Third-Party Intervention of Cooperation in Multilayer Networks

Hao Guo, Zhao Song, Matjaž Perc, *Member, IEEE*, Xuelong Li, *Fellow, IEEE*, and Zhen Wang, *Senior Member, IEEE*

***Abstract*—The conflicts in human societies have often been studied through evolutionary games. In social dilemmas, for example, individuals fair best if they defect, but the society is best off if everybody cooperates. Cooperation therefore often requires a mechanism or third parties to evolve and remain viable. To study how third parties affect the evolution of cooperation, we develop a novel game theoretic framework composed of two layers. One layer contains cooperators and defectors, while the other, the third-party layer, contains interveners. Interveners can be peacemakers, troublemakers, or a hybrid of these two. Focusing on two-player two-strategy games, we show that intervention, as an exogenous factor, can stimulate (or inhibit) cooperation by weakening (or strengthening) the dilemma strength of the game the disputant plays. Moreover, the outcome in the disputant layer that is triggered by intervention, in turn, stimulates its own evolution. We analyze the co-evolution of intervention and cooperation and find that even a minority of interveners can promote higher cooperation. By conducting stability analyses, we derive the conditions for the emergence of cooperation and intervention. Our research unveils the potential of third parties to control the evolution of cooperation.**

***Index Terms*—Cooperative systems, decision making, dynamics, game theory, networks.**

Manuscript received 15 March 2023; accepted 4 May 2023. Date of publication 3 July 2023; date of current version 17 October 2023. This work was supported in part by the National Science Fund for Distinguished Young Scholars under Grant 62025602; in part by the National Natural Science Foundation of China under Grant U22B2036, Grant 11931015, and Grant 81961138010; in part by New Cornerstone Science Foundation through the XPLORER PRIZE; in part by the Fok Ying-Tong Education Foundation, China, under Grant 171105; and in part by the Key Technology Research and Development Program of Science and Technology-Scientific and Technological Innovation Team of Shaanxi Province under Grant 2020TD-013. This article was recommended by Associate Editor L. C. Rego. *(Corresponding author: Zhen Wang.)*

Hao Guo and Zhao Song are with the School of Mechanical Engineering and the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China (e-mail: ghfreezing@outlook.com; songzhao@mail.nwpu.edu.cn).

Matjaž Perc is with the Faculty of Natural Sciences and Mathematics, University of Maribor, 2000 Maribor, Slovenia, also with the Department of Medical Research, China Medical University Hospital, China Medical University, Taichung 404332, Taiwan, also with Complexity Science Hub Vienna, 1080 Vienna, Austria, also with Alma Mater Europaea, 2000 Maribor, Slovenia, and also with the Department of Physics, Kyung Hee University, Seoul 02447, Republic of Korea (e-mail: matjaz.perc@gmail.com).

Xuelong Li is with the School of Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China (e-mail: xuelong_li@nwpu.edu.cn).

Zhen Wang is with the School of Mechanical Engineering, the School of Artificial Intelligence, Optics and Electronics, and the School of Cybersecurity, Northwestern Polytechnical University, Xi'an 710072, Shaanxi, China (e-mail: zhenwang0@gmail.com).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/TSMC.2023.3278048.

Digital Object Identifier 10.1109/TSMC.2023.3278048

I. Introduction

Conflicts that frequently occur in both society and engineering are consumptive and destructive. Although it can be solved privately by pairwise participants, third-party intervention plays a crucial role in mediating conflicts with public and transparent rules [1]. Of course, invalid or vicious interventions may activate conflicts. The mathematical portrait of the conflict between individual and collective interests can be simplified by the competition between cooperators and defectors [2], [3]. Cooperation [4], as an altruistic social behavior, can significantly benefit society. However, due to the cost, cooperation is constantly exploited by defection in a competitive environment. This poses a significant challenge to encouraging individuals to prioritize collective interest and stimulate cooperation.

Evolutionary game theory (EGT) [5], [6], [7], [8] provides a powerful tool for portraying social dilemmas, and has attracted growing interest across disciplines including computer science, mathematics, and statistics. In particular, two-player two-strategy (TPTS) games [9], [10], [11] are always employed to study conflicts between unfamiliar individuals, where each individual chooses a strategy (cooperation or defection) without knowing the opponent's strategy. Tools and concepts from EGT are also popular in characterizing and analyzing agents' preferences in different competitive behaviors [12], [13]. Based on EGT, social mechanisms, such as reward [14] and punishment [15], [16], [17], [18], as well as individual characteristics, such as memory [19], self-recommendation [20], [21], income redistribution [22], [23], and reputation [13], [24], are proposed to reveal the factors that contribute to the existence and maintenance of cooperation.

The rapidly developed network science provided a new direction for studying EGT [8], [25], [26]. Population game models with graphical strategy interactions have also attracted extensive research [27], [28]. Likewise, community networks play an important role in the stability of strategy evolution, where the interactions within a community are compact, and the interactions between communities are sparse [29]. The interdependent networks, controlled by coupling strength [30] and degree correlation [31], have shown effective influence on cooperative behavior. Furthermore, dynamic networked systems also provide new insights into understanding the emergence of cooperation [20], [21].

Nevertheless, what has been mentioned above only takes into account endogenous factors. To mediate conflicts, interventions from third parties play an essential role in human life. In the context of government policy, for example, interventions can reduce the negative impact on the environment while encouraging green production [32]. Furthermore, many experimental results suggest that third-party punishment is an important factor in explaining high levels of human cooperation [33], [34], [35]. Theoretical analysis also reveals a cost-effective external intervention for promoting fairness and cooperation in the prisoner's dilemma game (PDG) and ultimatum game [36], [37]. The optimal incentive that minimizes intervention cost while maximizing the benefit has also been explored in the context of public cooperation [38]. These works consider a single population model and study how to provide intervention in a cost-efficient way. However, they have a little discussion about the emergence of intervention and ignore that third parties are essentially a group and may be risky [39]. A failed intervention may have to bear the consequence of loss, which is a selfish reason that one gives up being an intervener and becomes a silencer. Therefore, it is natural to ask: How to develop a system to study the interplay of intervention and cooperation? How does intervention, as an external factor, control the evolution of cooperation? What is the reason for the emergence of intervention and cooperation?

In addition to examining the unilateral impact of interventions on conflicts [38], [40], it is also necessary to consider how intervention outcomes affect the behavior of third parties. It is believed that parents (or supervisors) can influence how their children perceive and respond to conflicts. By adopting various approaches, parents may either weaken or strengthen their children's attitudes toward conflict, or a combination of both. Another typical example is when conflicts arise between employees, the employer often acts as a mediator, as the benefits of intervention generally outweigh those of nonintervention. Simultaneously, a company may gain more if it exacerbates conflicts between other companies. This scenario also occurs between countries. Therefore, the underlying rewards gained from intervention are critical to motivating individuals or entities to get involved in the conflict.

We here address these questions by proposing a framework that couples third parties with disputant players to understand how outcome-based interveners affect the evolutionary dynamics in disputant players. Moreover, this framework involves two layers, one is the disputant layer, and the other is the third-party layer. Specifically, players in the disputant layer participate in TPTS games that an exogenous environment can control, i.e., the strategy of third parties. Meanwhile, the third-party layer contains interveners whose payoff is closely related to the evolutionary outcome in the disputant layer, i.e., the distribution of cooperation and defection. We answer our key research questions by analyzing the interplay between cooperation and intervention using replicator equations for infinitely large populations and Monte Carlo simulations (MCSs) for finitely large square lattices. We provide the condition where cooperation and intervention dominate the respective layer (see Theorem 1). Furthermore, complete cooperation in the disputant layer is not necessary for the dominance of intervention (see Theorems 3 and 6). The emergence of cooperation is influenced by the dilemma strength of the basic game if there is no intervention (see Theorems 2 and 5), or the strength of intervention if intervention exists (see Theorem 6). It is noteworthy that in some cases, a minority of interventions can actually encourage a majority of disputants to engage in cooperation. The simulation results obtained from finitely large square lattices provide more evidence to support our conclusions.

Our main contributions are summarized as follows.

1) We develop a novel evolutionary game theoretical framework to model the coupling effects between strategic conflicts and third-party intervention. This framework overcomes the limitation that only unidirectional relationships are considered in previous studies and allows for analyzing the dynamics of a coupled system.
2) The model enables us to explore the interplay between the intervener's type and income-preference pattern (IPP). We propose three types of interveners based on their effects on the dilemma strength, including peacemakers, troublemakers, and a hybrid of the two. Particularly, we demonstrate that peacemakers are effective at promoting cooperation. Furthermore, IPPs of intervention are crucial in shaping the coexistence of cooperation and intervention.
3) We show that intervention, by itself, can regulate individual decision making by monitoring strategic conflicts between disputants. The outcome in the disputant layer, which is instigated by intervention, in turn, stimulates its own evolution. This provides a new viewpoint for understanding the source of cooperation and intervention.
4) By analyzing the co-evolution dynamics of cooperation and intervention, we find various equilibria and derive their stability conditions. These include monostable states, such as co-extinction, co-dominance, and coexistence of cooperation and intervention, as well as bistable states under different IPPs. We then expand this system into networks with local interactions and develop an evolutionary game transition algorithm. Our research unveils the potential of third parties to control the evolution of cooperation.

The remainder of this article is organized as follows. In Section II, we give the notations and preliminaries. Section III formulates the system coupling problem of third party and human conflict. In Section IV, we give the model description and theory results of infinitely large well-mixed populations. In Section V, we provide the agent-based model and simulation results of square lattices. Finally, we conclude this article in Section VI.

II. Notations and Preliminaries

The notation of this article is summarized as follows. $\mathcal{S} = \{C, D\}$ and $\mathcal{A} = \{I, Q\}$ represent the strategy set of players in the disputant layer ($\mathcal{D}$) and the third-party layer ($\mathcal{T}$), respectively. Denote the payoff matrix of a TPTS game as

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

where mutual cooperation acquires a reward R, while mutual defection receives a punishment P. A cooperator obtains a sucker's payoff S if interacting with a defector who obtains

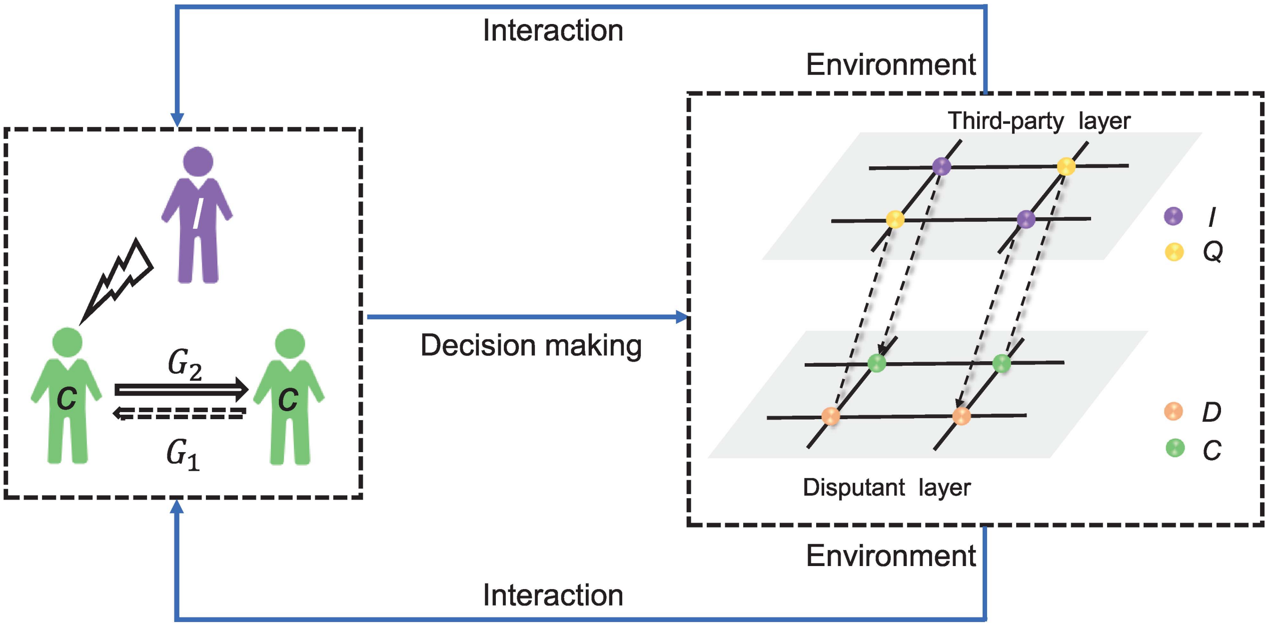


Fig. 1. Interactions with third-party intervention. Players in the disputant layer have two strategies, cooperation and defection. Players in the third-party layer have two strategies, intervention and silence. There is generally no restriction on whether the topologies of different layers are the same or different. Here, we take square lattices as an example. The coupling effect between two layers is divided into two parts: 1) interveners mediate conflicts between cooperation and defection and 2) interveners are rewarded according to the evolutionary outcomes of the disputant layer. Each player updates its strategy according to the payoff obtained in this interactive environment.

temptation T simultaneously. In detail, the game is a PDG if the parameters satisfy $T > R > P > S$; snowdrift game (SDG) if the parameters satisfy $T > R > S > P$; stag hunt game (SHG) if the parameters satisfy $R > T > P > S$; harmony game (HG) if the parameters satisfy $R > T, S > P$. To measure the strength of social dilemma, Wang et al. [41] rescaled these four parameters as two indicators, named risk-averting and gamble-intending dilemma, defined by $D_r = (P - S/R - P)$ and $D_g = (T - R/R - P)$, respectively. π_* represents the payoff of strategy $*$, and P_i is the payoff of player i. x and ϕ represent the fraction of cooperation in the disputant population and intervention in third-party population, respectively. Denote $\dot{x}$ and $\dot{\phi}$ as x's and ϕ's derivative with respect to time, respectively. J represents Jacobian in the stability analysis. $W_{\mathcal{S}_i \leftarrow \mathcal{S}_j}$ is the probability that player i imitates the strategy of j.

Denote $\mathcal{G} = \{\mathcal{V}, \mathcal{E}\}$ as a network, where $\mathcal{V} = \{1, 2, \ldots, N\}$ represents the node set, $\mathcal{E} \subseteq \mathcal{V} \times \mathcal{V}$ is link set. Let $a_{ij} \in \mathbb{R}$ be the element of adjacent matrix, if the ith player has a connection with jth player $a_{ij} = 1$; otherwise, $a_{ij} = 0$. Here, we consider an undirected and connected network, thus the degree of each node $k_i = \sum_{j=1}^{N} a_{ij}$. If $k_i = k_j \quad \forall i, j \in \mathcal{V}$, $\mathcal{G}$ is a homogeneous network. We call $\mathcal{G}$ as complete graph if $k_i = N - 1 \quad \forall i \in \mathcal{V}$. A complete graph with the same weight is also known as a well-mixed population in EGT. In particular, $N \to \infty$ means an infinitely large well-mixed population.

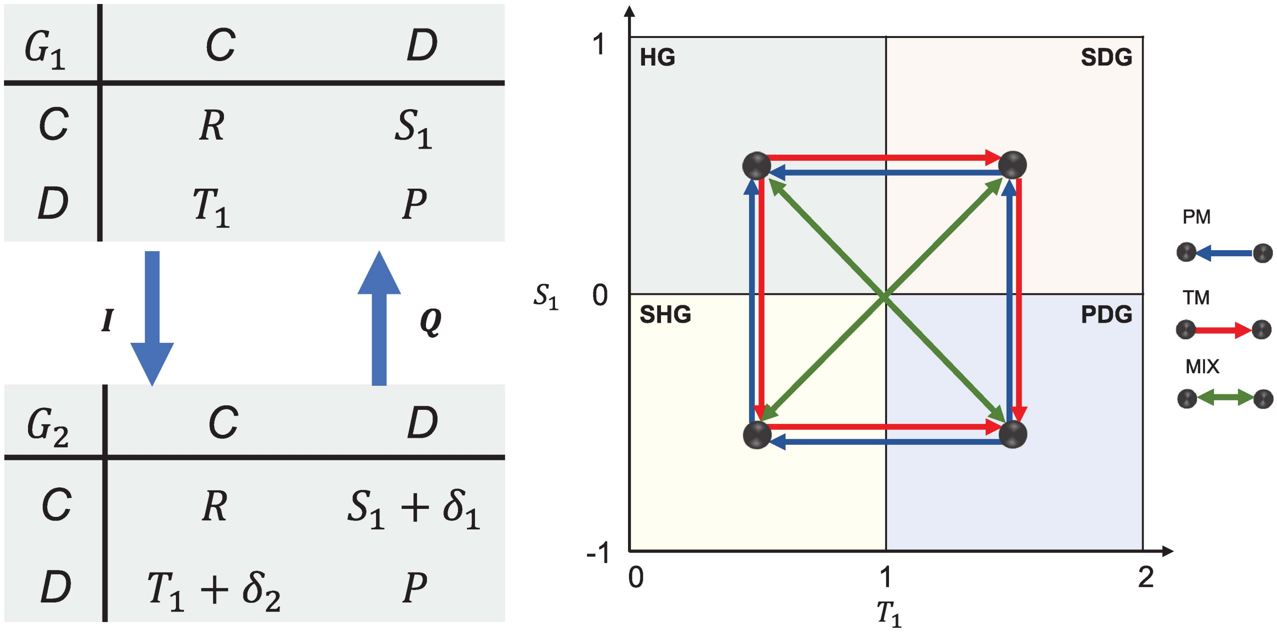
III. Problem Formulation

Since many conflict scenarios involve competition between cooperation and defection, we employ TPTS games [41]. In detail, players in the disputant layer have the same opportunity to choose cooperation (C) or defection (D) from set $\mathcal{S}$. Meanwhile, as an exogenous factor, players in the third-party layer can choose either intervention (I) or silence (Q) from set $\mathcal{A}$. Intervention to mediate conflicts between disputant players is rewarded according to the outcome of the disputants. Therefore, the system we study can be modeled by multilayer networks composed of disputant layer $\mathcal{D} = \mathcal{G}^{\mathcal{D}}$ and third-party layer $\mathcal{T} = \mathcal{G}^{\mathcal{T}}$. A sketch is given in Fig. 1, where we take square lattices as an example. Since nodes between two layers are one-to-one, the node sets are identical, and $\mathcal{V}^{\mathcal{D}} = \mathcal{V}^{\mathcal{T}}$. The edges between nodes in each particular layer of this system can be the same or different. Subsequently, the coupled effect can be depicted by an additional edge between two layers. This multilayer network is similar to an interconnected network where nodes have intraconnections within their own network, and interconnections with the other network [42]. The difference is that nodes between two layers are one-to-one. Specifically, an intervener controls the conflict by intervening in the game that its corresponding disputant play. Evolutionary outcomes related to this disputant then affect the payoff obtained by the intervener.

The basic game (G_1) involved in the disputant layer is given by the payoff matrix M_1. For simplicity yet without loss of generality, we here set $R = 1$ and $P = 0$ throughout this article [see Fig. 2(a)]

$$M_1 = \begin{pmatrix} 1 & S_1 \\ T_1 & 0 \end{pmatrix}. \tag{2}$$

As the term dilemma strength is closely associated with the equilibrium of the game [41], we here utilize it to measure the conflicts between disputant players. Two types of



(a) (b)

Fig. 2. Game transitions triggered by third parties. (a) If a player on the disputant layer corresponds with a silence strategy, it participates in the basic game G_1. On the other hand, the player supervised by a third-party intervention participates in the game G_2. (b) Types of the intervention. The intervener is called a peacemaker if it reduces the dilemma strength, a troublemaker if it strengthens the dilemma strength, and a mixer if these two happen simultaneously.

dilemma strength are employed, including D_r and D_g. The first term D_r measures a risk-averting dilemma, and the second term D_g measures a gamble-intending dilemma. The dilemma strength of G_1 therefore can be controlled by S_1 and T_1, namely, $D_{r1} = -S_1$ and $D_{g1} = T_1 - 1$. Generally, the higher the dilemma strength, the lower the cooperation rate. In our proposed model, players' gains in the disputant layer depend not only on their own and neighbors' strategies but also on third parties who act as exogenous environments and can trigger game transitions. Specifically, if supervised by an intervener, the player in the disputant layer will participate in another game (G_2) whose payoff matrix is determined as follows:

$$M_2 = \begin{pmatrix} 1 & S_2 \\ T_2 & 0 \end{pmatrix} \tag{3}$$

where $S_2 = S_1 + \delta_1$ and $T_2 = T_1 + \delta_2$. $-1 \le \delta_1, \delta_2 \le 1$ measure the strength of intervention. Subsequently, the dilemma strength of G_2 can be given by T_1, S_1, δ_1, and δ_2, namely, $D_{r2} = -S_2 = -S_1 - \delta_1$ and $D_{g2} = T_2 - 1 = T_1 + \delta_2 - 1$. The relationship between dilemma strengths is $D_{r2} = D_{r1} - \delta_1$, $D_{g2} = D_{g1} + \delta_2$. It is easy to see that the dilemma strength of G_2 is strengthened (weakened) if $\delta_1 < 0$ or $\delta_2 > 0$ ($\delta_1 > 0$ or $\delta_2 < 0$). A hybrid effect emerges if δ_1 and δ_2 have the same sign. According to the variation in dilemma strength [see Fig. 2(b)], here we define the type of the third-party intervener as a peacemaker (PM) if $\delta_1 > 0$ or $\delta_2 < 0$ (the dilemma strength of G_2 is weakened), as a troublemaker (TM) if $\delta_1 < 0$ or $\delta_2 > 0$ (the dilemma strength of G_2 is strengthened), and as a mixer (MIX) if $\delta_1 > 0$ and $\delta_2 > 0$, or $\delta_1 < 0$ and $\delta_2 < 0$ are satisfied. When it comes to third parties, the payoff of interveners is determined by two key factors: 1) the evolutionary outcome in the disputant layer, which can be reflected by strategy pairs between the corresponding player and its connected neighbors and 2) the IPP, which refers to the payoff received from distinct pairs. In contrast, the payoff of the silence strategy is fixed and does not depend on external conflicts.

IV. Infinitely Large Well-Mixed Population

A. Coupled Replicator Equation

We first consider infinitely large well-mixed populations where each player has the same probability of interacting with other players in the same population. Due to coupling with third parties, interactions between players in the disputant layer are influenced by the frequency of intervention. Therefore, the expected payoffs of cooperation and defection are given as follows:

$$\begin{aligned} \pi_C &= \phi(xR + (1-x)S_2) + (1-\phi)(xR + (1-x)S_1) \\ \pi_D &= \phi(xT_2 + (1-x)P) + (1-\phi)(xT_1 + (1-x)P) \end{aligned} \tag{4}$$

where x and $1-x$ mean the frequency of cooperation and defection in the disputant layer. ϕ and $1-\phi$ represent the fraction of intervention and silence in the third-party layer. The first term of the right-hand side represents the payoff received from G_2 (under intervention), while the second term means the payoff received from G_1 (without intervention). As stated above, we denote the payoff of the silence strategy by β. While the payoff of intervention is determined by the distribution of pairs between cooperators (CC-pair), pairs between cooperators and defectors (CD-pair), and pairs between defectors (DD-pair). Thus, the expected payoffs of intervention and silence are given as follows:

$$\pi_I = A_1 x^2 + 2A_2 x(1-x) + A_3(1-x)^2$$

$$\pi_Q = \beta \tag{5}$$

where A_1, A_2, and A_3 are the gains that intervener obtains from *CC*-, *CD*- and *DD*-pairs, respectively. Here, we define the IPP as *CC*-pair dominance if $\max(A_1, A_2, A_3) = A_1$, *CD*-pair dominance if $\max(A_1, A_2, A_3) = A_2$, and *DD*-pair dominance if $\max(A_1, A_2, A_3) = A_3$.

Replicator equation [29], [43], [44], [45] is a powerful tool to describe the evolutionary dynamics of collective behavior. Here, we illustrate the dynamics of this system by the fraction of cooperation x and intervention ϕ, satisfying the replicator equation as follows:

$$\begin{cases} \dot{x} = x(\pi_C - \bar{\pi}_1) := f(x, \phi) \\ \dot{\phi} = \phi(\pi_I - \bar{\pi}_2) := g(x, \phi) \end{cases} \tag{6}$$

where the dot means the derivative with respect to time. $\bar{\pi}_1$ and $\bar{\pi}_2$ represent the expected payoff of the disputant and third-party layers, respectively. Subsequently, the expected payoff can be calculated by

$$\begin{aligned} \bar{\pi}_1 &= x\pi_C + (1-x)\pi_D \\ \bar{\pi}_2 &= \phi\pi_I + (1-\phi)\pi_Q. \end{aligned} \tag{7}$$

By considering a mean-field (MF) description, ignoring the spatial topology and stochasticity in evolutionary dynamics, the trajectories of x and ϕ are determined by the expected payoff of cooperation and intervention, respectively. Note that there is no motivation to choose silence if $\min(A_1, A_2, A_3) > \beta$, because intervention is a gain-only option and $g(x, \phi) \geq 0$ is always true despite of x. On the other hand, there is no motivation to choose intervention if $\max(A_1, A_2, A_3) < \beta$, because silence is a gain-only option. Improved by [39], we consider a gain-and-loss scenario here to reflect the underlying risk that comes with intervention, namely, $\max(\pi_I) > \pi_Q$ and $\min(\pi_I) < \pi_Q$. In the remainder of this section, we first give general results and then discuss three special cases by fixing A_1, A_2, and A_3.

B. Equilibrium and Stability Analysis

By solving the coupled replicator equation given by (6), we can derive several fixed (or equilibrium) points.

1) $x = 0$ and $\phi = 0$, i.e., equilibrium $F_1 = (0, 0)$ which means co-extinction of C and I.
2) $x = 1$ and $\phi = 0$, i.e., equilibrium $F_2 = (1, 0)$ which means a polarized state with complete C and extinction of I.
3) $x = 0$ and $\phi = 1$, i.e., equilibrium $F_3 = (0, 1)$ which means a polarized state with complete I extinction of C.
4) $x = 1$ and $\phi = 1$, i.e., equilibrium $F_4 = (1, 1)$ which means co-dominance of C and I.
5) $x = (S_1/[S_1 + T_1 - 1])$ and $\phi = 0$, i.e., equilibrium $F_5 = (S_1/[S_1 + T_1 - 1], 0)$ which means the existence of C in the absence of I.
   Note that this equilibrium point exists if and only if $S_1 + T_1 \neq 1$ and $0 < x < 1$.
6) $x = ([S_1 + \delta_1]/[S_1 + T_1 + \delta_1 + \delta_2 - 1])$ and $\phi = 1$, i.e., equilibrium $F_6 = ([S_1 + \delta_1]/[S_1 + T_1 + \delta_1 + \delta_2 - 1], 1)$ which means the existence of C in the presence of I.
   Note that if and only if $S_1 + T_1 + \delta_1 + \delta_2 \neq 1$ and $0 < x < 1$, this equilibrium point exists. In addition, there are two interior equilibrium points that depend on the value of A_1, A_2, and A_3.
7) $x^* = ([A_3 - A_2 \pm \sqrt{A_2^2 + \beta(A_1 + A_3 - 2A_2) - A_1A_3}]/[A_1 - 2A_2 + A_3])$ and $\phi^* = ([S_1 - (S_1 + T_1 - 1)x^*]/[(\delta_1 + \delta_2)x^* - \delta_1])$, i.e., equilibrium F_7 and F_8 which mean the co-existence of C, D, I, and Q.

Note that if and only if $A_1 - 2A_2 + A_3 \neq 0$, $(\delta_1 + \delta_2)x^* - \delta_1 \neq 0$, $0 < x^* < 1$, and $0 < \phi^* < 1$, these equilibrium points exist. Solution of $\pi_C - \pi_D = 0$ and $\pi_I - \pi_Q = 0$ yields the interior equilibrium points.

To determine the stability of each fixed point, we use Lyapunov's indirect method. By doing so, Jacobian is given as follows:

$$J = \begin{bmatrix} \frac{\partial f(x,\phi)}{\partial x} & \frac{\partial f(x,\phi)}{\partial \phi} \\ \frac{\partial g(x,\phi)}{\partial x} & \frac{\partial g(x,\phi)}{\partial \phi} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \tag{8}$$

where

$$\begin{aligned} a_{11} &= [(3\delta_1 + 3\delta_2)\phi + 3S_1 + 3T_1 - 3]x^2 \\ &\quad + [(-4\delta_1 - 2\delta_2)\phi - 4S_1 - 2T_1 + 2]x + \delta_1\phi + S_1 \\ a_{12} &= x(x-1)[(\delta_1 + \delta_2)x - \delta_1] \\ a_{21} &= -2\phi(\phi - 1)[(A_1 - 2A_2 + A3)x + A_2 - A_3] \\ a_{22} &= -2(\phi - 0.5) \\ &\quad \Big[(A_1 - 2A_2 + A3)x^2 + (2A_2 - 2A_3)x + A_3 - \beta\Big]. \end{aligned} \tag{9}$$

Then, the characteristic function of the linear system is

$$\lambda^2 - \mathrm{Tr}\lambda + \Delta = 0 \tag{10}$$

where

$$\begin{aligned} \mathrm{Tr} &= a_{11} + a_{22} \\ \Delta &= |J| = a_{11}a_{22} - a_{12}a_{21}. \end{aligned} \tag{11}$$

Solving the characteristic roots of the characteristic equation yields $\lambda = ([\mathrm{Tr} \pm \sqrt{\mathrm{Tr}^2 - 4\Delta}]/2)$. The fixed point is asymptotically stable provided that the real part of all the characteristic roots is less than 0, i.e., $(Re(\lambda_k) < 0)$. This is equivalent to the condition that the trace of matrix J is less than 0 and the determinant is greater than 0, that is, $\mathrm{Tr} < 0$, $\Delta > 0$. It is worth noting that stability in this part refers to locally asymptotic stability [46]. Then, we showcase the following theorems.

Theorem 1: The equilibrium point (1, 1) is the stable state if $\delta_2 < 1 - T_1$ and $A_1 > \beta$.

Proof: The trace and determinant of equilibrium point (1, 1) are $T_2 - 1 - A_1 + \beta$ and $-(T_2 - 1)(A_1 - \beta)$. When $\delta_2 < 1 - T_1$ and $A_1 > \beta$, the trace and determinant satisfy $\mathrm{Tr} < 0$ and $\Delta > 0$. Thus, equilibrium point (1, 1) is stable. ■

Based on the parameters established within the proven range, this theorem shows that co-dominance of C and I can be achieved. Particularly, this condition has no requirement for the dilemma strength of basic game G_1 and the payoff from *CD*- and *DD*-pair. If the strength of intervention is powerful enough ($\delta_2 < 1 - T_1$) and the payoff from CC-pair is larger than that from choosing silence, cooperation and intervention can dominate their own layer. As intervention gains from CC-pair, the more cooperation in the disputant layer, the better for the evolution of intervention. This means that mitigating conflict in the disputant layer is beneficial for the evolution of cooperation and the profit of intervention.

Theorem 2: The equilibrium point (1, 0) is the stable state if $T_1 < 1$ and $A_1 < \beta$.

Proof: The trace and determinant of equilibrium point $(1\,,0)$ are $T_1 - 1 + A_1 - \beta$ and $(T_1 - 1)(A_1 - \beta)$. When $T_1 < 1$ and $A_1 < \beta$, the trace and determinant satisfy $\mathrm{Tr} < 0$ and $\Delta > 0$. Thus, the equilibrium point $(1\,,0)$ is stable. ■

We clarify that in the absence of third-party intervention, cooperation dominates the disputant layer only when $T_1 < 1$. The stability of this point is influenced effectively by the basic game G_1. Therefore, if playing PDG and SDG, disputants can never reach a complete cooperation state. On the other hand, in the case of cooperation dominating the disputant layer, intervention vanishes only when $A_1 < \beta$.

Corollary 1: Cooperation can dominate in the disputant layer if $1 - T_1 > \min(\delta_2, 0)$.

Incorporating Theorems 1 and 2, we can conclude that the domination of cooperation is fully determined by the value of $T_1 - 1$, δ_2, A_1, and β. If $1 - T_1 > \min(\delta_2, 0)$, cooperation can always dominate either $A_1 > \beta$ or $A_1 < \beta$. This indicates that besides the dilemma strength of G_1, intervention strength also plays an important role in deciding the domination of cooperation.

Theorem 3: The equilibrium point (0, 1) is the stable state if $\delta_1 < -S_1$ and $A_3 > \beta$.

Proof: The trace and determinant of equilibrium state $(0\,,1)$ are $S_1 + \delta_1 - A_3 + \beta$ and $-(S_1 + \delta_1)(A_3 - \beta)$. When $\delta_1 < -S_1$ and $A_3 > \beta$, the trace and determinant satisfy $\mathrm{Tr} < 0$ and $\Delta > 0$. Thus, the equilibrium point (0, 1) is stable. ■

This theorem reveals that in the presence of intervention, cooperation vanishes when $\delta_1 < -S_1$. It means that the stability of this equilibrium point is closely related to intervention strength. On the other hand, the complete intervention state relies on the payoff from DD-pair, i.e., the condition $A_3 > \beta$.

Theorem 4: The equilibrium point (0, 0) is the stable state if $S_1 < 0$ and $A_3 < \beta$.

Proof: The trace and determinant of equilibrium point (0, 0) are $S_1 + A_3 - \beta$ and $S_1(A_3 - \beta)$. When $S_1 < 0$ and $A_3 < \beta$, the trace and determinant satisfy $\mathrm{Tr} < 0$ and $\Delta > 0$. Thus, the equilibrium point (0, 0) is stable. ■

We emphasize that if and only if $S_1 < 0$ and $A_3 < \beta$, the co-extinction of C and I occurs. In contrast, this point will not be stable if the basic game that disputants participate in is SDG and HG. Taking Theorem 3 into consideration, the extinction condition of cooperation is determined by S_1, δ_1, A_3, and β.

Corollary 2: Given cooperation dominates (or is extinct) in the disputant layer, the dominance of intervention in the third-party layer relies on A_1 (or A_3) and β.

The previous discussion shows that when cooperation dominates the disputant layer, intervention can dominate the third-party layer if it benefits more from CC-pair than silence, i.e., $A_1 > \beta$. However, when defection dominates the disputant layer, intervention can dominate the third-party layer if it benefits more from DD-pair than silence, i.e., $A_3 > \beta$. In addition to these equilibrium points, we also find two boundary solutions, namely, F_5 and F_6.

Theorem 5: The equilibrium point $([S_1/S_1 + T_1 - 1], 0)$ is the stable state if $S_1 > 0, T_1 > 1$ and $U_1 = (A_3 - \beta)(T_1 - 1)^2 + 2S_1(A_2 - \beta)(T_1 - 1) + S_1^2(A_1 - \beta) < 0$.

Proof: The trace and determinant of equilibrium point $([S_1/S_1 + T_1 - 1], 0)$ are $([U_1 - S_1(T_1 - 1)(S_1 + T_1 - 1)]/[(S_1 + T_1 - 1)^2])$ and $-([(T_1 - 1)S_1U_1]/[(S_1 + T_1 - 1)^3])$. When $S_1 > 0, T_1 > 1$ and $U_1 < 0$, the trace and determinant satisfy $\mathrm{Tr} < 0$ and $\Delta > 0$. Thus, the equilibrium point $([S_1]/[S_1 + T_1 - 1]), 0)$ is stable. ■

We showcase that in the absence of intervention, cooperation is possible, but it must build upon $S_1 > 0$, $T_1 > 1$, and $U_1 > 0$. It means only when disputants play SDG, cooperation may exist in this system.

Theorem 6: The equilibrium point $([S_1 + \delta_1]/[S_1 + T_1 + \delta_1 + \delta_2 - 1], 1)$ is the stable state if $\delta_1 > -S_1, \delta_2 > 1 - T_1$ and $U_2 = (A_3 - \beta)(T_2 - 1)^2 + 2S_2(A_2 - \beta)(T_2 - 1) + S_2^2(A_1 - \beta) > 0$.

Proof: The trace and determinant of equilibrium point $([S_1 + \delta_1]/[S_1 + T_1 + \delta_1 + \delta_2 - 1], 1)$ are $-([U_2 + S_2(T_2 - 1)(S_2 + T_2 - 1)]/[(S_2 + T_2 - 1)^2])$ and $([S_2U_2(T_2 - 1)]/[(S_2 + T_2 - 1)^3])$. When $\delta_1 > -S_1, \delta_2 > 1 - T_1$, and $U_2 > 0$, the trace and determinant satisfy $\mathrm{Tr} < 0$ and $\Delta > 0$. Thus, the equilibrium point $([S_1 + \delta_1]/[S_1 + T_1 + \delta_1 + \delta_2 - 1], 1)$ is stable. ■

This theorem reveals that under the dominance of intervention, regardless of the basic game G_1, if $\delta_1 > -S_1, \delta_2 > 1 - T_1$, $U_2 > 0$ are satisfied, cooperation and defection coexist.

Theorem 7: The interior equilibrium point (x^*, ϕ^*) is the stable state if $V_1 = (\delta_1 + \delta_2)\phi^* + S_1 + T_1 - 1 > 0$ and $U_3 = [\delta_1 - (\delta_1 + \delta_2)x^*][2(A_1 - 2A_2 + A_3)x^* + 2(A_2 - A_3)] < 0$ with $0 < x^* < 1$ and $0 < \phi^* < 1$.

Proof: The interior equilibrium point (x^*, ϕ^*) has an associated Jacobian

$$J^* = \begin{bmatrix} x(1-x)\frac{\partial h_1(x,\phi)}{\partial x} & x(1-x)\frac{\partial h_1(x,\phi)}{\partial \phi} \\ \phi(1-\phi)\frac{\partial h_2(x)}{\partial x} & 0 \end{bmatrix}_{(x^*,\phi^*)} \tag{12}$$

where $h_1(x, \phi) = [(\delta_1 - \delta_2)\phi - S_1 - T_1 + 1]x + \delta_1\phi + S_1$ and $h_2(x) = (A_1 - 2A_2 + A_3)x^2 + 2(A_2 - A_3)x + A_3 - \beta$. As $0 < x^* < 1$ and $0 < \phi^* < 1$, it is easy to derive that the trace $\mathrm{Tr} < 0$ and $\Delta > 0$ if $V_1 > 0$ and $U_3 < 0$. ■

Thus, far, we have clarified the stability condition of each equilibrium point. In order to study the gain-and-loss scenario, we assume that the silence strategy relies on a fixed payoff $\beta = 2$. Subsequently, denote $A_1 = 4, A_2 = 1, A_3 = 0$ as CC-pair dominance intervention, where interveners receive the largest payoff A_1 from CC-pair. As shown in Fig. 3, compared with the silence strategy, the advantage of intervention changes with the frequency of cooperation (the outcome of the disputant layer). Similarly, denote CD-pair dominance intervention as $A_1 = 0, A_2 = 8, A_3 = 0$, and DD-pair dominance intervention as $A_1 = 0, A_2 = 1, A_3 = 4$. Since the valid parameter space of our model is $-1 \le S_1, S_2 \le 1$, $0 \le T_1, T_2 \le 2$, and $0 \le x, \phi \le 1$, we only discuss results that satisfy this space. Without a specific statement, we obtain the following results by fixing $T_1 = 1.1$ and $S_1 = -0.1$.

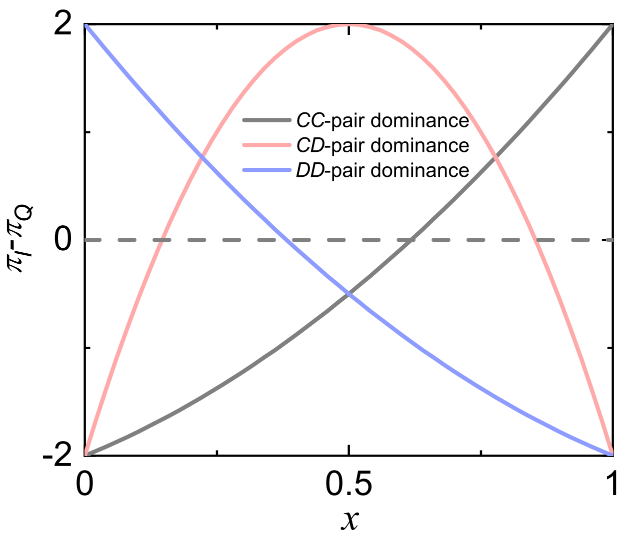


Fig. 3. Payoff difference of intervention and silence as a function of cooperation rate. For the *CC*-pair dominance pattern (black line), the payoff difference depends mainly on the fraction of cooperation. The larger the cooperation rate, the higher the payoff difference. *DD*-pair dominance pattern produces the opposite results (blue line). For the *CD*-pair dominance, the optimal payoff difference is obtained at $x = 0.5$.

1) CC-Pair Dominance Pattern: In the case of $A_1 = 4$, $A_2 = 1$, and $A_3 = 0$, (5) can be rewritten as

$$\begin{aligned} \pi_I &= 4x^2 + 2x(1-x) \\ \pi_Q &= \beta. \end{aligned} \tag{13}$$

There is only one interior equilibrium $F_7 = ([\sqrt{5} - 1/2], [S_1 - (S_1 + T_1 - 1)x_1]/[(\delta_1 + \delta_2)x_1 - \delta_1])$, where $x_1 = (\sqrt{5} - 1/2)$. Its stability condition can be obtained according to Theorem 7. Hereafter, in Fig. 4(a), we showcase a phase diagram as a function of (δ_1, δ_2) pair. Under these values, it is easy to deduce that equilibrium points $F2 = (1, 0)$ and $F_3 = (0, 1)$ are unstable. While equilibrium point $F_1 = (0, 0)$ is always stable. For $\delta_1 = \delta_2 = 0$, the disputant layer plays a PDG regardless of the intervention strategy, so $F_1 = (0, 0)$ is the unique asymptotically stable equilibrium.

When $\delta_1 < 0$ and $\delta_2 > 0$, the intervention behaves as TM type, amplifying the dilemma strength in the disputant layer. Consequently, in the upper left part of the diagram, there are still only two available strategies: 1) defection for disputants and 2) silence for third parties. When $\delta_1 > 0$ and $\delta_2 < 0$, intervention manifests as PM type, weakening the dilemma strength of disputant layer. Particularly in $\delta_2 < -0.1$ domain, co-dominance of C and I emerge. Therefore, we showcase a bistable area that contains two stable equilibrium points. As shown in Fig. 4(b), which equilibrium point the system falls is closely related to the initial value of C and I. It is worth noting that the interior fix point here is a saddle whose eigenvalues of characteristic function are real roots with opposite signs. From a geometric perspective, there exist orbits approaching and moving away from the saddle point simultaneously. Furthermore, a small region containing equilibrium points F_1 and F_6 is also triggered by PM type intervention. The result states that compared with TM type, PM type intervention is particularly good at stimulating cooperation.

Turning Attention to MIX Type: When $\delta_1 < 0$ and $\delta_2 < 0$ (lower left region), the game played by disputants under intervention shifts toward the SHG. Since the Nash equilibria of SHG are (C, C) and (D, D), when $\delta_2 < -0.1$, the system undoubtedly enters the bistable state. When $\delta_1 > 0$ and $\delta_2 > 0$ (upper right region), the game played by disputants under intervention shifts toward the SDG. Cooperation can be maintained only when $\delta_1 > 0.1$ and $U_2 > 0$, i.e., F_1F_6 bistable region. Otherwise, point F_1 will be the unique asymptotically stable equilibrium.

Furthermore, through fixing $\delta_1 = 0.8$, the results in Fig. 4(c) show that intervention with *CC*-pair dominance pattern keeps the same evolutionary orientation with cooperation. There exist bistable states of F_1F_4 and F_1F_6, and a monostable state of F_1. In particular, discontinuous and continuous phase transitions emerge, respectively, in disputant and third-party layers as δ_2 increases. The results so far demonstrate that cooperation is promoted when intervention emerges in the third-party layer with a larger frequency. It is natural to ask whether a minority of interventions can stimulate significant increases in cooperation. We will address this doubt in the following parts.

2) CD-Pair Dominance Pattern: In the case of $A_1 = 0$, $A_2 = 8$, and $A_3 = 0$, (5) can be rewritten as

$$\begin{aligned} \pi_I &= 16x(1-x) \\ \pi_Q &= \beta. \end{aligned} \tag{14}$$

There are two interior equilibrium points $F_7 = ([2 + \sqrt{2}/4], [S_1 - (S_1 + T_1 - 1)x_2]/[(\delta_1 + \delta_2)x_2 - \delta_1])$ and $F_8 = ([2 - \sqrt{2}/4], [S_1 - (S_1 + T_1 - 1)x_3]/[(\delta_1 + \delta_2)x_3 - \delta_1])$, where $x_2 = (2 + \sqrt{2}/4)$ and $x_3 = (2 - \sqrt{2}/4)$. We can get their stability conditions according to Theorem 7. Fig. 5(a) reveals the distribution of asymptotically stable equilibrium points in δ_1-δ_2 parameter space. As given by Theorems 1–4, $F_1 = (0, 0)$ is always stable, while equilibrium points F_2, F_3, and F_4 are unstable in these parameter settings. Furthermore, we also find regions where equilibrium points F_6 and F_7 are stable. In detail, when intervention behaves as PM type, there is a bistable region that satisfies the stability condition for point F_7. This indicates partial intervention can promote cooperation in the disputant layer. Here, equilibria are sensitive to the initial frequency of C and I. As shown in Fig. 5(b), the system falls into equilibrium point F_1 when the initial frequencies of the cooperation and intervention are low, or F_7 when the initial conditions are applicable. The result reveals that PM type intervention can still promote cooperation under *CD*-pair dominance pattern. Another cooperation existence area is MIX type (upper right corner), i.e., the effect of intervention transforms the PDG into an SDG. Since intervention benefits more from the coexistence of cooperation and defection, F_6 becomes stable in a large proportion of this region.

Then, we showcase that a minority of interventions can stimulate a majority of cooperation in *CD*-pair dominance pattern. As shown in Fig. 5(c), cooperation in the disputant layer is higher than intervention when $\delta_2 \leq 0.02$. It answers our concerns about whether cooperation can be triggered by a small fraction of intervention. Furthermore, intervention can

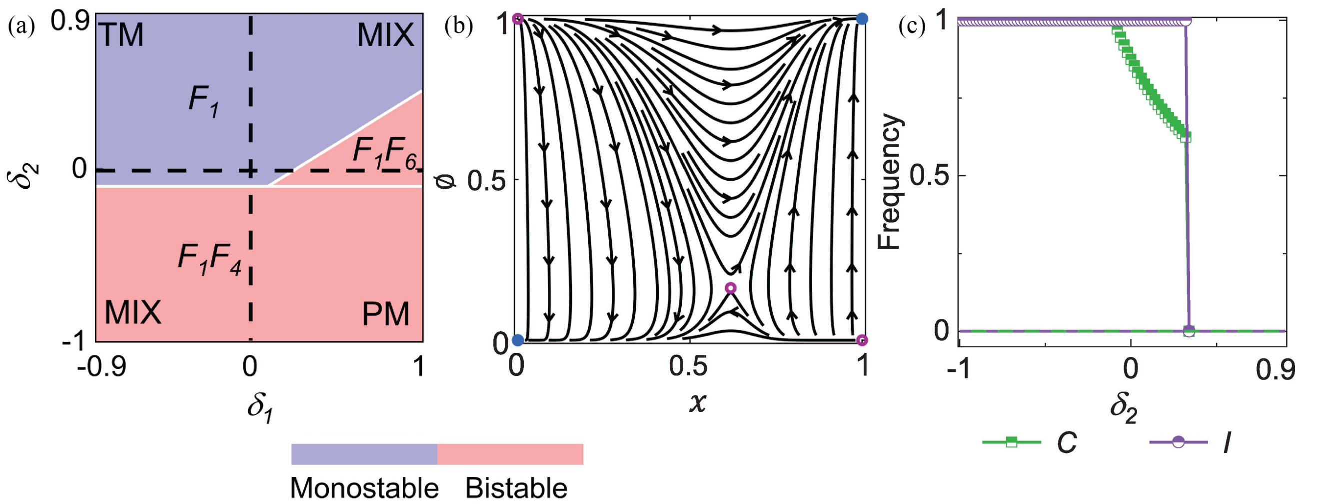


Fig. 4. Phase diagram of cooperation and intervention in δ_1–δ_2 space under *CC*-pair dominance pattern. (a) Compared with TM type, PM type intervention is more conducive to the evolution of cooperation. For MIX type, when the game is transferred to an SHG under the intervention, the conflict layer is likely to change to a fully cooperative state. When the intervention effect is an SDG, the conflict layer can maintain a state of coexistence of cooperation and defection. (b) Within a bistable region, the initial frequency of cooperation and intervention plays a key role in the equilibrium that the system reaches. Solid and open dots represent stable and other fixed points, respectively. (c) Intervention keeps the same evolutionary orientation with cooperation. Parameters are fixed to $\delta_1 = 0.8$ and $\delta_2 = -0.5$ in panel (b), and $\delta_1 = 0.8$ in panel (c).

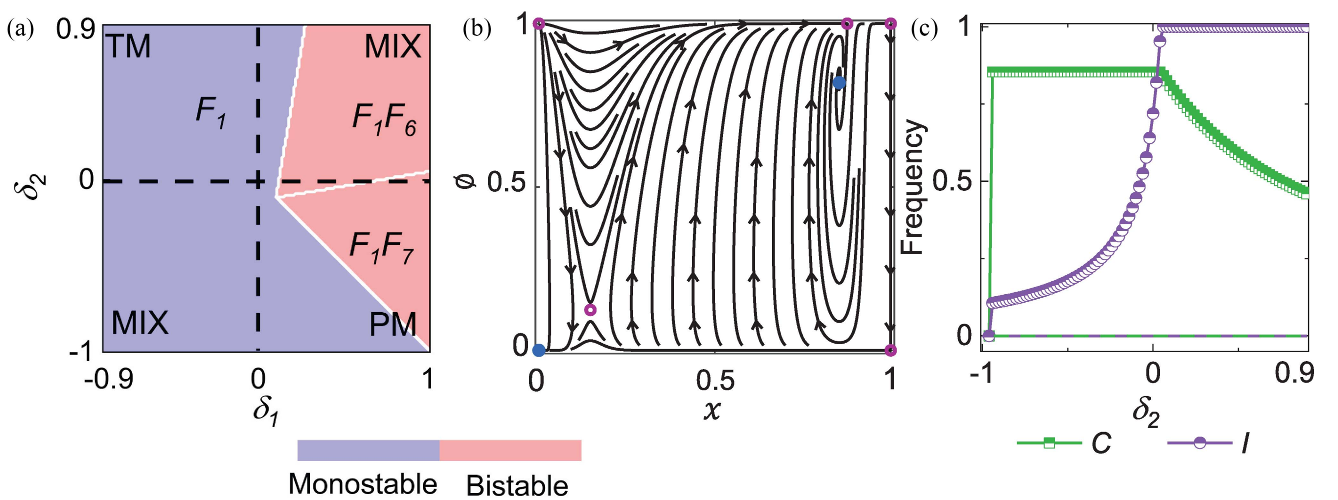


Fig. 5. Phase diagram of cooperation and intervention in δ_1–δ_2 space under *CD*-pair dominance pattern. (a) Cooperation is promoted when intervention behaves as PM and MIX types (right top corner). (b) Within a bistable region, which equilibrium the system falls depends on the initial frequency of cooperation and intervention. (c) Intervention dominates the third-party layer when cooperation and defection are sufficiently mixed. Parameters are fixed to $\delta_1 = 0.95$ and $\delta_2 = 0.02$ in panel (b), and $\delta_1 = 0.95$ in panel (c).

easily become dominant when cooperation and defection are sufficiently mixed if intervention receives more payoff from the *CD*-pair.

3) DD-Pair Dominance Pattern: In the case of $A_1 = 0$, $A_2 = 1$, and $A_3 = 4$, (5) can be rewritten as

$$\pi_I = 2(1-x)(2-x)$$
$$\pi_Q = \beta. \tag{15}$$

Subsequently, we can derive an interior equilibrium point $F_7 = ([3-\sqrt{5}/2], [S_1 - (S_1 + T_1 - 1)x_4]/[(\delta_1+\delta_2)x_4 - \delta_1])$, where $x_4 = (3-\sqrt{5}/2)$. Its stability condition is obtained according to Theorem 7. Given $\delta_2 = 0.5$, we show how asymptotically stable equilibria change with δ_1 in Fig. 6(a). With the increase of δ_1, the asymptotically stable equilibrium moves from F_3 to F_6 and finally to F_7. It reveals that cooperation and intervention remain in opposite evolutionary orientations, i.e., cooperation (intervention) is promoted (prohibited) as the increase of δ_1. In particular, a minority of interventions can stimulate a higher level of cooperation with suitable δ_1. In a monostable state, the equilibrium is insensitive to the initial values [see Fig. 6(b)]. This is further evidenced by assigning the initial values of pair (x, ϕ) as (0.1, 0.1), (0.2, 0.2), $\cdots$, (0.9, 0.9) in Fig. 6(c). After finite steps, the evolution of cooperation (top) and intervention (bottom) eventually reaches a unique stable state.

V. Extension to Square Lattices

Having seen the highly nontrivial interplay of intervention and cooperation in replicator dynamics (RDs), in this section, we will consider the effect of relaxing the infinitely large well-mixed hypothesis by allowing finitely large populations. In particular, unlike interactions between all actors (well-mixed populations), structures with local interactions

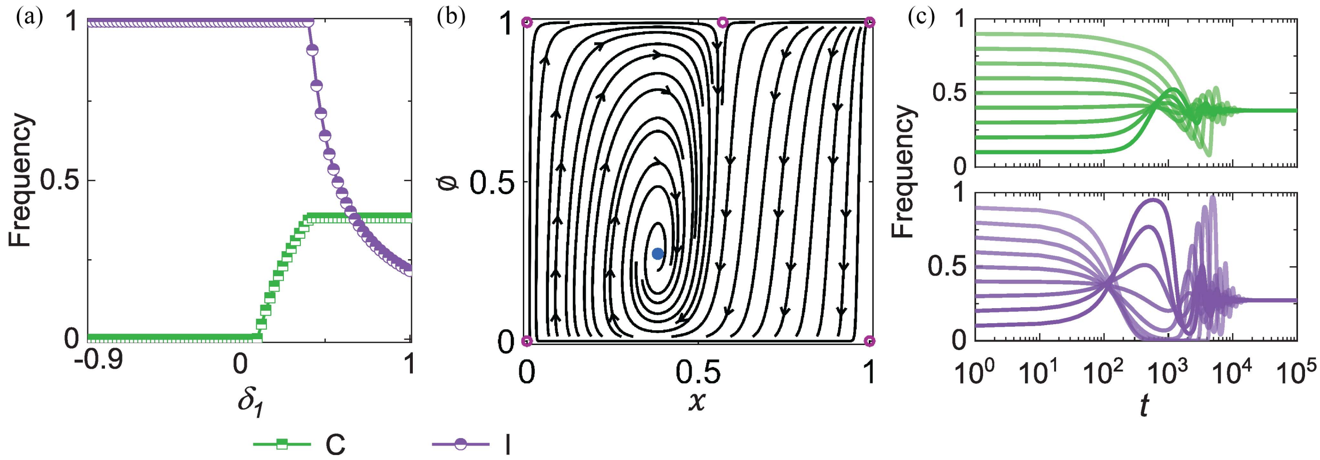


Fig. 6. Co-evolution of cooperation and intervention in δ_1–δ_2 space under *DD*-pair dominance pattern. (a) Intervention and cooperation maintain opposite evolutionary orientations. (b) Within a monostable region, the interior equilibrium is globally stable regardless of the initial conditions. (c) Equilibria of cooperation (top) and intervention (bottom) are insensitive to initial values. The parameter is fixed to $\delta_1 = 0.9$ in panels (b) and (c).

also play a crucial role in strategic conflict [8]. In doing so, there is a widely used updating rule in the literature. That is the Fermi rule [47], where each player imitates one of their opponent's strategies with a probability given by the Fermi function. Note that RD is determinate, and the variation of the population is linear in the payoff difference. Similar to RD, the Fermi function is also a function of payoff difference. The strategies of players with high payoffs are more likely to spread. Unlike RD, the Fermi rule can study the effects of temperature or selection intensity.

A. Agent-Based Model

Each layer adopts a square lattice with periodic boundaries and von Newmann neighbors [47]. For the disputant layer, the strategy of player $i^{\mathcal{D}}$ is represented by $\mathcal{S}_{i\mathcal{D}} = (1, 0)^T$ for cooperation, $\mathcal{S}_{i\mathcal{D}} = (0, 1)^T$ for defection. For the third-party layer, the strategy of player $i^{\mathcal{T}}$ is represented by $\mathcal{A}_{i\mathcal{T}} = (1, 0)^T$ for intervention, $\mathcal{A}_{i\mathcal{T}} = (0, 1)^T$ for silence. Interacting with all neighbors, each player $i^{\mathcal{D}}$ in the disputant layer receives a payoff as follows:

$$P_{i\mathcal{D}} = \alpha \sum_{y^{\mathcal{D}} \in \Omega_{i\mathcal{D}}} \mathcal{S}_{i\mathcal{D}} M_1 \mathcal{S}_{y\mathcal{D}} + (1-\alpha) \sum_{y^{\mathcal{D}} \in \Omega_{i\mathcal{D}}} \mathcal{S}_{i\mathcal{D}} M_2 \mathcal{S}_{y\mathcal{D}} \quad (16)$$

where $\Omega_{i\mathcal{D}}$ represents the neighbor set of player $i^{\mathcal{D}}$. $\alpha = 0$ if player $i^{\mathcal{D}}$ corresponds to an intervener ($\mathcal{A}_{i\mathcal{T}} = (1, 0)^T$), and $\alpha = 1$ otherwise. For the third-party layer, player $i^{\mathcal{T}}$ with silence strategy receives a fixed payoff $k_i^{\mathcal{D}}\beta$. The payoff of player $i^{\mathcal{T}}$ with intervention strategy is determined by the number of *CC*-pair (N_{CC}), *CD*-pair or DC-pair (N_{CD}), *DD*-pair (N_{DD}) between $i^{\mathcal{D}}$ and its neighbors in the disputant layer. As depicted in Section IV, each *CC*-, *CD*-, and *DD*-pair brings payoff A_1, A_2, and A_3 to intervention, respectively. Subsequently, the cumulative payoff is given as follows:

$$P_{i\mathcal{T}} = \begin{cases} A_1 N_{CC} + A_2 N_{CD} + A_3 N_{DD}, & \text{if } \mathcal{A}_{i\mathcal{T}} = I \\ k_{i\mathcal{D}}\beta, & \text{if } \mathcal{A}_{i\mathcal{T}} = Q. \end{cases} \quad (17)$$

Evolutionary games are depicted by MCS, including the following steps (see Algorithm 1): a randomly selected player i acquires payoff via (16) if it belongs to disputant layer, or via (17) if it belongs to third-party layer. Then, we randomly select one of i's neighbors, say j, and get its payoff in a similar way. Finally, i imitates j's strategy with a probability determined by the Fermi function

$$W_{\mathcal{S}_i \leftarrow \mathcal{S}_j}(P_i, P_j) = \frac{1}{1 + e^{-(P_j - P_i)/K}} \quad (18)$$

where K^{-1} represents the intensity of selection. Since it has been well studied [47], we parameterize it as 0.1. To ensure

---

Algorithm 1: Evolutionary Games With Third-Party Intervention

```
Input: the payoff matrix, the step of MCS Λ
for each i on the square lattice do
    if i ∈ 𝒟 then
        Initialize player i with a strategy from set 𝒮 randomly;
    else
        Initialize player i with a strategy from set 𝒜 randomly;
    end
end
t ← 1;
while t < Λ do
    m ← 1;
    while m < L × L do
        Select a layer u ∈ {𝒟, 𝒯} randomly;
        Select a player i^u from layer u and one of its neighbor
        y^u randomly;
        Calculate the payoff of i^u and y^u by Eq. 16 if u = 𝒟,
        and by Eq. 17 otherwise;
        Player i^u imitates the strategy of y^u with a probability
        given by Eq. 18;
        Select a player i^v from layer v (v is different from u)
        and one of its neighbor y^v randomly;
        Calculate the payoff of i^v and y^v by Eq. 17 if v = 𝒯,
        and by Eq. 16 otherwise;
        Player i^v imitates the strategy of y^v with a probability
        given by Eq. 18;
        m ← m + 1;
    end
    t ← t + 1;
end
```

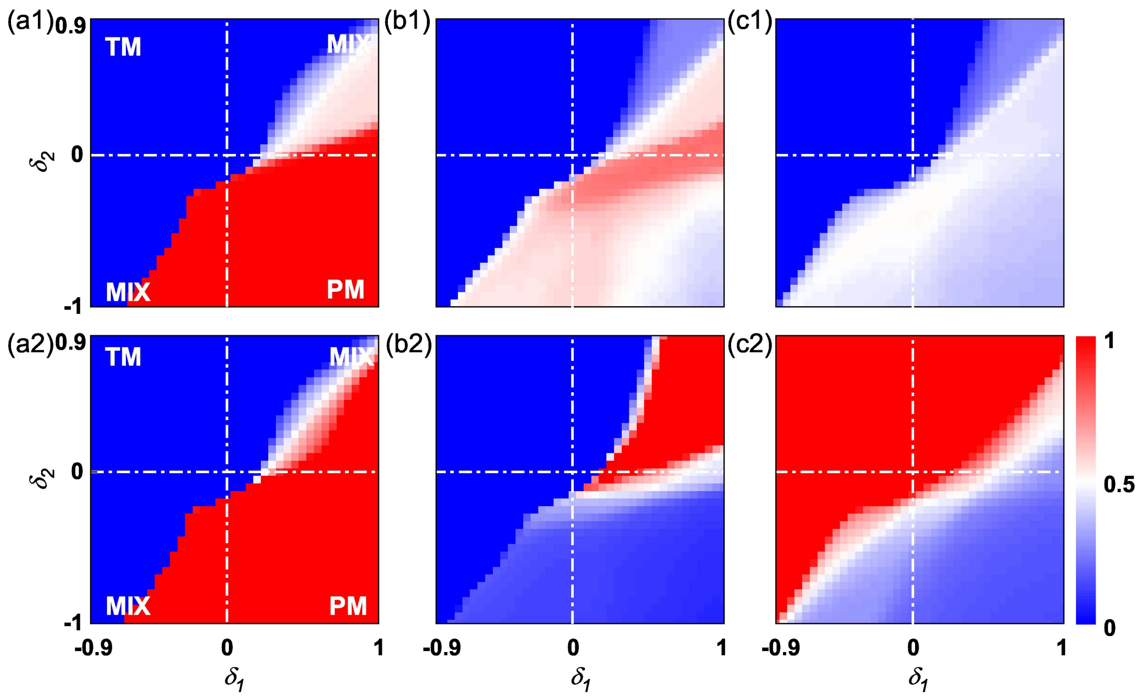


Fig. 7. Cooperation under PM type intervention is more prosperous than that under a TM type. Changing the IPP of intervention can effectively control the equilibrium of this coupled system. From the first to the third column, it shows the frequency of cooperation (top row) and intervention (bottom row) under *CC*-pair, *CD*-pair, and *DD*-pair dominance patterns, respectively. (a1) and (a2) Cooperation and intervention keep the same evolution orientation. (b1) Cooperation reaches optimal value even if only a mere fraction of intervention in the third-party layer. (b2) Intervention dominates the third-party layer when cooperation mixes sufficiently with defection. (c1) Cooperation and intervention evolve in opposite directions. In particular, regardless of the frequency of intervention, cooperation cannot dominate disputant layer. (c2) Intervention dominates the third-party layer when defection prevails. The color code represents the frequency of cooperation and intervention. Parameters are obtained as $T_1 = 1.1$, $S_1 = -0.1$, and $K = 0.1$.

the accuracy of the results, we calculate the average frequency of each strategy over 3000 MCS steps after entering a convergent state. Without the specific declaration, the square lattice consists of 200×200 players. Furthermore, to avoid finite size effects, we test scales of 100×100 and 300×300, with almost identical results.

B. Phase Diagram

So far, we have revealed the evolutionary dynamics in well-mixed populations with intervention by third parties through MF theory. We are now attempting to explain how spatial structure affects the coupling between cooperation and intervention. To have a comprehensive overview, we provide the phase diagrams of C and I in δ_1–δ_2 space (see Fig. 7). Consistent with well-mixed populations, the parameter space where cooperation thrives varies with the IPP of intervention. Without the supervision of interveners, players in the disputant layer participate in PDG with $T_1 = 1.1, S_1 = -0.1$. Since PDG has been well studied on the square lattice, there is no doubt that cooperation disappears under these parameters. However, this situation will be changed if we consider third-party intervention. It is worth noting that interactions in the disputant layer entirely follow matrix M_2 if intervention dominates the third-party layer, whereas it follows matrix M_1 if the third-party layer evolves into a full Q state. When the intervening IPP is *CC*-pair dominance, PM type is more likely to stimulate cooperation than TM type [Fig. 7(a1)]. Turning our attention to MIX type, there is a discontinuous phase transition in the lower left of panels (a1) and (a2), but a continuous phase transition in the upper right corners. Since intervention gains from *CC*-pair but loses from others, intervention keeps the same evolutionary orientation as cooperation, i.e., when cooperation thrives, intervention thrives; when cooperation declines, intervention declines. Following the *CD*-pair dominance pattern, interveners spring up as *CD*-pair increases. Therefore, intervention dominates the third-party layer if cooperation and defection are mixed sufficiently [see Fig. 7(b1) and (b2)]. When the IPP is *DD*-pair dominance, selecting intervention is better if there exists more *DD*-pair in the disputant layer [see Fig. 7(c1) and (c2)], revealing a completely opposite evolutionary orientation of C and I.

VI. Conclusion and Discussion

In this article, we develop a novel framework to address the co-evolution of cooperation and third-party intervention. Although evidence has proven that third parties play an inevitable role in the emergence and maintenance of cooperative behavior [48], [49], they have not addressed the emergence of intervention. Different from these studies that consider only one population, we model the interplay between human conflicts and third parties by a coupled system, including disputant and third-party layers. Another difference is that the intervention in this article is risky rather than considering intervening in a cost-effective way [36], [40]. We showcase seven theorems and implement three special cases by considering gain-and-loss forms, including *CC*-, *CD*-, and *DD*-pair dominance patterns. Furthermore, according to the utility on the dilemma strength between disputants, we propose three types of interveners: 1) peacemakers; 2) troublemakers; and 3) mixers. Instead of choosing to intervene, players in the third-party layer can also keep silent to avoid the risk of loss. Through the analysis of coupled replicator equations, we show that peacemakers are particularly effective at promoting cooperation. Interestingly, a mere fraction of intervention can stimulate higher cooperation in *CD*- and *DD*-pair dominance patterns. On the other hand, complete cooperation is not necessary for complete intervention. Moreover, we find monostable states such as co-extinction, co-dominance, and co-existence of cooperation and intervention, as well as bi-stable states. Then, by developing an evolutionary algorithm in large-scale square lattices, we reproduce the co-extinction, co-dominance, and co-existence of cooperation and intervention. Our research revealed the condition under which intervention emerges and how intervention controls the equilibria in the conflict. Similar to environment feedback [46], [50], feedback between third-party and player's strategies provides the potential for studying the linkage between exogenous intervention and human behavior. Without intervention, cooperators in conflict with defectors are less likely to win the game with a larger dilemma strength. However, strong positive intervention (especially peacemaker) enables cooperation to dominate defection. This unveils the potential of third parties to control the evolution of cooperation.

Under this framework, several attractive avenues for future work still exist. One of the most concerning directions is evaluating the cost efficiency of this kind of risk-bearing intervention. Previous studies have proposed a promising framework for solving the cost-efficiency problem [36], [40]. To do so, we must formulate a scheme that takes into account both the cost of intervention and the benefit of this system, including the increased cooperation rate in the disputant layer and the increased payoff in the third-party layer. Moreover, it is crucial to consider implementing this scheme with a limited number of interveners, rather than relying on global intervention. On the other hand, to evaluate the utility of intervention, we need to consider samaritan interveners who do not change their behavior over time [51]. One possible way is to assign part of third parties as permanent interveners permanently. In light of this, it is natural to expect whether the location of a samaritan intervener (how many cooperators it corresponds to) has a significant impact on the evolution of cooperation [37]. Furthermore, with the framework proposed in this article, there is still room for improvement, such as the time-delay effect [52] in well-mixed populations and time scale [53], [54] in the different layers. Incorporating control theory into cooperative systems is also an exciting research direction [38], [55]. Although cooperation can be effectively promoted by third-party intervention, defection still dominates the network under sufficiently strong temptation. In order to investigate how to facilitate large-scale cooperation, we need to relax the hypotheses further and even organize human experiments.

Acknowledgment

The authors grateful to Shupeng Gao for useful discussions.

References

[1] R. S. Burt and M. Knez, "Kinds of third-party effects on trust," *Rational. Soc.*, vol. 7, no. 3, pp. 255–292, 1995.
[2] A. Rapoport and A. M. Chammah, *Prisoner's Dilemma: A Study in Conflict and Cooperation*, Ann Arbor, MI, USA: Univ. Michigan Press, 1965.
[3] A. Rapoport, *Game Theory as a Theory of Conflict Resolution*. Dordrecht, The Netherlands: D. Reidel, 1974.
[4] S. A. West, A. S. Griffin, and A. Gardner, "Social semantics: Altruism, cooperation, mutualism, strong reciprocity and group selection," *J. Evol. Biol.*, vol. 20, no. 2, pp. 415–432, 2007.
[5] J. Farrell and R. Ware, "Evolutionary stability in the repeated prisoner's dilemma," *Theor. Popul. Biol.*, vol. 36, no. 2, pp. 161–166, 1989.
[6] R. Cressman, "Evolutionary stability in the finitely repeated prisoner's dilemma game," *J. Econ. Theory*, vol. 68, no. 1, pp. 234–248, 1996.
[7] J. Hofbauer and K. Sigmund, *Evolutionary Games and Population Dynamics*. Cambridge, U.K.: Cambridge Univ. Press, 1998.
[8] M. A. Nowak and R. M. May, "Evolutionary games and spatial chaos," *Nature*, vol. 359, no. 6398, pp. 826–829, 1992.
[9] G. Szabó and G. Fath, "Evolutionary games on graphs," *Phys. Rep.*, vol. 446, nos. 4–6, pp. 97–216, 2007.
[10] M. Doebeli and C. Hauert, "Models of cooperation based on the prisoner's dilemma and the snowdrift game," *Ecol. Lett.*, vol. 8, no. 7, pp. 748–766, 2005.
[11] J. M. Pacheco, F. C. Santos, M. O. Souza, and B. Skyrms, "Evolutionary dynamics of collective action in N-person stag hunt dilemmas," *Proc. Royal Soc. B*, vol. 276, no. 1655, pp. 315–321, 2009.
[12] Y. Yang and X. Li, "Towards a snowdrift game optimization to vertex cover of networks," *IEEE Trans. Cybern.*, vol. 43, no. 3, pp. 948–956, Jun. 2013.
[13] C. Tang, A. Li, and X. Li, "When reputation enforces evolutionary cooperation in unreliable MANETs," *IEEE Trans. Cybern.*, vol. 45, no. 10, pp. 2190–2201, Oct. 2015.
[14] Z. Wang, M. Jusup, L. Shi, J.-H. Lee, Y. Iwasa, and S. Boccaletti, "Exploiting a cognitive bias promotes cooperation in social dilemma experiments," *Nat. Commun.*, vol. 9, no. 1, pp. 1–7, 2018.
[15] A. Dreber, D. G. Rand, D. Fudenberg, and M. A. Nowak, "Winners don't punish," *Nature*, vol. 452, no. 7185, pp. 348–351, 2008.
[16] Y. Jiao, T. Chen, and Q. Chen, "Probabilistic punishment and reward under rule of trust-based decision-making in continuous public goods game," *J. Theor. Biol.*, vol. 486, Feb. 2020, Art. no. 110103.
[17] A. Szolnoki and M. Perc, "Second-order free-riding on antisocial punishment restores the effectiveness of prosocial punishment," *Phys. Rev. X*, vol. 7, no. 4, 2017, Art. no. 41027.
[18] S. Gao, J. Du, and J. Liang, "Evolution of cooperation under punishment," *Phys. Rev. E*, vol. 101, no. 6, 2020, Art. no. 62419.
[19] W. Lu, J. Wang, and C. Xia, "Role of memory effect in the evolution of cooperation based on spatial prisoner's dilemma game," *Phys. Lett. A*, vol. 382, nos. 42–43, pp. 3058–3063, 2018.
[20] B. Wu, H. J. Park, L. Wu, and D. Zhou, "Evolution of cooperation driven by self-recommendation," *Phys. Rev. E*, vol. 100, no. 4, 2019, Art. no. 42303.
[21] J. Du and Z. Wu, "Evolutionary dynamics of cooperation in dynamic networked systems with active striving mechanism," *Appl. Math. Comput.*, vol. 430, Oct. 2022, Art. no. 127295.
[22] J. Du and B. Wang, "Evolution of global cooperation in multi-level threshold public goods games with income redistribution," *Front. Phys.*, vol. 6, p. 67, Jul. 2018.
[23] J. Du, "Redistribution promotes cooperation in spatial public goods games under aspiration dynamics," *Appl. Math. Comput.*, vol. 363, Dec. 2019, Art. no. 124629.
[24] J. Gross and C. K. De Dreu, "The rise and fall of cooperation through reputation and group polarization," *Nat. Commun.*, vol. 10, no. 1, pp. 1–10, 2019.
[25] S. Assenza, J. Gómez-Gardeñes, and V. Latora, "Enhancement of cooperation in highly clustered scale-free networks," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 78, no. 1, 2008, Art. no. 17101.
[26] X. Chen, Å. Brännström, and U. Dieckmann, "Parent-preferred dispersal promotes cooperation in structured populations," *Proc. Royal Soc. B*, vol. 286, no. 1895, 2019, Art. no. 20181949.

[27] S. Tan, Y. Wang, and A. V. Vasilakos, "Distributed population dynamics for searching generalized Nash equilibria of population games with graphical strategy interactions," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 52, no. 5, pp. 3263–3272, May 2022.
[28] S. Tan, Z. Fang, Y. Wang, and J. Lü, "Consensus-based multipopulation game dynamics for distributed Nash equilibria seeking and optimization," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 53, no. 2, pp. 813–823, Feb. 2023.
[29] J. Zhang, Y. Zhu, and Z. Chen, "Evolutionary game dynamics of multiagent systems on multiple community networks," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 50, no. 11, pp. 4513–4529, Nov. 2020.
[30] Z. Wang, A. Szolnoki, and M. Perc, "Optimal interdependence between networks for the evolution of cooperation," *Sci. Rep.*, vol. 3, no. 1, pp. 1–7, 2013.
[31] Z. Wang, L. Wang, and M. Perc, "Degree mixing in multilayer networks impedes the evolution of cooperation," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 89, no. 5, 2014, Art. no. 52813.
[32] A. Hafezalkotob, "Direct and indirect intervention schemas of government in the competition between green and non-green supply chains," *J. Clean. Prod.*, vol. 170, pp. 753–772, Jan. 2018.
[33] N. A. Nakashima, E. Halali, and N. Halevy, "Third parties promote cooperative norms in repeated interactions," *J. Exp. Soc. Psychol.*, vol. 68, pp. 212–223, Jan. 2017.
[34] P. Lergetporer, S. Angerer, D. Glätzle-Rützler, and M. Sutter, "Third-party punishment increases cooperation in children through (misaligned) expectations and conditional cooperation," *Proc. Nat. Acad. Sci. U S A*, vol. 111, no. 19, pp. 6916–6921, 2014.
[35] N. Nikiforakis and H. Mitchell, "Mixing the carrots with the sticks: Third party punishment and reward," *Exp. Econ.*, vol. 17, no. 1, pp. 1–23, 2014.
[36] T. A. Han and L. Tran-Thanh, "Cost-effective external interference for promoting the evolution of cooperation," *Sci. Rep.*, vol. 8, no. 1, 2018, Art. no. 15997.
[37] T. Cimpeanu, C. Perret, and T. A. Han, "Cost-efficient interventions for promoting fairness in the ultimatum game," *Knowl. Based Syst.*, vol. 233, Dec. 2021, Art. no. 107545.
[38] S. Wang, X. Chen, and A. Szolnoki, "Exploring optimal institutional incentives for public cooperation," *Commun. Nonlinear Sci. Numer. Simulat.*, vol. 79, Dec. 2019, Art. no. 104914.
[39] N. Halevy and E. Halali, "Selfish third parties act as peacemakers by transforming conflicts and promoting cooperation," *Proc. Nat. Acad. Sci. U S A*, vol. 112, no. 22, pp. 6937–6942, 2015.
[40] M. H. Duong and T. A. Han, "Cost efficiency of institutional incentives for promoting cooperation in finite populations," *Proc. Royal Soc. A*, vol. 477, no. 2254, 2021, Art. no. 20210568.
[41] Z. Wang, S. Kokubo, M. Jusup, and J. Tanimoto, "Universal scaling for the dilemma strength in evolutionary games," *Phys. Life Rev.*, vol. 14, pp. 1–30, Sep. 2015.
[42] Z. Wang, L. Wang, A. Szolnoki, and M. Perc, "Evolutionary games on multilayer networks: A colloquium," *Eur. Phys. J. B*, vol. 88, pp. 1–15, May 2015.
[43] C. Hauert, S. De Monte, J. Hofbauer, and K. Sigmund, "Replicator dynamics for optional public good games," *J. Theor. Biol.*, vol. 218, no. 2, pp. 187–194, 2002.
[44] A. Antonioni, L. A. Martinez-Vaquero, C. Mathis, L. Peel, and M. Stella, "Individual perception dynamics in drunk games," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 99, no. 5, 2019, Art. no. 52311.
[45] S. Tan, Y. Wang, Y. Chen, and Z. Wang, "Evolutionary dynamics of collective behavior selection and drift: Flocking, collapse, and oscillation," *IEEE Trans. Cybern.*, vol. 47, no. 7, pp. 1694–1705, Jul. 2017.
[46] J. S. Weitz, C. Eksin, K. Paarporn, S. P. Brown, and W. C. Ratcliff, "An oscillating tragedy of the commons in replicator dynamics with game-environment feedback," *Proc. Nat. Acad. Sci. U S A*, vol. 113, no. 47, pp. E7518–E7525, 2016.
[47] C. Hauert and G. Szabó, "Game theory and physics," *Amer. J. Phys.*, vol. 73, no. 5, pp. 405–414, 2005.
[48] M. Fiedler and E. Haruvy, "The effect of third party intervention in the trust game," *J. Behav. Exp. Econ.*, vol. 67, pp. 65–74, Apr. 2017.
[49] B. R. House et al., "Social norms and cultural diversity in the development of third-party punishment," *Proc. Royal Soc. B*, vol. 287, no. 1925, 2020, Art. no. 20192794.
[50] A. Szolnoki and X. Chen, "Environmental feedback drives cooperation in spatial social dilemmas," *Europhys. Lett.*, vol. 120, no. 5, 2018, Art. no. 58001.
[51] A. Kumar, V. Capraro, and M. Perc, "The evolution of trust and trustworthiness," *J. R. Soc. Interface*, vol. 17, no. 169, 2020, Art. no. 20200491.
[52] F. Yan, X. Chen, Z. Qiu, and A. Szolnoki, "Cooperator driven oscillation in a time-delayed feedback-evolving game," *New J. Phys.*, vol. 23, no. 5, 2021, Art. no. 53017.
[53] X. Xu, Z. Rong, Z. Tian, and Z.-X. Wu, "Timescale diversity facilitates the emergence of cooperation-extortion alliances in networked systems," *Neurocomputing*, vol. 350, pp. 195–201, Jul. 2019.
[54] F. L. Pinheiro, J. M. Pacheco, and F. C. Santos, "Stable leaders pave the way for cooperation under time-dependent exploration rates," *Roy. Soc. Open Sci.*, vol. 8, no. 2, 2021, Art. no. 200910.
[55] H. Tembine, E. Altman, R. El-Azouzi, and Y. Hayel, "Evolutionary games in wireless networks," *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 40, no. 3, pp. 634–646, Jun. 2010.

Hao Guo received the Ph.D. degree in industrial engineering from Northwestern Polytechnical University, Xi'an, China, in 2022.

His research interests include evolutionary game theory and multiagent systems.

Zhao Song received the B.S. degree in automation from Northwestern Polytechnical University, Xi'an, China, in 2018, where she is currently pursuing the Ph.D. degree in industrial engineering with the School of Mechanical Engineering.

Her current research interests include game theory and multiagent systems.

Matjaž Perc (Member, IEEE) received the Ph.D. degree from the University of Maribor, Maribor, Slovenia.

He is currently a Professor of Physics with the University of Maribor.

Prof. Perc is also the 2015 recipient of the Young Scientist Award for Socio and Econophysics from the German Physical Society, and the 2017 USERN Laureate. In 2018, he received the Zois Award, which is the highest national research award in Slovenia. Since 2021, he has been the Vice Dean of Natural Sciences at the European Academy of Sciences and Arts. He is a member of Academia Europaea and the European Academy of Sciences and Arts, and among top 1% most cited physicists according to 2020, 2021, and 2022 Clarivate Analytics data. In 2019, he became a Fellow of the American Physical Society.

Xuelong Li (Fellow, IEEE) is currently a Full Professor with the School of Computer Science and Artificial Intelligence, Optics and Electronics, Northwestern Polytechnical University, Xi'an, China.

Zhen Wang (Senior Member, IEEE) received the Ph.D. degree from Hong Kong Baptist University, Hong Kong, in 2014.

He is currently a Full Professor with Northwestern Polytechnical University, Xi'an, China. His current research interests include artificial intelligence, network science, data mining, and multiagent learning games.

Dr. Wang is a member of Academia Europaea and the European Academy of Sciences and Arts.