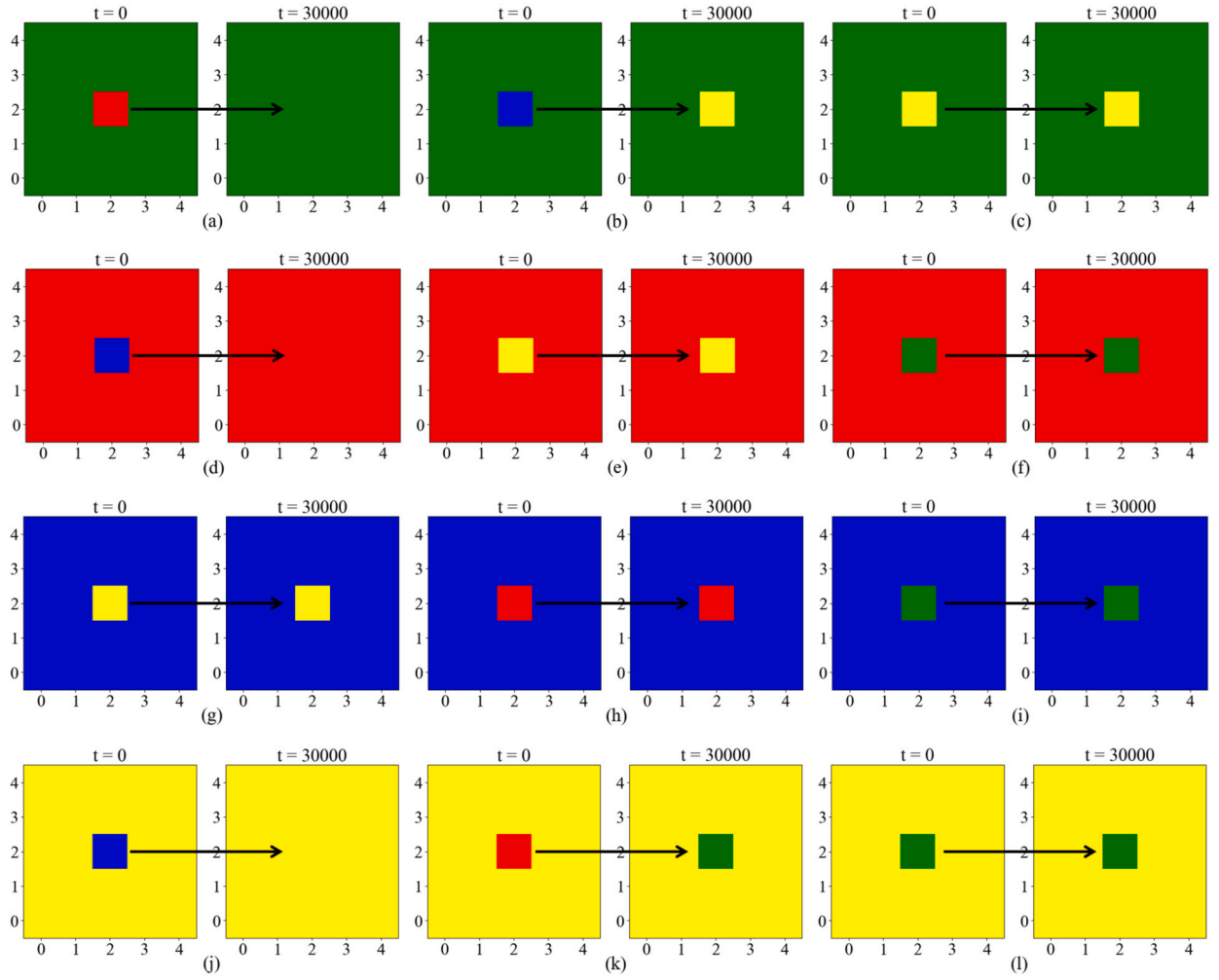

Fig. 16. The individual-to-cluster contact mode. (a-c) show the green cluster, (d-f) show the red cluster, (g-i) show the blue cluster, and (j-l) show the yellow cluster. The results shown were obtained with the following values: $r_1 = 3.0$, $r_2 = 2.8$, $c = 1$, $\gamma = 0.9$, $\alpha = 0.1$, $\epsilon = 0$, and $L = 5$.

the Q-value in the more valuable game. Ultimately, the blue individual will transition into a yellow individual. Similarly, the blue individual is found to transition into yellow in the yellow cluster (Fig. 16(j)).

• **The blue individual transitions into the red individual.** As shown in the Fig. 16(d), in the environment where red individuals form clusters, it can be observed that in the more valuable game, since only the blue individual chooses cooperation, Q_c in the more valuable game of red individuals decreases to a negative value, ultimately leading the blue individual to transition into a red individual.

Other isolated individuals remain unchanged in different strategy populations. Without loss of generality, consider the case of the yellow individual in the green cluster shown in the Fig. 16(c). Due to the defection behavior of the green individuals, the yellow individual is almost always in the less valuable game, which prevents Q_c in the more valuable game from being updated, thereby maintaining its strategy.

Based on the aforementioned experiments, it is observed that the blue individual transitions into the yellow and red individual, while the red individual transitions into the green individual, thereby validating all the findings presented in the Fig. 12.

Individual-to-individual contact. For ease of analysis, we uniformly distribute each individual in a 4×4 grid to ensure that every pair of individual types is in contact. As shown in the Fig. 17, individuals of each type remain stable, with no strategy changes. This stability mainly arises from the fact that, regardless of the environment state of each individual, there are at least two cooperative individuals in the group involved in the public goods game centered around the individual, which helps maintain the individual's strategy unchanged. Without loss of generality, let's take the blue individual located at (1,1) as an example: if in a more valuable game, there is a blue individual (self) and a yellow individual located at (1,0) cooperating within the group centered around it in the public goods game; if in a less valuable game, the group consists of a blue individual (self) and two red individuals located at (0,1) and (1,2) cooperating. This ensures that Q_c of the blue individual remains positive in both the more valuable and less valuable games, meaning that the blue individual's strategy does not change during strategy interactions.

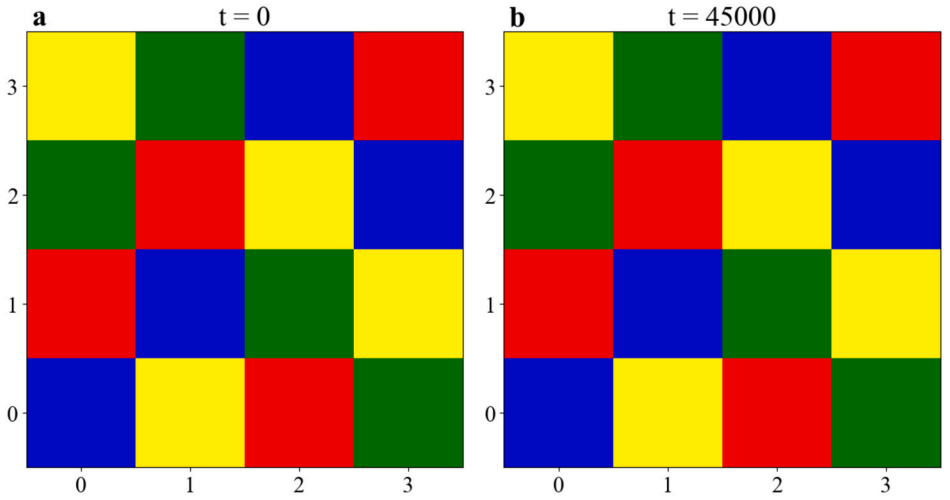

Fig. 17. The individual-to-individual contact mode. The figure illustrates the spatial distribution at two time steps: **(a)** $t = 0$ and **(b)** $t = 45000$, where individuals are distributed uniformly. The results shown were obtained with the following values: $r_1 = 3.0$, $r_2 = 2.8$, $c = 1$, $\gamma = 0.9$, $\alpha = 0.1$, $\epsilon = 0$, and $L = 4$.

Through the analysis of three different contact modes, we discover that the maintenance of cooperation is collectively facilitated via different mechanisms. In the case of cluster-to-cluster contact, cooperative clusters effectively resist the erosion of defection strategies due to their high internal payoffs and stable environmental states. Specifically, the yellow region can only invade the blue individuals at the blue-green boundaries, and has little impact on the internal structure of the cooperative clusters. In the case of cluster-to-individual contact, under certain conditions, the blue individual transitions into the yellow or red individual, and the red individual transitions into the green individual. In the case of individual-to-individual contact, since each individual ensures that there are at least two cooperative neighbors in the public goods game, this uniform spatial distribution provides the most stable foundation for maintaining cooperation. In conclusion, the strategy transitions observed in the three contact modes confirm the results shown in the Fig. 12. Additionally, the analysis of the yellow isolation zone also explains the intrinsic mechanism behind the stability of cooperation, further elaborating on the reasons why cooperation can emerge.

3.3. Extended experiments

To verify the positive effect of information perception on the emergence of cooperation, we conduct extended experiments in six aspects: (1) synchronous update interactions; (2) different r-value differences between the more valuable game and the less valuable game; (3) a more stringent environmental transition mechanism; (4) different Q-learning state paradigms; (5) a larger von Neumann neighborhood range; and (6) different network topologies.

(1) Synchronous update interactions.

Given that the update mechanism plays a crucial role in shaping the dynamics of individual strategies, we explore its impact by examining the synchronous update mechanism in the extended experiments.

In each round t, individuals select the action with the highest Q-value according to their current states, based on the information in their Q-tables. After all individuals take their actions, they obtain the payoff π_i and update the corresponding Q-values simultaneously [53]. As shown in the Fig. 18, the threshold for SE is 4.18, the threshold for TNP is 4.38, and the threshold for TP is 3.14, which is lower than the first two systems, thereby validating the positive effect of information perception on the emergence of cooperation. Furthermore, when compared with the results from asynchronous updating in this study, the cooperation threshold with synchronous updating is smaller, indicating that synchronous updating enables all individuals to adjust their strategies simultaneously based on the latest environmental information. This helps each individual in the system recognize the greater benefits of cooperation at the same time, facilitating the propagation and stabilization of cooperative behavior.

(2) Impact of the difference in r values between the more valuable game and the less valuable game.

Since the magnitude of the r value difference between the more valuable game and the less valuable game directly affects an individual's decision-making process, we further investigate the impact of this difference on cooperative behavior.

• **The difference of 0.5.** Increasing the difference in r between the more valuable game and the less valuable game to 0.5. As shown in the Fig. 19, the threshold for SE is 4.50, the threshold for TNP is 5.00, and the threshold for TP is 4.38, which is lower than the first two systems, demonstrating the crucial role of information perception in promoting the emergence of cooperation. When compared with the results from a difference of 0.2 in this study, it is observed that the thresholds for TNP and TP are delayed. This is because for the same r_1, a larger r difference corresponds to a smaller r_2, thus increasing the difficulty of cooperation emergence.

• **The difference of 0.1.** Decreasing the difference in r between the more valuable game and the less valuable game to 0.1. As shown in the Fig. 20, the threshold for SE is 4.50, the threshold for TNP is 4.60, and the threshold for TP is 4.00, which is lower than the first two systems, confirming the significant impact of environmental information perception on the evolution of cooperation.

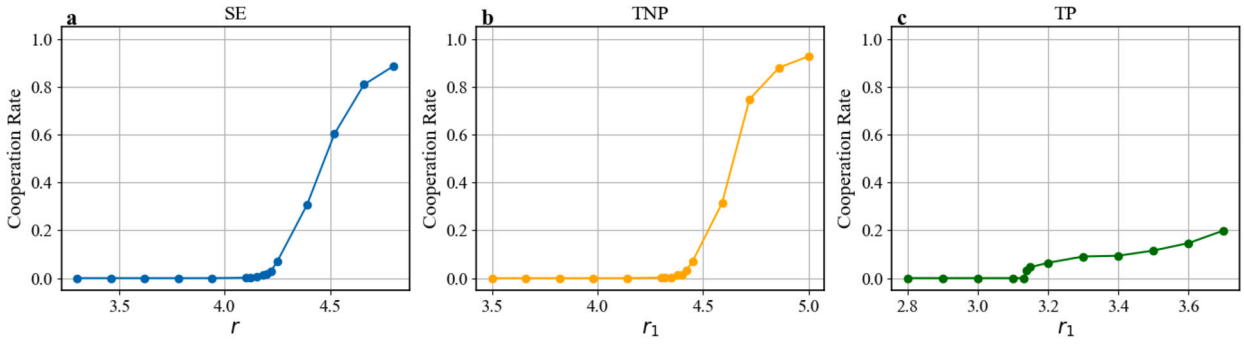

Fig. 18. Cooperation rate at convergence versus the synergy factor r in three systems with synchronous updates. (a) **SE:** The cooperation emergence threshold is approximately $r_c = 4.18$. (b) **TNP:** The cooperation emergence threshold is approximately $r_c = 4.38$. (c) **TP:** The cooperation emergence threshold is approximately $r_c = 3.14$. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

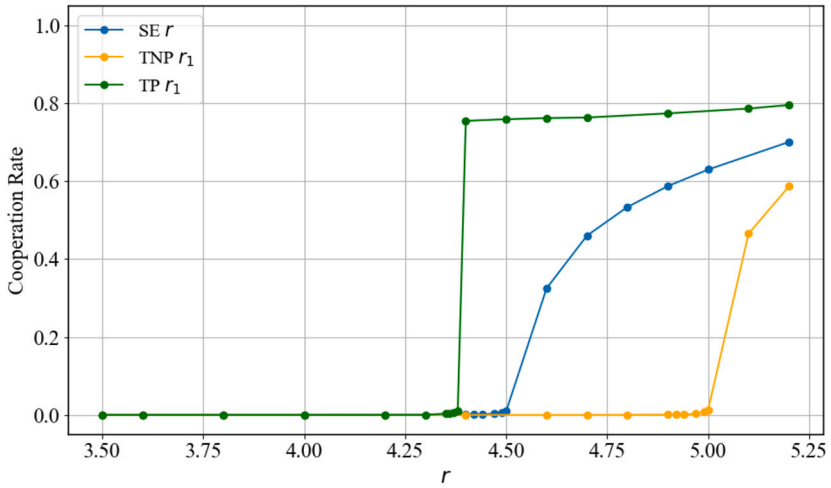

Fig. 19. Cooperation rate at convergence versus the synergy factor r in three systems with a difference in r of 0.5. The cooperation emergence threshold is approximately $r_c = 4.50$ for SE, $r_c = 5.00$ for TNP, and $r_c = 4.38$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

When compared with the results from a difference of 0.2 in this study, it is observed that the thresholds for TNP and TP are advanced. This is because for the same r_1, a smaller r difference corresponds to a larger r_2, thus lowering the difficulty of cooperation emergence.

(3) A more stringent environmental transition mechanism.

Since the experiments in this study adopt a mild transition mechanism, it is necessary to further investigate the impact of a more stringent transition mechanism on individual strategies and cooperative behavior.

Modifying the transition mechanism between the more valuable game and the less valuable game as shown in the Fig. 21: when an individual has four cooperative neighbors, the system enters the more valuable game in the next round; when the number of cooperative neighbors is fewer than four, the system will be the less valuable game in the next round [34]. As shown in the Fig. 22, the threshold for SE is 4.50, the threshold for TNP is 4.70, and the threshold for TP is 4.11, which is lower than the first two systems, thereby validating the positive effect of information perception on the emergence of cooperation. Furthermore, when compared with the original transitioning mechanism in this study, the results show little difference, indicating that the strictness of the transitioning mechanism does not have a significant impact on the emergence of cooperation.

(4) More diverse set of states.

Since Q-learning can have various state set configurations, and different state sets influence an individual's decision-making process, we further adopt two different state sets for the extended experiments.

• **Neighbor Behavior Relationship.** Based on the proportion of cooperators and defectors among the neighbors, i.e., $n_C > n_D$, $n_C = n_D$, and $n_C < n_D$, the state set is defined as $S = \{s'_1, s'_2, s'_3\}$ [53]. As shown in the Fig. 23, the threshold for TP is 2.51, which is lower than the threshold of 2.80 for SE, thereby demonstrating the crucial role of information perception in promoting the emergence of cooperation. When compared to the results from the original state setting, it is observed that the thresholds are advanced. This may be because the neighbor behavior relationships setting allows individuals to more accurately infer group cooperation trends, thus triggering reciprocal strategies earlier.

• **Self-regarding.** The individual's state is no longer dependent on the proportion of cooperators among the neighbors, but is instead determined by the action taken in the previous time step, i.e., self-regarding, with $S = \{C, D\}$ [54]. As shown in the Fig. 24, the

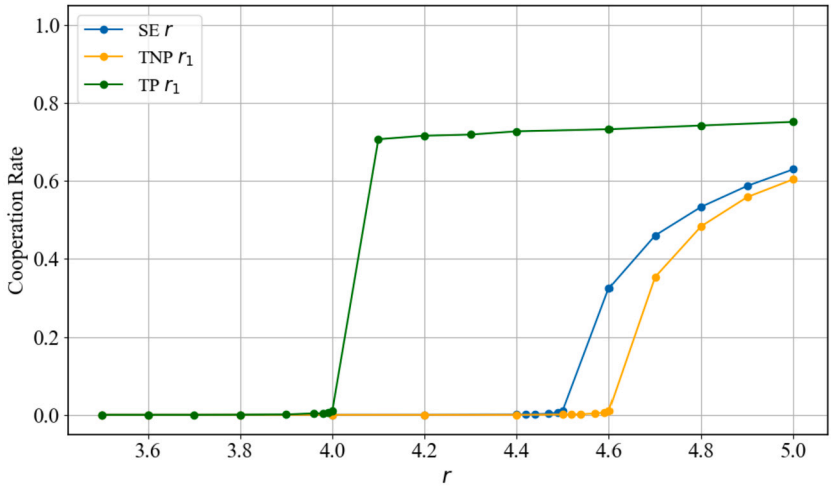

Fig. 20. Cooperation rate at convergence versus the synergy factor r in three systems with a difference in r of 0.1. The cooperation emergence threshold is approximately $r_c = 4.50$ for SE, $r_c = 4.60$ for TNP, and $r_c = 4.00$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

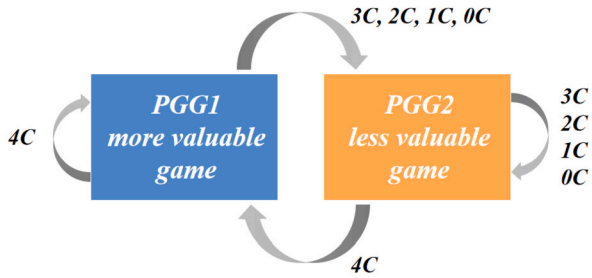

Fig. 21. A more stringent two-state stochastic public goods game model. PGG1 (left, in blue) represents the more valuable game, and **PGG2** (right, in orange) represents the less valuable game.

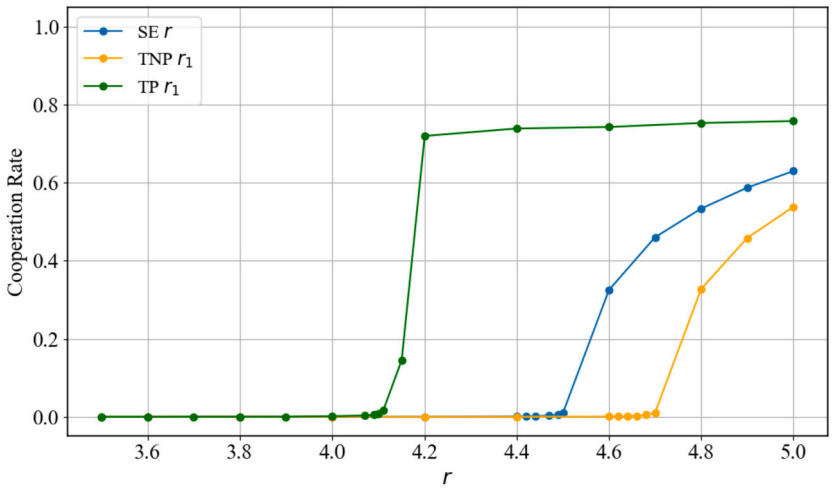

Fig. 22. Cooperation rate at convergence versus the synergy factor r in three systems with a more stringent environmental transition mechanism. The cooperation emergence threshold is approximately $r_c = 4.50$ for SE, $r_c = 4.70$ for TNP, and $r_c = 4.11$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

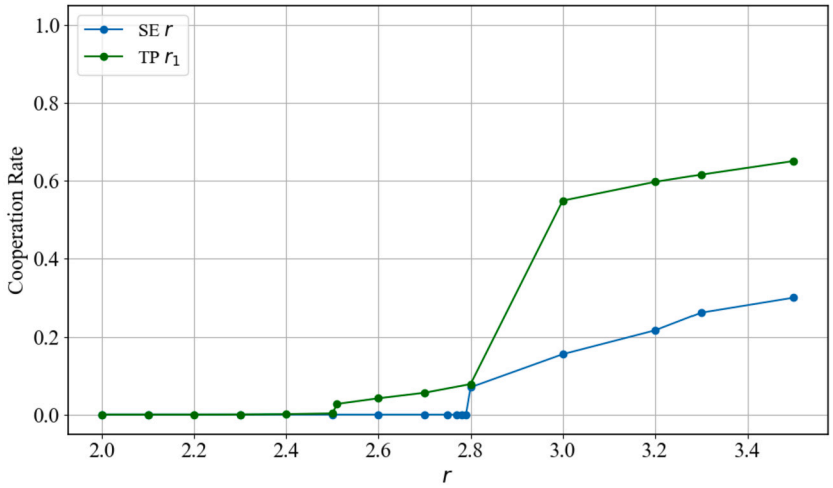

Fig. 23. Cooperation rate at convergence versus the synergy factor r in three systems with the state set $S = \{s_1', s_2', s_3'\}$. The cooperation emergence threshold is approximately $r_c = 2.80$ for SE, and $r_c = 2.51$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

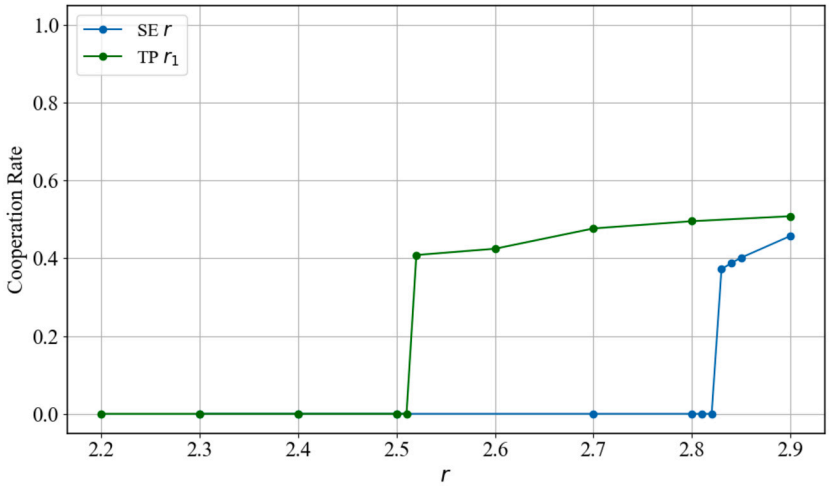

Fig. 24. Cooperation rate at convergence versus the synergy factor r in three systems with the state set $S = \{C, D\}$. The cooperation emergence threshold is approximately $r_c = 2.83$ for SE, and $r_c = 2.52$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

threshold for TP is 2.52, which is lower than the threshold of 2.83 for SE, thereby confirming the significant impact of environmental information perception on the evolution of cooperation. When compared to the results from the original state setting, it is observed that the thresholds are advanced. This may be because the self-regarding setting, by strengthening the direct connection between individual actions and feedback, reduces the interference from group dynamics, thereby decreasing strategy uncertainty and improving the efficiency of cooperation evolution.

(5) A larger von Neumann neighborhood range.

Since the experiments in this study use a von Neumann neighborhood with $\rho = 1$, it is necessary to further investigate the impact of a larger von Neumann neighborhood range on individual strategies and cooperation behaviors.

By increasing the von Neumann neighborhood range to $\rho = 2$, the number of neighbors for each individual increases from the original four to twelve. As shown in the Fig. 25, the threshold for SE is 9.38, the threshold for TNP is 9.58, and the threshold for TP is 7.67, which is lower than the first two systems, demonstrating the critical role of information perception in promoting the emergence of cooperation. Furthermore, when compared to the results with the von Neumann neighborhood $\rho = 1$ in this study, it is observed that the thresholds for all three systems have been delayed. This is because increasing the von Neumann neighborhood range results in a larger number of neighbors for each individual, thereby weakening the clustering effect to some extent. With more interactions between individuals, the emergence of cooperation becomes more difficult.

(6) Different network topologies. Given that the experiments in this paper focus solely on regular networks, it is necessary to further explore the emergence of cooperation in more complex real-world social structures.

The scale-free network. As shown in the Fig. 26, the threshold for SE is 7.22, the threshold for TNP is 7.42, and the threshold for TP is 6.76, which is lower than the first two systems, thereby validating the positive effect of information perception on the

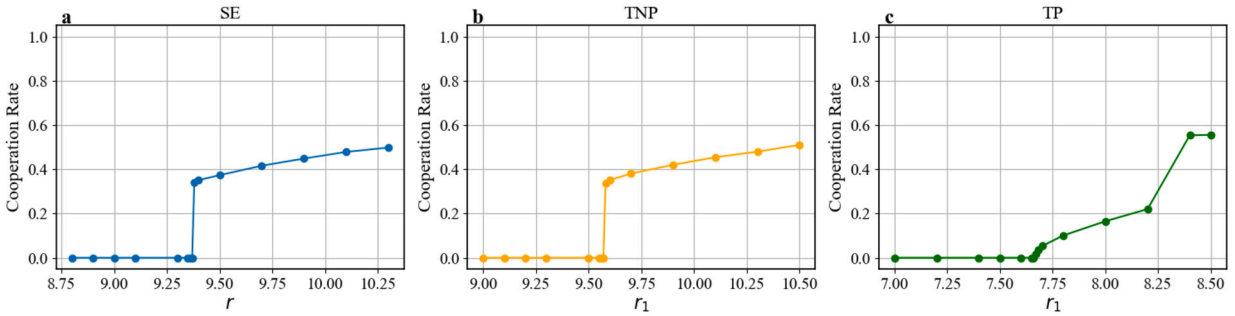

Fig. 25. Cooperation rate at convergence versus the synergy factor r in three systems with von Neumann neighborhood $\rho = 2$. (a) **SE:** The cooperation emergence threshold is approximately $r_c = 9.38$. (b) **TNP:** The cooperation emergence threshold is approximately $r_c = 9.58$. (c) **TP:** The cooperation emergence threshold is approximately $r_c = 7.67$. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

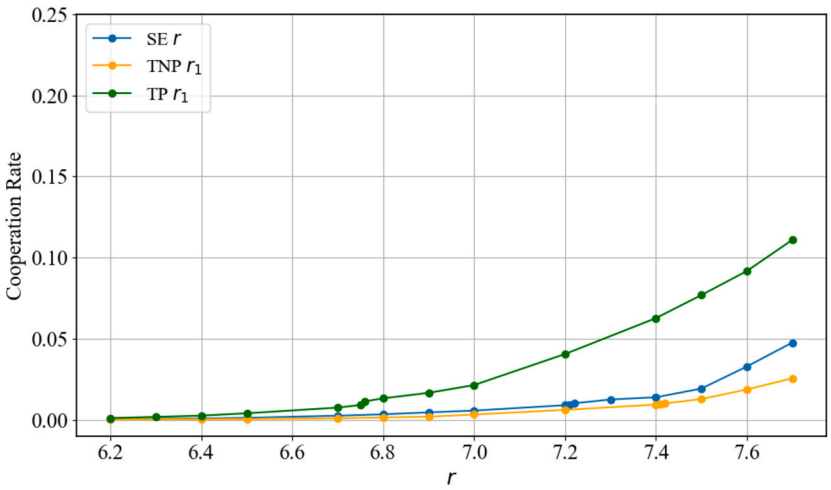

Fig. 26. Cooperation rate at convergence versus the synergy factor r in three systems with the scale-free network. The cooperation emergence threshold is approximately $r_c = 7.22$ for SE, $r_c = 7.42$ for TNP, and $r_c = 6.76$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

emergence of cooperation. At the same time, when compared to the results from the regular network in this study, it is observed that the thresholds for all three systems are delayed. The reason is that scale-free networks exhibit a topology with a long-tail degree distribution, where most nodes have few connections, while a small number of nodes have many connections. This uneven structure reduces the spread and scope of cooperation, thereby inhibiting its emergence.

The small-world network. As shown in the Fig. 27, the threshold for SE is 4.14, the threshold for TNP is 4.34, and the threshold for TP is 3.80, which is lower than the first two systems, demonstrating the critical role of information perception in promoting the emergence of cooperation. Furthermore, when compared to the results from the regular network in this study, it is observed that the thresholds for all three systems have been advanced. This is because the high clustering and short path characteristics of the small-world network structure facilitate the rapid spread of information, resulting in more frequent interactions between individuals and promoting the emergence of cooperation.

4. Conclusion and discussion

This study introduces the stochastic game framework and explores the impact of environmental information perception ability on the evolution of cooperation in the public goods game under a two-state environmental transition mechanism. Compared to SE, we find that the cooperation threshold in TNP is higher. This is because, in TNP, the upper limit of Q_c is lower than Q_d, which causes individuals to tend to choose defection, and thus each individual makes strategy choices only in the less valuable game until a higher r value is reached, at which point cooperation will emerge. In contrast, cooperation emerges earlier in TP. This is because, in TP, individuals can perceive environmental changes in real-time and adjust their decisions based on the current environmental state. Through microscopic analysis, it is found that in cluster-to-cluster contact, the yellow isolation zone phenomenon exists, which allows individuals within the isolation zone to effectively resist the erosion of defection strategies, thus revealing the underlying mechanism of cooperation stability. In the case of individual-to-individual uniform contact, each individual ensures at least two cooperative individuals within the group participating in the public goods game, thereby ensuring that Q_c of each individual remains positive, which in turn stabilizes cooperative behavior in the system.

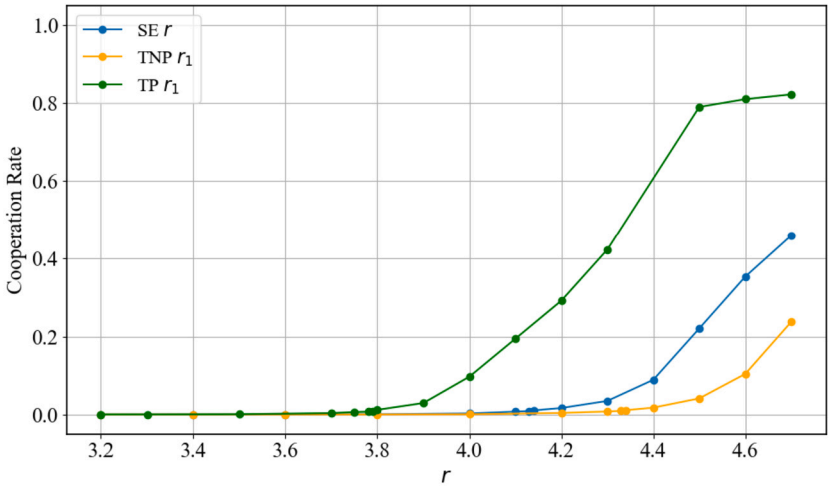

Fig. 27. Cooperation rate at convergence versus the synergy factor r in three systems with the small-world network. The cooperation emergence threshold is approximately $r_c = 4.14$ for SE, $r_c = 4.34$ for TNP, and $r_c = 3.80$ for TP. The results shown were obtained with the following values: $\alpha = 0.1$, $\gamma = 0.9$, $c = 1.0$, $\epsilon = 0.01$, and $L = 100$.

To validate the robustness of the experimental results, we conduct extended experiments from six dimensions: (1) synchronous update interactions; (2) different r-value differences between the more valuable game and the less valuable game; (3) a more stringent environmental transition mechanism; (4) different Q-learning state paradigms; (5) a larger von Neumann neighborhood range; and (6) different network topologies. The results of these experiments all validate the positive effect of environmental information perception on the emergence of cooperation.

Overall, this study emphasizes the important role of environmental information perception in promoting the evolution of cooperation, offering a new theoretical perspective for understanding cooperation mechanisms in stochastic environments. Building on this foundation, future research will incorporate active exclusion strategies to systematically investigate their similarities and differences with environmental feedback mechanisms in fostering cooperation, thereby deepening our understanding of the dynamics underlying collective behavior. Additionally, future research will consider incorporating more detailed ecological or economic mechanisms to model environmental changes, further refining the theoretical framework of the model.

Data availability

Data will be made available on request.

References

[1] R.L. Trivers, The evolution of reciprocal altruism, Q. Rev. Biol. 46 (1) (1971) 35–57.
[2] R. Axelrod, W.D. Hamilton, The evolution of cooperation, Science 211 (4489) (1981) 1390–1396.
[3] E. Fehr, I. Schurtenberger, Normative foundations of human cooperation, Nat. Hum. Behav. 2 (7) (2018) 458–468.
[4] M.A. Nowak, K. Sigmund, Evolutionary dynamics of biological games, Science 303 (5659) (2004) 793–799.
[5] E. Pennisi, On the origin of cooperation, 2009.
[6] M.A. Nowak, Five rules for the evolution of cooperation, Science 314 (5805) (2006) 1560–1563.
[7] C. Hauert, M. Holmes, M. Doebeli, Evolutionary games and population dynamics: maintenance of cooperation in public goods games, Proc. Royal Soc. B, Biol. Sci. 273 (1600) (2006) 2565–2571.
[8] S. Kurokawa, Y. Ihara, Emergence of cooperation in public goods games, Proc. Royal Soc. B, Biol. Sci. 276 (1660) (2009) 1379–1384.
[9] A. Szolnoki, M. Perc, Competition of tolerant strategies in the spatial public goods game, New J. Phys. 18 (8) (2016) 083021.
[10] M. Olson Jr, The Logic of Collective Action: Public Goods and the Theory of Groups, with a new preface and appendix, vol. 124, Harvard University Press, 1971.
[11] E. Fehr, S. Gächter, Altruistic punishment in humans, Nature 415 (6868) (2002) 137–140.
[12] D. Han, S. Yan, D. Li, The evolutionary public goods game model with punishment mechanism in an activity-driven network, Chaos Solitons Fractals 123 (2019) 254–259.
[13] X. Chen, A. Szolnoki, M. Perc, Probabilistic sharing solves the problem of costly punishment, New J. Phys. 16 (8) (2014) 083016.
[14] K. Xie, T. Liu, The regulation of good and evi promotes cooperation in public goods game, Appl. Math. Comput. 478 (2024) 128844.
[15] A. Szolnoki, M. Perc, Second-order free-riding on antisocial punishment restores the effectiveness of prosocial punishment, Phys. Rev. X 7 (4) (2017) 041027.
[16] D.G. Rand, A. Dreber, T. Ellingsen, D. Fudenberg, M.A. Nowak, Positive interactions promote public cooperation, Science 325 (5945) (2009) 1272–1275.
[17] T. Sasaki, S. Uchida, Rewards and the evolution of cooperation in public good games, Biol. Lett. 10 (1) (2014) 20130903.
[18] M. Milinski, D. Semmann, H.-J. Krambeck, Reputation helps solve the 'tragedy of the commons', Nature 415 (6870) (2002) 424–426.
[19] F.P. Santos, F.C. Santos, J.M. Pacheco, Social norm complexity and past reputations in the evolution of cooperation, Nature 555 (7695) (2018) 242–245.
[20] A. Szolnoki, M. Perc, Conditional strategies and the evolution of cooperation in spatial public goods games, Phys. Rev. E, Stat. Nonlinear Soft Matter Phys. 85 (2) (2012) 026104.
[21] J. Quan, X. Yang, X. Wang, Continuous spatial public goods game with self and peer punishment based on particle swarm optimization, Phys. Lett. A 382 (26) (2018) 1721–1730.
[22] L. Yang, Z. Xu, L. Zhang, D. Yang, Benefits of intervention in spatial public goods games, Phys. Lett. A 382 (48) (2018) 3470–3475.

[23] A. Szolnoki, M. Perc, G. Szabó, Topology-independent impact of noise on cooperation in spatial public goods games, Phys. Rev. E, Stat. Nonlinear Soft Matter Phys. 80 (5) (2009) 056109.
[24] M. Perc, J. Gómez-Gardenes, A. Szolnoki, L.M. Floría, Y. Moreno, Evolutionary dynamics of group interactions on structured populations: a review, J. R. Soc. Interface 10 (80) (2013) 20120997.
[25] D.G. Rand, M.A. Nowak, Human cooperation, Trends Cogn. Sci. 17 (8) (2013) 413–425.
[26] F.C. Santos, M.D. Santos, J.M. Pacheco, Social diversity promotes the emergence of cooperation in public goods games, Nature 454 (7201) (2008) 213–216.
[27] M. Mäs, H.H. Nax, A behavioral study of "noise" in coordination games, J. Econ. Theory 162 (2016) 195–208.
[28] W. Barfuss, J.F. Donges, J. Kurths, Deterministic limit of temporal difference reinforcement learning for stochastic games, Phys. Rev. E 99 (4) (2019) 043305.
[29] L.S. Shapley, Stochastic games, Proc. Natl. Acad. Sci. 39 (10) (1953) 1095–1100.
[30] A. Szolnoki, X. Chen, Environmental feedback drives cooperation in spatial social dilemmas, Europhys. Lett. 120 (5) (2018) 58001.
[31] J. Quan, M. Zhang, Y. Zhou, X. Wang, J.-B. Yang, Dynamic scale return coefficient with environmental feedback promotes cooperation in spatial public goods game, J. Stat. Mech. Theory Exp. 2019 (10) (2019) 103405.
[32] X. Wang, F. Fu, Eco-evolutionary dynamics with environmental feedback: cooperation in a changing world, Europhys. Lett. 132 (1) (2020) 10001.
[33] D. Lyu, H. Liu, C. Deng, X. Wang, Promotion of cooperation in a structured population with environmental feedbacks, Chaos, Interdiscip. J. Nonlinear Sci. 34 (12) (2024).
[34] C. Hilbe, Š. Šimsa, K. Chatterjee, M.A. Nowak, Evolution of cooperation in stochastic games, Nature 559 (7713) (2018) 246–249.
[35] L. Yang, L. Zhang, Environmental feedback in spatial public goods game, Chaos Solitons Fractals 142 (2021) 110485.
[36] X. Ma, J. Quan, X. Wang, Evolution of cooperation with nonlinear environment feedback in repeated public goods game, Appl. Math. Comput. 452 (2023) 128056.
[37] M. Nowak, K. Sigmund, A strategy of win-stay, lose-shift that outperforms tit-for-tat in the prisoner's dilemma game, Nature 364 (6432) (1993) 56–58.
[38] M. Kleshnina, C. Hilbe, Š. Šimsa, K. Chatterjee, M.A. Nowak, The effect of environmental information on evolution of cooperation in stochastic games, Nat. Commun. 14 (1) (2023) 4153.
[39] X. Wang, Z. Yang, G. Chen, Y. Liu, Enhancing cooperative evolution in spatial public goods game by particle swarm optimization based on exploration and q-learning, Appl. Math. Comput. 469 (2024) 128534.
[40] Y. Shen, Y. Ma, H. Kang, X. Sun, Q. Chen, Learning and propagation: evolutionary dynamics in spatial public goods games through combined q-learning and Fermi rule, Chaos Solitons Fractals 187 (2024) 115377.
[41] C.D. Brummitt, H. Delventhal, M. Retzlaff, Packard snowflakes on the von Neumann neighborhood, J. Cell. Autom. 3 (1) (2008) 57–80.
[42] J. Qin, Y. Chen, W. Fu, Y. Kang, M.M. Perc, Neighborhood diversity promotes cooperation in social dilemmas, IEEE Access 6 (2017) 5003–5009.
[43] K.L. Chung, Markov Chains, Springer-Verlag, New York, 1967.
[44] S. Chib, Markov chain Monte Carlo methods: computation and inference, Handb. Econom. 5 (2001) 3569–3649.
[45] C.J.C.H. Watkins, Learning from delayed rewards (1989).
[46] C.J. Watkins, P. Dayan, Q-learning, Mach. Learn. 8 (1992) 279–292.
[47] B. Schönfisch, A. De Roos, Synchronous and asynchronous updating in cellular automata, Biosystems 51 (3) (1999) 123–143.
[48] H.J. Blok, B. Bergersen, Synchronous versus asynchronous updating in the "game of life", Phys. Rev. E 59 (4) (1999) 3876.
[49] H. Zhang, T. An, P. Yan, K. Hu, J. An, L. Shi, J. Zhao, J. Wang, Exploring cooperative evolution with tunable payoff's loners using reinforcement learning, Chaos Solitons Fractals 178 (2024) 114358.
[50] M. Wunder, M.L. Littman, M. Babes, Classes of multiagent q-learning dynamics with epsilon-greedy exploration, in: Proceedings of the 27th International Conference on Machine Learning (ICML-10), 2010, pp. 1167–1174.
[51] C. Hauert, S. De Monte, J. Hofbauer, K. Sigmund, Replicator dynamics for optional public good games, J. Theor. Biol. 218 (2) (2002) 187–194.
[52] L. Fan, Z. Song, L. Wang, Y. Liu, Z. Wang, Incorporating social payoff into reinforcement learning promotes cooperation, Chaos, Interdiscip. J. Nonlinear Sci. 32 (12) (2022).
[53] G. Zheng, J. Zhang, S. Deng, W. Cai, L. Chen, Evolution of cooperation in the public goods game with q-learning, Chaos Solitons Fractals 188 (2024) 115568.
[54] L. Wang, D. Jia, L. Zhang, P. Zhu, M. Perc, L. Shi, Z. Wang, Lévy noise promotes cooperation in the prisoner's dilemma game with reinforcement learning, Nonlinear Dyn. 108 (2) (2022) 1837–1845.