

Bifurcation and Multistability in a Collective-Risk Public Goods Game with Quorum-Activated Protection

Shulan Li^a, Xiaogang Li^b, Lei Shi^{b,*}, Jun He^{c,*} and Attila Szolnoki^d

^a*School of Accounting and Auditing, Yunnan University of Finance and Economics, Kunming, 650221, China*

^b*School of Statistics and Mathematics, Yunnan University of Finance and Economics, Kunming, 650221, China*

^c*School of Engineering, Kunming University, Kunming, 650214, China*

^d*Institute of Technical Physics and Materials Science, Centre for Energy Research, Budapest, 1525, Hungary*

ARTICLE INFO

Keywords:

public goods game
collective risk
quorum-activated protection
replicator dynamics
saddle-node bifurcation

ABSTRACT

redCollective-failure risk can promote cooperation when cooperative behavior reduces the probability of a shared loss. Existing studies usually embed risk mitigation in the contribution itself, whereas many real systems rely on a separate costly and non-excludable protective activity that requires a participation quorum. Whether such protection can evolve and persist remains unclear. Here, we formulate an N -player public-goods game with ordinary cooperators, defectors, and protective cooperators. Defectors endogenously increase a saturating probability of collective failure, whereas protective cooperators incur an additional cost and reduce this risk for all group members once their number reaches the quorum. Replicator analysis and numerical bifurcation calculations yield three main findings. First, protective cooperation gains an advantage when the focal participant completes the quorum and the resulting expected loss reduction exceeds its additional cost. This advantage varies non-monotonically with group composition. Second, when a single protective cooperator cannot activate protection, protection effectiveness does not enter its first-order invasion fitness at full defection; greater effectiveness cannot overcome the rare-invasion barrier when full defection is locally asymptotically stable. Third, a boundary saddle-node bifurcation creates a saddle and a locally asymptotically stable coexistence equilibrium of defectors and protective cooperators. When full cooperation and full defection are also locally asymptotically stable, three attractors coexist, and the numerically observed evolutionary outcome depends on the initial population composition. These results distinguish the effectiveness of activated protection from its ability to become established from rarity.

1. Introduction

Cooperation enables individuals to produce collective benefits that cannot be obtained independently, yet it is vulnerable to free riding. Explaining how cooperation emerges and persists despite this conflict is a central problem in evolutionary theory [1, 2]. The public goods game (PGG) provides a standard representation of this dilemma. Cooperators pay a personal cost to generate a shared return, whereas defectors obtain the return without contributing [3, 4]. When the individual return from contributing is lower than its cost, selection favors defection even though universal cooperation would yield a better collective outcome. The PGG therefore captures the evolutionary basis of the tragedy of the commons [5–7].

Evolutionary game theory has identified several mechanisms that can offset the individual advantage of defection. Five canonical routes are kin selection [8], direct reciprocity [9], indirect reciprocity [10], network reciprocity [11], and group selection [12]. These mechanisms promote cooperation through distinct evolutionary channels. Kin selection relies on genetic relatedness, direct reciprocity on repeated interactions, indirect reciprocity on reputation, network reciprocity on structured populations, and group selection on differences in success among groups [13–16]. Incentive mechanisms provide an alternative route by directly modifying the payoff difference between cooperation and defection. Costly punishment and reward can promote cooperation [17–19], although their effectiveness may be limited by implementation costs and second-order free riding [20–22]. Institutional and behavioral rules, including voluntary participation [23], optional exit [24, 25], and social exclusion [26–28], provide further routes by regulating entry into collective interactions or access to shared returns. Together, these studies show how assortment, incentives, and

*Corresponding authors

 shi_lei65@hotmail.com (L. Shi); hejun@kmu.edu.cn (J. He)

participation rules can favor cooperation in ordinary public goods games [29, 30]. A different problem arises when defection not only reduces public-good provision but also increases the probability of a collective failure that affects every group member.

The collective-risk social dilemma formalizes this shared-loss structure by imposing a probabilistic loss when a group fails to reach a collective target [31, 32]. Experiments show that sufficiently severe risk can increase contributions, although high risk or scientific uncertainty alone does not guarantee collective success [33, 34]. Evolutionary analyses further show that contribution targets create coordination barriers and that risk magnitude, group size, and target strictness can generate multiple evolutionary outcomes [35–38]. **In spatial populations, group-performance-dependent risk can help avert collective failure, although excessively demanding targets may weaken the cooperative effect of stronger risk feedback [39].** Subsequent extensions have examined inequality, intergenerational sustainability, communication, insurance, and feedback between population behavior and the risk environment [40–46]. **Communication of group sentiment and intended investment can also reverse inaction by improving coordination under shared risk [47].** More recent studies have broadened this institutional perspective by introducing centralized monitoring and reporting, competing risk-sharing arrangements, adaptive collective targets, and joint feedback among cooperation, risk, and cost [48–52]. These studies demonstrate that institutional responses and endogenous feedback can reorganize the equilibrium structure and generate initial-condition-dependent outcomes. Nevertheless, typical public-good provision is generally not separated from a distinct, cost-bearing, and non-excludable protective activity whose effectiveness depends on its own participation quorum.

In many real systems, preventing catastrophic loss often requires protective activities that are fundamentally distinct from ordinary public-good provision. This type of protection entails additional operating costs and activates only once participation reaches a prescribed quorum, after which it reduces risk for everyone regardless of their contribution. For instance in cross-border epidemic control, ordinary contributions typically fund routine public-health services. Protective cooperation, by contrast, finances joint disease surveillance, interoperable early-warning systems, and shared laboratory capacity. The participation quorum defines the minimum network of regions required for reliable detection. A similar logic applies to regional disaster prevention. Here, protective cooperation builds connected flood-control and monitoring infrastructure beyond standard public services. These systems activate only when enough jurisdictions participate, yet their protective shield is entirely non-excludable. Because such protection is inherently costly, quorum-dependent, and non-excludable, the evolutionary conditions required for its emergence remain a challenging open question.

To address this gap, we develop an N -player evolutionary public-goods game that explicitly separates risk mitigation from ordinary contributions. The population consists of ordinary cooperators, defectors, and protective cooperators. Both ordinary and protective cooperators contribute to the public good, while protective cooperators also incur an operating cost. The probability of collective failure increases and gradually saturates with the number of defectors, making risk endogenous to group composition. Protection reduces this probability for every group member, including defectors, but becomes active only when the number of protective cooperators reaches a prescribed threshold. We derive the expected payoffs under random matching, establish the local invasion and stability conditions of the three-strategy replicator system, and use numerical branch tracking to locate changes in its equilibrium structure. The results show that the advantage of protective cooperation varies non-monotonically with group composition because it is concentrated in configurations in which the focal individual completes the quorum. **When activation requires more than one protective cooperator, a rare protective cooperator cannot activate the mechanism, so changing protection effectiveness does not alter its first-order invasion fitness at full defection. If full defection is locally asymptotically stable, even a highly effective protective mechanism cannot overcome the resulting rare-invasion barrier.**

Sufficiently effective protection can nevertheless create a locally asymptotically stable coexistence state at a positive protective-cooperator frequency through a boundary saddle-node bifurcation while full defection remains locally asymptotically stable. When full cooperation is also locally asymptotically stable, the resulting tristability makes the numerically observed outcome depend on the initial population composition.

2. Model and methods

2.1. Population and strategies

Consider an infinite, well-mixed population whose members repeatedly form groups of size $N \geq 3$. Each individual uses one of three strategies: ordinary cooperation (C), defection (D), or protective cooperation (S). Let x , y , and z

denote their population frequencies. The state space is the simplex

$$\Delta_2 = \left\{ (x, y, z) \in \mathbb{R}_{\geq 0}^3 \mid x + y + z = 1 \right\}.$$

Under random matching, the strategies of the other group members are sampled independently from these frequencies.

Within any realized group, individuals adopting either strategy C or S contribute a fixed cost $c > 0$ to a public pool, whereas D -players contribute nothing. This aggregate contribution is multiplied by a synergy factor r and distributed equally among all N group members. By imposing the standard social-dilemma condition $1 < r < N$, any individual contribution increases the collective payoff ($r > 1$) yet yields a net personal deficit ($r/N < 1$) in the hypothetical absence of collective risk. Furthermore, beyond the baseline contribution c , strategy S incurs an independent operating cost $k > 0$ per interaction, strictly regardless of whether the protective mechanism is ultimately triggered.

2.2. Collective risk and quorum-activated protection

Crucially, defection endogenously generates a shared risk of collective failure. For a realized group containing d defectors, the failure probability is

$$p(d) = 1 - e^{-\gamma d} = 1 - \rho^d, \quad \gamma > 0, \quad (1)$$

where γ measures the sensitivity of collective failure risk to the number of defectors and $\rho = e^{-\gamma} \in (0, 1)$ is the multiplicative survival factor associated with one defector.

This exponential specification assumes that each additional defector multiplies the groups probability of avoiding failure by a constant factor ρ . This formulation naturally keeps the failure probability bounded between zero and one, yielding $p(d+1) - p(d) = (1 - \rho)\rho^d > 0$. Consequently, the collective risk strictly increases with the number of defectors, yet its marginal impact diminishes as failure becomes increasingly certain. Saturation is therefore a highly appropriate assumption for scenarios where additional free-riding inflicts less marginal damage on a group that is already severely exposed to collective collapse.

Crucially, the core pivotality mechanism driven by the quorum rule does not rely on this specific exponential form. For any alternative bounded and monotonically increasing risk function satisfying $p(0) = 0$, the strategic advantage of a focal protective cooperator emerges precisely when its participation completes the quorum. Similarly, the structural rare-invasion barrier remains universally robust: whenever activation requires multiple cooperators, a single mutant cannot trigger the mechanism, meaning the actual protection effectiveness drops out of the first-order invasion fitness entirely. However, the exact closed-form simplifications, specific equilibrium coordinates, and numerical bifurcation thresholds do remain dependent on the chosen exponential specification.

If collective failure occurs, every group member suffers the same additive loss $L > 0$, independently of the public-good return, and no transfers occur between strategies. Collective protection is activated if and only if the number of protective cooperators reaches an integer quorum $M \in \{2, \dots, N - 1\}$. Once activated, protection reduces the baseline failure probability by the fraction $\omega \in [0, 1]$. To avoid notational overlap with repeated-game models, we explicitly note that throughout the manuscript, ω exclusively denotes protection effectiveness. A group containing d defectors therefore faces the residual failure probability $p(d)$ when the quorum is not reached and $(1 - \omega)p(d)$ when it is reached. Because the reduction applies to every group member, including defectors, activated protection is non-excludable rather than a targeted sanction.

We restrict the quorum to $M \in \{2, \dots, N - 1\}$ to focus on the nondegenerate coordination problem. If $M = 1$, a single mutant can activate protection, so ω enters its first-order invasion fitness near full defection. If $M = N$, activation requires an all- S group, which contains no defectors and has baseline risk $p(0) = 0$; protection is therefore redundant.

2.3. Payoffs in a realized group

Let $n = N - 1$ be the number of co-players of a focal individual, and let i , j , and m be the numbers using strategies C , D , and S , respectively. Thus, $i + j + m = n$.

For a focal C - or D -player, protection is activated when at least M co-players use strategy S , so $m \geq M$. A focal S -player counts toward the quorum and therefore requires only $m \geq M - 1$ protective co-players.

Letting $\mathbb{I}_{\mathcal{E}}$ denote the standard indicator function – yielding 1 if event $\mathcal{E}$ occurs and 0 otherwise – the expected payoffs for the three strategies, conditional on the realized co-player composition, are formulated as

$$\begin{cases} \pi_C = \frac{rc(i+m+1)}{N} - c - Lp(j) [1 - \omega \mathbb{I}_{m \geq M}], \\ \pi_D = \frac{rc(i+m)}{N} - Lp(j+1) [1 - \omega \mathbb{I}_{m \geq M}], \\ \pi_S = \frac{rc(i+m+1)}{N} - c - k - Lp(j) [1 - \omega \mathbb{I}_{m \geq M-1}]. \end{cases} \quad (2)$$

In each payoff, the first term is the focal player's share of the public-good return. Strategies C and S pay the contribution cost c , and strategy S also pays the operating cost k . A focal defector increases the total number of defectors from j to $j+1$, so its failure probability is $p(j+1)$. The final term is the expected collective loss, with the indicator function imposing the appropriate quorum condition for each focal strategy.

2.4. Expected payoffs under random matching

Under random matching, the focal individual's co-player composition satisfies $(i, j, m) \sim \text{Multinomial}(n; x, y, z)$. The expected payoff of strategy $\ell \in \{C, D, S\}$ is $P_\ell = \mathbb{E}[\pi_\ell(i, j, m)]$. These multinomial expectations are evaluated exactly in Appendix A.

Define the expected non-failure factor $q = x + z + \rho y = 1 - (1 - \rho)y$. A randomly sampled co-player contributes a factor 1 if it uses C or S and a factor ρ if it uses D , so the expected factor is q . Independence of the n draws then gives $q^n = \mathbb{E}[\rho^j] = \mathbb{E}[e^{-\gamma j}]$, where $j \sim \text{Binomial}(n, y)$.

Furthermore, to capture the threshold-dependent activation of the protective mechanism, we introduce the risk-weighted activation terms for a generic integer threshold $h \in \{0, \dots, n\}$ as follows:

$$T_h = \mathbb{E}[p(j) \mathbb{I}_{m \geq h}], \quad T_h^D = \mathbb{E}[p(j+1) \mathbb{I}_{m \geq h}]. \quad (3)$$

These quantities are risk-weighted activation terms rather than activation probabilities, because each threshold-satisfying composition is weighted by its state-dependent failure probability. Thus, ωT_h and ωT_h^D give the expected reductions in failure probability, whereas $L\omega T_h$ and $L\omega T_h^D$ give the corresponding expected losses avoided by protection. Their finite-sum expressions are derived in Appendix A.2.

By setting $h = 0$, we effectively remove the activation filter, which yields:

$$T_0 = 1 - q^n, \quad T_0^D = 1 - \rho q^n. \quad (4)$$

In this context, T_0 signifies the expected baseline unprotected failure probability faced by a focal contributor, whereas T_0^D incorporates the marginal risk increment induced by a focal defector.

Aggregating these mathematical components, the expected population-level payoffs for the three strategies are formulated as:

$$\begin{cases} P_C = \frac{rc[1 + n(x+z)]}{N} - c - L [T_0 - \omega T_M], \\ P_D = \frac{rcn(x+z)}{N} - L [T_0^D - \omega T_M^D], \\ P_S = \frac{rc[1 + n(x+z)]}{N} - c - k - L [T_0 - \omega T_{M-1}]. \end{cases} \quad (5)$$

Structurally, the first component in each payoff equation captures the expected share of the public-good return, the second component accounts for the strategy-specific investment costs, and the final bracketed term reflects the residual expected collective loss after accounting for the protective mechanism. Notably, the activation index shifts from M to $M-1$ exclusively for a focal protective cooperator (P_S), accurately reflecting the endogenous property that the focal individual's own participation counts toward satisfying the critical activation quorum.

2.5. Selection gradients

To analytically isolate the marginal protection benefit uniquely conferred by strategy S , we define the pivotal threshold probability as $\Psi_M := T_{M-1} - T_M$. Because the events $m \geq M-1$ and $m \geq M$ differ exclusively at the discrete boundary $m = M-1$, this quantity can be explicitly expanded as

$$\begin{aligned}\Psi_M &= \mathbb{E}[p(j) \mathbb{I}_{m=M-1}] \\ &= \binom{n}{M-1} z^{M-1} [(1-z)^{n-M+1} - (x+\rho y)^{n-M+1}].\end{aligned}\quad (6)$$

The quantity Ψ_M represents the risk-weighted probability that the focal protective cooperator completes the activation threshold. Consequently, the payoff difference between protective and ordinary cooperators reduces to the exact identity

$$P_S - P_C = -k + L\omega\Psi_M. \quad (7)$$

Here, $L\omega\Psi_M$ is the expected collective loss avoided because the focal player is pivotal for activation. Protective cooperation is favored over ordinary cooperation if and only if this expected pivotal benefit exceeds the fixed operating cost k .

To facilitate the subsequent dynamical analysis on the state space, we eliminate the frequency of defectors by substituting $y = 1 - x - z$. From this point onward, the variables q , R_M , and Ψ_M denote their corresponding composite functions evaluated on the reduced two-dimensional simplex, ensuring that all partial derivatives with respect to x and z are strictly taken within this constrained domain. We then define the selection gradients relative to defection as

$$A(x, z; \omega) = P_C - P_D, \quad B(x, z; \omega) = P_S - P_D. \quad (8)$$

For notational brevity, all structural parameters other than the protection effectiveness ω are suppressed from the function arguments.

To express these selection gradients in closed form, we introduce the baseline marginal payoff parameter $a = \frac{rc}{N} - c$ alongside the threshold-dependent expectation term $R_M = \mathbb{E}[\rho^j \mathbb{I}_{m \geq M}]$. As formally derived in Appendix B, substituting these components yields the exact closed-form expressions:

$$A(x, z; \omega) = a + L(1 - \rho) [q^n - \omega R_M], \quad (9)$$

$$B(x, z; \omega) = A(x, z; \omega) - k + L\omega\Psi_M. \quad (10)$$

The constant a captures the direct payoff consequence of replacing a defector with a contributor in the absence of collective risk. The bracketed term in $A(x, z; \omega)$ gives the associated reduction in expected collective loss, including groups in which protection is already active. The function $B(x, z; \omega)$ additionally incorporates the operating cost k and the risk-weighted pivotal benefit $L\omega\Psi_M$ specific to protective cooperation.

2.6. Replicator dynamics

To model the evolutionary trajectories of the competing strategies, we first establish the mean population payoff, defined as $\bar{P} = xP_C + yP_D + zP_S$. By absorbing an arbitrary positive selection rate into the time scale without loss of generality, the temporal evolution of the strategy frequencies is governed by the standard replicator equations [53, 54]:

$$\begin{cases} \dot{x} = x(P_C - \bar{P}), \\ \dot{y} = y(P_D - \bar{P}), \\ \dot{z} = z(P_S - \bar{P}), \end{cases} \quad (11)$$

where the overdot denotes differentiation with respect to evolutionary time.

Because the population frequencies must rigorously sum to unity, eliminating the defector frequency via $y = 1 - x - z$ allows us to mathematically project the dynamics onto the reduced two-dimensional state space, formally defined as the simplex $\Delta_{xz} = \{(x, z) \in \mathbb{R}_{\geq 0}^2 \mid x + z \leq 1\}$. By substituting the previously derived selection gradients

from Eq. (8) into this framework, the continuous-time evolution on the reduced simplex can be explicitly expressed as the planar system:

$$\begin{cases} \dot{x} = x [(1-x)A(x, z; \omega) - zB(x, z; \omega)], \\ \dot{z} = z [(1-z)B(x, z; \omega) - xA(x, z; \omega)]. \end{cases} \quad (12)$$

Structurally, the boundaries of this simplex constitute invariant manifolds. Specifically, the coordinate edges $x = 0$ and $z = 0$ remain strictly invariant because the evolutionary growth rate of any strategy is inherently scaled by its own frequency, thereby preventing an absent strategy from spontaneously emerging. Similarly, the diagonal boundary spanning $x + z = 1$ is dynamically invariant, as this condition corresponds to a population entirely devoid of defectors ($y = 0$), rendering $\dot{y} = 0$ within the original full system in Eq. (11). Together, Equations (9), (10), and (12) form the closed, rigorously defined planar dynamical system that serves as the analytical foundation for all subsequent stability and bifurcation analyses.

2.7. Equilibria and local stability

To facilitate a rigorous stability analysis of the dynamical system, we first rewrite the planar equations (12) in the generalized vector field form $\dot{x} = F(x, z; \omega)$ and $\dot{z} = G(x, z; \omega)$, and correspondingly define the associated Jacobian matrix of the system as

$$J(x, z; \omega) = \begin{pmatrix} F_x & F_z \\ G_x & G_z \end{pmatrix}. \quad (13)$$

The explicit analytical expressions for the general Jacobian, alongside its specific reductions evaluated at both boundary and interior equilibria, are systematically derived in Appendix C.1.

Following standard nonlinear dynamical systems theory, an equilibrium point is classified as locally asymptotically stable if and only if both eigenvalues of its evaluated Jacobian matrix possess strictly negative real parts. Conversely, the presence of at least one eigenvalue with a positive real part renders the equilibrium unstable. In non-hyperbolic scenarios where the leading eigenvalue possesses a zero real part, first-order linearization becomes structurally inconclusive, thereby necessitating the examination of higher-order terms or center manifold reductions to rigorously ascertain the system's local topological behavior.

The homogeneous populations consisting entirely of defectors, ordinary cooperators, and protective cooperators are denoted by $\mathbf{D} = (0, 1, 0)$, $\mathbf{C} = (1, 0, 0)$, and $\mathbf{S} = (0, 0, 1)$. They correspond to the vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$ of the reduced simplex, respectively.

2.7.1. Homogeneous equilibria

At the full-defection equilibrium $\mathbf{D}$, we define the baseline boundary selection gradient as $A_D := A(0, 0; \omega) = a + L(1 - \rho)\rho^n$. Evaluating the Jacobian at this state yields the two transversal invasion eigenvalues:

$$\lambda_C^D = A_D, \quad \lambda_S^D = A_D - k. \quad (14)$$

The first eigenvalue gives the initial growth rate of a rare ordinary cooperator in an otherwise defecting population, whereas the second gives the corresponding growth rate of a rare protective cooperator. Consequently, full defection is locally asymptotically stable if and only if $A_D < 0$, is a saddle when $0 < A_D < k$, and is an unstable node when $A_D > k$. The boundary equality cases are nonhyperbolic and are analyzed in Appendix C.3.

Neither eigenvalue in Eq. (14) depends on ω . Because a single mutant protective cooperator cannot trigger the mechanism, the actual protection effectiveness has no bearing on the first-order invasion directions at full defection. However, the evolutionary consequence of this mathematical fact depends strictly on the baseline dynamics. When $A_D < 0$, the full-defection equilibrium is locally asymptotically stable, meaning this structural deficit translates into a permanent rare-invasion barrier across all admissible values of ω .

Similarly, at the full-cooperation equilibrium $\mathbf{C}$, we define the analogous boundary gradient $A_C := A(1, 0; \omega) = a + L(1 - \rho)$. The corresponding invasion eigenvalues governing the introduction of alternative strategies are determined as

$$\lambda_D^C = -A_C, \quad \lambda_S^C = -k. \quad (15)$$

Thus, the full-cooperation state achieves local asymptotic stability precisely when $A_C > 0$, and it degenerates into a saddle point when $A_C < 0$. In the critical scenario where $A_C = 0$, the state becomes nonhyperbolic; Appendix C.3 formally establishes its instability by analyzing the flow dynamics strictly along the invariant C – D edge.

Finally, at the full-protective-cooperation equilibrium S , the exact invasion eigenvalues are determined as

$$\lambda_C^S = k, \quad \lambda_D^S = k - a - L(1 - \rho)(1 - \omega). \quad (16)$$

Because $\lambda_C^S = k > 0$ for every $k > 0$, the full-protective-cooperation equilibrium is locally unstable. It acts as a saddle when $\lambda_D^S < 0$ and an unstable node when $\lambda_D^S > 0$. When $\lambda_D^S = 0$, the equilibrium becomes nonhyperbolic yet remains unstable since $\lambda_C^S > 0$.

2.7.2. Boundary equilibria

Restricting the evolutionary dynamics to the invariant C – D edge where $z = 0$, the differential equation governing the fraction of ordinary cooperators simplifies to $\dot{x} = x(1 - x)A(x, 0; \omega)$. As shown in Appendix C.4, the restricted gradient $A(x, 0; \omega)$ is strictly increasing in x . Consequently, a unique nontrivial edge equilibrium, denoted by $E_{CD} = (x_{CD}, 1 - x_{CD}, 0)$, exists if and only if $A_D < 0 < A_C$. Its tangential eigenvalue is positive, whereas its transverse protective-cooperator invasion eigenvalue is $-k < 0$; hence, E_{CD} is a saddle. Along the invariant edge, this saddle separates the one-dimensional basins of full cooperation and full defection.

Turning to the invariant C – S edge, the complete absence of defectors implies that the collective failure probability functionally vanishes ($p(0) = 0$), which correspondingly nullifies the pivotal threshold probability ($\Psi_M = 0$). Under these ideal conditions, the payoff differential identity in Eq. (7) tightly reduces to $P_S - P_C = -k < 0$. This structural asymmetry guarantees that no nontrivial C – S equilibrium can exist for any strictly positive operating cost ($k > 0$); thus, the evolutionary flow along this boundary is unidirectionally driven from S toward C . The degenerate, cost-free scenario where $k = 0$ entails structural sensitivities and is separately addressed in Appendix C.5.

Along the invariant D – S edge, any nontrivial coexistence equilibrium must take the generic form $E_{DS} = (0, 1 - z_{DS}, z_{DS})$ for $0 < z_{DS} < 1$, and must strictly satisfy the root condition $B(0, z_{DS}; \omega) = 0$. By linearizing the system around this boundary state, the transverse and tangential eigenvalues are respectively given by

$$\lambda_{\perp} = A(0, z_{DS}; \omega), \quad \lambda_{\parallel} = z_{DS}(1 - z_{DS})B_z(0, z_{DS}; \omega). \quad (17)$$

The transverse eigenvalue $\lambda_{\perp}$ gives the invasion growth rate of a rare ordinary cooperator, whereas the tangential eigenvalue $\lambda_{\parallel}$ governs perturbations along the invariant edge. Assuming hyperbolicity, the D – S equilibrium is locally asymptotically stable if and only if $A(0, z_{DS}; \omega) < 0$ and $B_z(0, z_{DS}; \omega) < 0$. It is a saddle when the two eigenvalues have opposite signs, while the vanishing of either eigenvalue makes first-order linearization inconclusive.

A critical topological transition, characterized as a boundary saddle–node bifurcation, unfolds when two distinct nontrivial D – S equilibria mutually approach and coalesce. Analytically, this bifurcation is pinpointed by the simultaneous satisfaction of the equilibrium and tangency conditions:

$$B(0, z_{DS}; \omega) = 0, \quad B_z(0, z_{DS}; \omega) = 0. \quad (18)$$

To ensure that the tangential zero eigenvalue remains geometrically simple within the full two-dimensional planar system, the transverse non-degeneracy condition must hold:

$$A(0, z_{DS}; \omega) \neq 0. \quad (19)$$

With ω as the bifurcation parameter and all other parameters fixed, a nondegenerate codimension-one fold also requires

$$B_{zz}(0, z_{DS}; \omega) \neq 0, \quad B_{\omega}(0, z_{DS}; \omega) \neq 0. \quad (20)$$

The rigorous derivation of these bifurcation criteria, grounded in the scalar vector field properly restricted to the D – S edge, is systematically detailed in Appendix C.7.

2.7.3. Interior equilibria

By definition, a fully polymorphic (interior) equilibrium, denoted as $E_* = (x^*, y^*, z^*)$, requires the strictly positive coexistence of all three strategies ($x^*, y^*, z^* > 0$). Within the replicator dynamics framework, this mutual coexistence necessitates the exact equalization of all three expected payoffs, which is algebraically equivalent to the simultaneous vanishing of the respective selection gradients:

$$A(x^*, z^*; \omega) = 0, \quad B(x^*, z^*; \omega) = 0. \quad (21)$$

As systematically detailed in Appendix C.8, the Jacobian matrix evaluated at such an interior state admits a mathematically tractable factorization. Specifically, it can be analytically shown that the determinant of this evaluated Jacobian, $\det J_*$, simplifies to the exact form

$$\det J_* = x^* y^* z^* (A_x B_z - A_z B_x) \Big|_{E_*}. \quad (22)$$

Applying standard linear stability criteria, a hyperbolic interior equilibrium achieves local asymptotic stability if and only if its Jacobian simultaneously possesses a positive determinant ($\det J_* > 0$) and a negative trace ($\text{tr } J_* < 0$). Conversely, the state structurally manifests as a saddle point whenever its determinant is strictly negative ($\det J_* < 0$), and acts as a locally unstable source (or unstable spiral) if both the determinant and the trace are positive ($\det J_* > 0$ and $\text{tr } J_* > 0$). In the critical event of a singular Jacobian matrix ($\det J_* = 0$), or a pure imaginary pair of eigenvalues ($\text{tr } J_* = 0$ with $\det J_* > 0$), the equilibrium degenerates into a nonhyperbolic state, rendering first-order linearization dynamically inconclusive.

A prospective topological transition involving an interior saddle-node bifurcation is mathematically pinpointed when an equilibrium candidate simultaneously satisfies the payoff equality in Eq. (21) and exhibits a singular Jacobian matrix. Because the strategy frequencies x^*, y^*, z^* are strictly positive in the interior domain, this singularity condition is algebraically equivalent to the vanishing of the determinant of the underlying selection gradients:

$$(A_x B_z - A_z B_x) \Big|_{E_*} = 0. \quad (23)$$

To ensure this zero eigenvalue remains geometrically simple within the planar system, the trace of the Jacobian must concurrently remain non-zero:

$$\text{tr } J_* \neq 0. \quad (24)$$

While these specific criteria successfully isolate an equilibrium with a simple zero eigenvalue, they are inherently insufficient to definitively prove the occurrence of a nondegenerate saddle-node bifurcation. To rigorously confirm this structural transition, the requisite parameter transversality and quadratic nondegeneracy conditions are independently formulated and computationally verified in Appendix D.3.

Finally, Table 1 systematically catalogues the diverse equilibrium classes subsequently employed in the numerical bifurcation analysis. Unless explicitly stated otherwise, it should be understood that all stability classifications provided herein strictly pertain to local topological properties within the reduced two-dimensional strategy simplex.

3. Numerical bifurcation results

Unless stated otherwise, the numerical branch tracking and phase-space calculations use the baseline parameter set

$$\begin{aligned} N = 5, \quad M = 2, \quad r = 2.3, \quad c = 1, \\ k = 0.4, \quad L = 4, \quad \gamma = 1.4. \end{aligned} \quad (25)$$

These baseline values serve strictly as an illustrative tool to map the bifurcation structures inherently admitted by the model, rather than reflecting a specific empirical calibration. It is important to emphasize the boundary between our analytical and numerical results. The equilibrium existence and local-stability conditions derived in Section 2 hold generally across the entire admissible parameter space. By contrast, the subsequent numerical outputs—specifically the reported equilibrium coordinates, fold locations, eigenvalues, phase portraits, and basin geometries—apply exclusively to this selected parameter set unless explicitly stated otherwise.

Table 1

Existence conditions and local classification of the equilibrium classes on the strategy simplex.

Equilibrium class	Existence condition	Local classification within the simplex
$D = (0, 1, 0)$	All admissible parameters	Locally asymptotically stable if $A_D < 0$; a saddle if $0 < A_D < k$; an unstable node if $A_D > k$; nonhyperbolic and unstable if $A_D = 0$ or $A_D = k$.
$C = (1, 0, 0)$	All admissible parameters	Locally asymptotically stable if $A_C > 0$; a saddle if $A_C < 0$; nonhyperbolic and unstable if $A_C = 0$.
$S = (0, 0, 1)$	All admissible parameters	Unstable because $\lambda_C^S = k > 0$; a saddle if $\lambda_D^S < 0$; an unstable node if $\lambda_D^S > 0$; nonhyperbolic and unstable if $\lambda_D^S = 0$.
E_{CD}	$A_D < 0 < A_C$	A saddle for $k > 0$, with one positive tangential eigenvalue and the negative transverse eigenvalue $-k$.
Nontrivial C – S equilibrium	None for $k > 0$; an equilibrium continuum exists for $k = 0$.	For $k > 0$, the edge flow is directed from S toward C . For $k = 0$, the edge is neutral in the tangential direction and its transverse stability is parameter dependent.
$E_{DS} = (0, 1 - z_{DS}, z_{DS})$	$B(0, z_{DS}; \omega) = 0$, with $0 < z_{DS} < 1$	The transverse and tangential eigenvalues are $\lambda_{\perp} = A(0, z_{DS}; \omega)$ and $\lambda_{\parallel} = z_{DS}(1 - z_{DS})B_z(0, z_{DS}; \omega)$, respectively. The equilibrium is locally asymptotically stable if both eigenvalues are negative, an unstable node if both are positive, and a saddle if they have opposite signs. It is nonhyperbolic if either eigenvalue vanishes.
$E_* = (x^*, y^*, z^*)$	$A(x^*, z^*; \omega) = B(x^*, z^*; \omega) = 0$, with $x^*, y^*, z^* > 0$.	For a hyperbolic equilibrium, locally asymptotically stable if $\text{tr } J_* < 0$ and $\det J_* > 0$; locally unstable if $\text{tr } J_* > 0$ and $\det J_* > 0$; and a saddle if $\det J_* < 0$. The equilibrium is nonhyperbolic if $\det J_* = 0$, or if $\text{tr } J_* = 0$ with $\det J_* > 0$.

3.1. Two saddle–node bifurcations

Figure 1 shows how the equilibrium set changes with ω . Panel (a) reports the interior equilibria, and panel (b) reports equilibria on the invariant D – S edge ($x^* = 0$ and $y^* = 1 - z^*$). The pure C and D vertices are omitted because their locations are independent of ω and both remain locally asymptotically stable under the baseline parameters. The branch search, stability classification, and fold tests are described in Appendix D.

The first bifurcation, SN1, is a codimension-one saddle–node bifurcation in the interior of the simplex at

$$(x^*, y^*, z^*, \omega_{\text{SN1}}) = (0.2994, 0.4272, 0.2734, 0.2882).$$

Using our parameter-stepped root-search procedure, no interior equilibrium was detected below ω_{SN1} within the searched parameter interval. Crossing SN1 creates a saddle and an unstable node. At the fold, the Jacobian has a simple zero eigenvalue and a positive second eigenvalue, while the numerical normal-form coefficients are nonzero. Since neither branch is locally attracting, the analytical fold calculation solely confirms that SN1 introduces no new attractor. Beyond this local analytical result, our numerical phase portraits demonstrate that this bifurcation globally reorganizes the stable-manifold and separatrix geometry.

The second bifurcation, SN2, occurs on the invariant D – S edge at

$$(x^*, y^*, z^*, \omega_{\text{SN2}}) = (0, 0.7282, 0.2718, 0.5254).$$

At this point, the boundary equilibrium and tangency conditions hold:

$$B(0, z^*; \omega_{\text{SN2}}) = 0, \quad B_z(0, z^*; \omega_{\text{SN2}}) = 0.$$

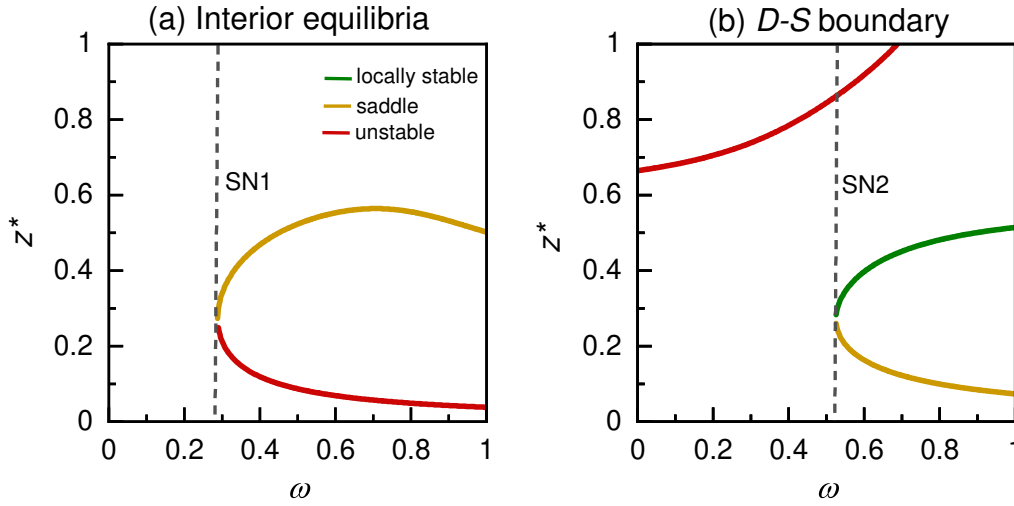


Figure 1: Only the boundary saddle-node creates a locally stable protective equilibrium at positive protective-cooperator frequency, whereas the interior saddle-node adds no attractor. Equilibrium branches with respect to protection effectiveness ω under the baseline parameters in Eq. (25). Panel (a) shows the protective-cooperator frequency z^* of the interior equilibria, and panel (b) shows equilibria on the invariant D – S edge. Green, orange, and red curves denote locally asymptotically stable, saddle, and unstable branches, respectively. The dashed vertical lines mark the interior saddle-node SN1 at $\omega_{\text{SN1}} \approx 0.2882$ and the boundary saddle-node SN2 at $\omega_{\text{SN2}} \approx 0.5254$. The unstable D – S branch reaches an edge transcritical bifurcation at $\omega \approx 0.6881$. The pure- C and pure- D equilibria are omitted because their locations and local classifications do not change with ω under the baseline parameters.

The derivative tests in Appendix D.3 confirm that SN2 is nondegenerate. For $\omega > \omega_{\text{SN2}}$, the fold produces two D – S equilibria. The lower- z branch is a saddle. On the upper- z branch, both the transverse eigenvalue $\lambda_{\perp} = A(0, z^*; \omega)$ and the tangential eigenvalue $\lambda_{\parallel} = z^*(1 - z^*)B_z(0, z^*; \omega)$ are negative, so this branch is locally asymptotically stable within the simplex. Because the pure C and D vertices remain locally stable, the baseline system changes from bistability to tristability after SN2. A separate unstable D – S branch reaches the pure- S vertex at $\omega \approx 0.6881$. Because this vertex collision does not change the number of local attractors, it is not involved in the formation of the protective state.

3.2. Phase-space reorganization and tristability

Figure 2 shows the corresponding changes in the simplex flow. At $\omega = 0.20$ [panel (a)], C and D are locally asymptotically stable, whereas S is a saddle with eigenvalues -1.4709 and 0.4000 . The observed trajectories approach either C or D , with the stable manifold of the C – D edge saddle separating their basins. On the defector-free C – S edge, Eq. (7) reduces to $P_S - P_C = -k < 0$, so the edge flow points from S toward C .

At $\omega = 0.40$ [panel (b)], SN1 has created the interior saddle and unstable node, while SN2 has not yet occurred. The stable manifold of the interior saddle changes the numerically observed separatrix geometry, but the only local attractors remain C and D .

At $\omega = 0.70$ [panel (c)], after SN2, the locally asymptotically stable upper D – S equilibrium is

$$E_{DS}(0.70) = (0, 0.5488, 0.4512).$$

Equation (17) gives $\lambda_{\perp} = -0.4229$ and $\lambda_{\parallel} = -0.5349$. Both are real and negative, so E_{DS} is a locally asymptotically stable node and the linearized relaxation near this equilibrium is non-oscillatory.

Taken together, Figs. 1 and 2 establish the fundamental dynamical sequence under our baseline parameters. The system begins with full cooperation and full defection forming a simple bistable pair. An interior saddle-node bifurcation then generates a saddle and an unstable node. While this numerically alters the separatrix geometry, it adds no new attractor. As the parameter ω increases, a subsequent boundary saddle-node bifurcation creates a saddle alongside a locally asymptotically stable equilibrium where defectors and protective cooperators coexist. Because the pure states of full cooperation and full defection both maintain their stability, this boundary emergence effectively

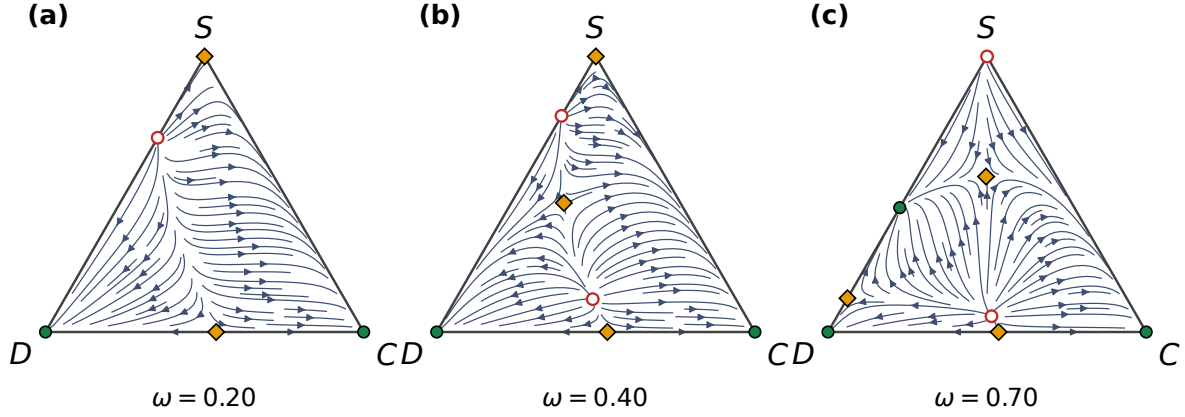


Figure 2: The numerical phase portraits reveal that the interior saddle-node bifurcation reorganizes the separatrix geometry without introducing a new attractor. In contrast, the boundary saddle-node bifurcation transitions the system from bistability to tristability by generating a locally stable D-S equilibrium. Replicator phase portraits on the equilateral strategy simplex under the baseline parameters at $\omega = 0.20$ in panel (a), $\omega = 0.40$ in panel (b), and $\omega = 0.70$ in panel (c). The vertices C , D , and S represent the three homogeneous states, and arrowed integral curves indicate forward-time flow. Filled green circles, orange diamonds, and open red circles denote locally asymptotically stable equilibria, saddles, and unstable equilibria, respectively.

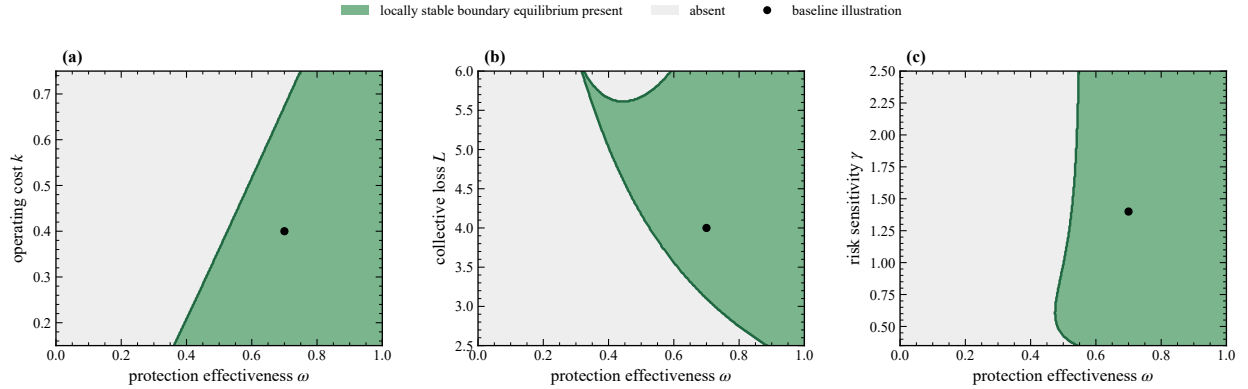


Figure 3: Pointwise classification of locally asymptotically stable, nontrivial D-S equilibria in three two-parameter planes. Panel (a) shows the (ω, k) plane at $L = 4$ and $\gamma = 1.4$, panel (b) shows the (ω, L) plane at $k = 0.4$ and $\gamma = 1.4$, and panel (c) shows the (ω, γ) plane at $k = 0.4$ and $L = 4$. Green points contain at least one boundary root satisfying $\lambda_1 < 0$ and $\lambda_{\parallel} < 0$, whereas no such root is detected at gray points. Solid contours are numerically extracted boundaries between these two classifications and need not represent a single saddle-node locus. Black points mark the parameter combinations used for the phase portrait at $\omega = 0.70$. The maps classify equilibrium existence and local stability, not basin size.

converts the system from bistability to tristability. This clear sequential progression establishes the structural baseline for the parameter comparisons that follow.

3.3. Parameter dependence of locally stable boundary protection

We next examine how operating cost, collective loss, and risk sensitivity modify the parameter region in which a locally asymptotically stable, nontrivial D-S equilibrium exists. For each sampled pair (ω, η) , where $\eta \in \{k, L, \gamma\}$, we computed all nontrivial roots of

$$B(0, z; \omega, \eta) = 0, \quad 0 < z < 1,$$

and evaluated their transverse and tangential eigenvalues,

$$\lambda_{\perp} = A(0, z; \omega, \eta), \quad \lambda_{\parallel} = z(1 - z)B_z(0, z; \omega, \eta).$$

A parameter pair is classified as “present” only when at least one boundary root satisfies $\lambda_{\perp} < 0$ and $\lambda_{\parallel} < 0$. The resulting regions therefore describe the existence of a locally asymptotically stable boundary equilibrium, not its basin size. The pointwise classification procedure is described in Appendix D.5.

Figure 3(a) shows that increasing the operating cost k shifts the lower boundary of the stable region toward greater protection effectiveness. The boundary is close to linear over the reported interval. This pattern is consistent with the affine dependence

$$B(0, z; \omega, k) = B_0(z) - k + \omega B_1(z),$$

although the boundary is not exactly linear because the equilibrium frequency z changes along it. At the baseline value $k = 0.4$, the lower boundary passes through $\omega \approx 0.5254$, in agreement with SN2 in Fig. 1.

The stability region responds to collective loss in a more complex manner, as shown in Fig. 3(b). Over the lower and intermediate portions of the sampled L interval, increasing L extends the stable region toward smaller values of ω . At sufficiently large L , however, the boundary between the stable and absent regions can no longer be characterized by a single threshold curve. For example, at $L = 6$, the sampled grid detects a locally stable boundary equilibrium for approximately

$$0.3175 \leq \omega \leq 0.3225 \quad \text{and} \quad 0.5925 \leq \omega \leq 1,$$

with no such equilibrium detected between these intervals. Thus, larger collective loss can reorganize the boundary equilibrium structure rather than merely lower a unique effectiveness threshold. This pattern arises numerically because L changes all risk-dependent terms in the boundary payoff difference, not only the pivotal benefit $L\omega\Psi_M$. **Because the classification in Figure 3(b) relies exclusively on pointwise root solving and stability evaluation rather than formal two-parameter branch continuation, it does not track the identity of individual equilibrium branches as L varies. Consequently, the two intervals observed at $L = 6$ must simply be interpreted as disconnected parameter regions where at least one locally stable boundary equilibrium is detected. They do not by themselves establish that both intervals belong to the identical branch, nor do they prove a re-entrant stability transition along a single continuous curve. Nevertheless, this topological feature highlights an important practical insight: increasing the collective loss does not merely shift a single, unique protection-effectiveness threshold.**

The minimum protection effectiveness associated with the lower stability boundary varies non-monotonically with risk sensitivity, as shown in Fig. 3(c). As γ increases, this boundary first shifts toward smaller values of ω , reaches its minimum near $\gamma \approx 0.61$, and then moves back toward larger values of ω . This reversal is consistent with the saturation of the risk function $p(d) = 1 - e^{-\gamma d}$. Once failure is already likely for the relevant group compositions, further increases in γ produce progressively smaller changes in the risk differences governing selection. The numerical map documents this non-monotonic pattern over the reported parameter interval, although the proposed saturation-based explanation is not an analytical proof of its cause.

3.4. Quorum size shifts the stability threshold

The quorum size M is discrete, so we compare separate boundary calculations rather than vary it as a continuous parameter. For $N = 5$ and the remaining baseline parameters, Table 2 shows that the critical effectiveness increases from 0.5254 at $M = 2$ to 0.8321 at $M = 4$. The route to stability also changes. At $M = 2$, the stable branch is created directly at the boundary fold SN2. At $M = 3$ and $M = 4$, a D – S branch exists before it becomes stable; local asymptotic stability is acquired when the transverse C -invasion eigenvalue $\lambda_{\perp} = A(0, z)$ crosses zero.

Changing M replaces Ψ_M by the risk-weighted probability of encountering exactly $M - 1$ protective co-players. It therefore changes both the location and magnitude of the pivotality term in the simplex. For the examined values, a stricter quorum requires greater protection effectiveness and can change the bifurcation mechanism by which the boundary equilibrium becomes stable. This comparison does not establish monotonicity for other group sizes or parameter sets.

Table 2Critical protection effectiveness ω_c and the dynamical route to local asymptotic stability across discrete quorum sizes M .

M	ω_c	z_c	Local transition
2	0.5254	0.2718	Boundary saddle–node
3	0.6608	0.7935	Transverse invasion crossing
4	0.8321	0.9541	Transverse invasion crossing

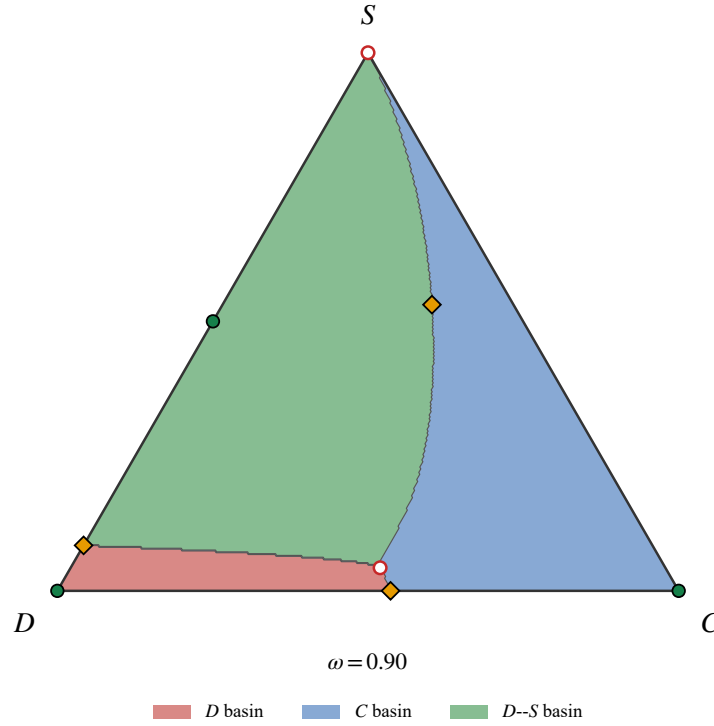


Figure 4: Numerically observed basins at $\omega = 0.90$ under the baseline parameters. Red, blue, and green initial states approach full defection, full cooperation, and the locally asymptotically stable D – S boundary equilibrium, respectively. Filled green circles, orange diamonds, and open red circles denote locally asymptotically stable equilibria, saddles, and unstable equilibria. **Terminal states were assigned only when their Euclidean distance from a candidate attractor previously identified by the equilibrium and stability analysis was below 10^{-6} .** The exact pure- S state remains unclassified. The colored regions are finite-grid numerical basin estimates rather than a global basin classification.

3.5. Numerically observed basin geometry after SN2

Local asymptotic stability does not determine which equilibrium is approached from a given initial composition. We therefore integrated the replicator system from a uniform barycentric grid at $\omega = 0.90$. At this parameter value, the locally asymptotically stable boundary equilibrium is $E_{DS}(0.90) = (0, 0.4992, 0.5008)$, with eigenvalues -0.8428 and -0.4836 .

Figure 4 shows the numerically observed basins of full defection, full cooperation, and $E_{DS}(0.90)$. The sampled basin boundaries pass through the reported saddles and are consistent with their stable manifolds. The protective boundary equilibrium attracts a positive fraction of the sampled initial conditions, but its basin does not contain the full-defection vertex. Under the reported parameters, initial compositions on the defector-heavy side of the observed separatrix approach full defection rather than the protective boundary state.

The basin calculation classified 45 450 of the 45 451 initial states. The only unclassified state was the exact pure- S equilibrium, which remains at that unstable equilibrium and does not approach any of the three attractors used in the

basin assignment. Repeating the calculation with half the integration step and, separately, with a longer integration horizon produced no classification changes among the resolved grid points. These results support the displayed numerical partition for the stated parameters, but they do not constitute a global convergence theorem.

4. Discussion and conclusion

This study is concerned not with whether protection can reduce collective risk after activation, but with how a threshold-dependent and non-excludable protective activity enters the evolutionary process. Analysis of the endogenous-risk public-goods game shows that threshold activation generates qualitatively different selection effects across population compositions. The payoff advantage of protective over ordinary cooperation does not increase continuously with the abundance of protective cooperators. Instead, it is concentrated in pivotal configurations where the focal individual completes the activation threshold. This composition dependence creates a dynamical separation between the rare-strategy stage and the established protective state. Near full defection, rare protective cooperators cannot activate the mechanism, so protection effectiveness does not alter their local invasion direction. Once protective cooperators have an established presence, however, the same protection effectiveness can substantially reshape the nonlinear selection gradient and create a locally asymptotically stable coexistence state of defectors and protective cooperators through a boundary saddle-node bifurcation. **The emergence of this new attractor occurs without destabilizing the full-defection state. Consequently, under our reported baseline parameters, both the protective coexistence equilibrium and full defection operate as local attractors. Because full cooperation also retains its local asymptotic stability, the strategy simplex effectively harbors three distinct local attractors. As demonstrated by our numerical basin partition, the system's terminal state for these parameters is highly sensitive to the initial population composition. It is important to emphasize that this finite-grid basin geometry reflects a numerical observation rather than a generalized global convergence theorem spanning the entire admissible parameter space.**

The dynamical results can be understood through a mechanism that links individual payoff differences to population-level bifurcation structure. Ordinary and protective cooperators make the same public-good contribution, so selection between them depends only on the operating cost continuously paid by protective cooperators and the expected loss avoided when they complete the activation threshold. Their payoff difference is therefore

$$P_S - P_C = -k + L\omega\Psi_M,$$

where

$$\Psi_M = \mathbb{E}[p(j) \mathbb{I}_{m=M-1}]$$

is not the unconditional probability that the group reaches the activation threshold. Instead, it is the risk-weighted probability that the focal protective cooperator becomes the M th participant and thereby activates protection. The condition $m = M - 1$ ensures that the focal individual is indispensable for activation, while the risk weight $p(j)$ ensures that this pivotal role produces a benefit only when defectors generate a positive failure risk. This definition first explains why the advantage of protective cooperation depends on group composition. When protective cooperators are rare, the focal individual is unlikely to encounter $M - 1$ protective co-players. As the population approaches the pure-protection state, the threshold is usually satisfied without the focal individual, while the declining number of defectors also drives $p(j)$ toward zero. The protective benefit therefore does not increase continuously with protective-cooperator frequency. Instead, it is concentrated in intermediate group compositions where failure risk and pivotality are simultaneously substantial. The same structure explains why protection effectiveness does not enter the first-order invasion dynamics at full defection. For $M \geq 2$, Ψ_M scales with z^{M-1} near full defection, and the replicator equation for the protective strategy contains an additional multiplicative frequency factor z . The growth effect generated by protection effectiveness therefore enters the dynamical system at order z^M or higher and does not appear in the first-order Jacobian at full defection. Increasing ω can amplify the benefit of protection after activation, but it cannot change the local invasion direction of a rare protective cooperator near full defection. **This structural deficit implies that when full defection is locally stable ($A_D < 0$), the absence of a first-order activation benefit creates a formidable rare-invasion barrier. An otherwise highly effective protective technology can remain permanently out of reach if the population cannot overcome the initial participation deficit.** Once z moves away from zero, pivotal configurations occur with non-negligible probability, allowing $L\omega\Psi_M$ to compete with the fixed cost k . On the defector-protective-cooperator boundary, this competition enters the zero-growth condition $B(0, z; \omega) = 0$. As ω increases, the protective benefit

changes the nonlinear dependence of this condition on z . When the additional tangency condition $B_z(0, z; \omega) = 0$ is satisfied, the boundary equilibria undergo a saddle–node bifurcation. The resulting stable coexistence reflects the feedback between risk generation and risk mitigation. More defectors raise the failure probability and thereby increase the potential value of protection. More protective cooperators suppress the failure probability, but their operating cost does not disappear as risk declines. If defectors vanish completely, then $p(0) = 0$ and protection produces no additional benefit, allowing ordinary cooperators to replace protective cooperators through their cost advantage. This feedback explains why protective cooperation may fail to become established from rarity yet remain stable in coexistence with defection in population compositions that retain sufficient exposure to collective risk.

Existing studies have generally approached cooperation under collective risk along two broad lines. Collective-risk models link public-good contributions directly to risk mitigation, so that reaching an aggregate contribution target prevents a shared loss that otherwise occurs with an exogenously specified probability [33, 36–38]. Related cooperation models alter strategy-specific payoffs or redistribute exposure to collective losses through punishment [17, 18], social exclusion [26], or insurance [43]. Recent research has extended these approaches by considering centralized monitoring and reporting, competing risk-sharing arrangements, adaptive governance, and coevolutionary feedback among cooperation, risk, and cost [48–52]. These approaches emphasize aggregate contributions, targeted incentives, institutional adaptation, or continuously changing risk environments, but they do not separate ordinary public-good provision from a distinct non-excludable protective activity activated by its own participation quorum.

The present model separates failure-risk generation, public-good contribution, and collective protection into three related but distinct processes. Defection endogenously raises the failure probability, protective cooperators bear an additional cost beyond their public-good contribution, and activated protection benefits all group members, including defectors. Risk mitigation therefore enters evolutionary selection through pivotality in risk-bearing groups rather than directly through aggregate contributions or targeted transfers. Protective cooperation may consequently fail to invade because it cannot activate the mechanism when rare, yet form a locally asymptotically stable coexistence state through the creation of a new attractor at a positive frequency without requiring full defection to lose its local stability. This result extends the analysis of threshold-dependent protection beyond the rare-invasion criterion to local bifurcations and basin geometry. It also distinguishes the effectiveness of protection after activation, its capacity to form from a small number of participants, and its stable maintenance after establishment. Cross-border epidemic surveillance provides a relevant example because its effectiveness depends on shared national surveillance and reporting capacities. Improving monitoring or early-warning technology can increase the protective value of an established system, while its initial formation still depends on the operating cost, activation threshold, and initial participation structure.

The preceding analysis clarifies when collective protection can emerge and persist, but the dynamical stability of a protective state does not imply that it maximizes social welfare. Evolutionary prevalence and social welfare need not coincide when cooperative behavior entails additional costs [55]. In the present model, a higher frequency of protective cooperators may increase the probability of activation and reduce expected collective losses, but it also raises the operating costs borne by the population. Moreover, if defectors disappear, the failure risk vanishes because $p(0) = 0$, while protective cooperators continue to pay the operating cost k . Strategy frequencies, protection activation, local stability, and basin size therefore characterize different aspects of the evolutionary dynamics and do not by themselves identify the welfare-maximizing outcome. A formal welfare assessment would require comparing the mean population payoff across feasible states, which lies beyond the local-stability analysis undertaken here.

Some of the model assumptions naturally limit the broader validity of our conclusions. The infinite well-mixed population excludes spatial clustering and network heterogeneity. Local assortment may allow protective cooperators that are globally rare to satisfy the activation threshold within particular groups, but it may also produce spatial lock-in between protected and unprotected regions. The deterministic replicator dynamics do not capture demographic drift, mutation, or noise-induced transitions across basin boundaries in finite populations. Players are also assumed to have homogeneous contribution costs, operating costs, and protective capacities, while participation recognition and threshold activation are error free. Cost heterogeneity, recognition errors, and activation failure may alter both pivotality and local stability. In addition, the numerical branch tracking and bifurcation diagnostics identify local bifurcations, and the numerically observed basins apply only to the reported parameter ranges rather than constituting a global convergence result. Future research can extend the mechanism to complex networks, spatial lattices, and finite populations, examine how stochasticity, heterogeneity, and local assortment reshape the invasion barrier and basin structure, and allow the activation threshold, operating cost, and protection effectiveness to coevolve with population composition and environmental risk.

Appendix A. Derivation of the expected payoffs

This appendix derives the expected payoffs from the composition-dependent payoffs in Eq. (2). We first evaluate the risk-weighted activation terms and then assemble the public-good and collective-loss components.

A.1. Random matching

Among the $n = N - 1$ co-players of a focal individual, the numbers (i, j, m) of ordinary cooperators, defectors, and protective cooperators follow the multinomial distribution

$$\Pr(i, j, m) = \frac{n!}{i!j!m!} x^i y^j z^m,$$

subject to the strict composition constraint $i + j + m = n$. Hence, for any focal strategy $\ell \in \{C, D, S\}$, the expected population-level payoff is formulated as

$$P_\ell = \sum_{\substack{i, j, m \geq 0 \\ i + j + m = n}} \Pr(i, j, m) \pi_\ell(i, j, m). \quad (26)$$

All expressions below are exact analytical evaluations of this finite multinomial sum.

A.2. Risk-weighted activation terms

Starting from Eq. (3), the risk-weighted activation term for a focal contributor unfolds as

$$T_h = \sum_{m=h}^n \sum_{j=0}^{n-m} \frac{n!}{(n-m-j)!j!m!} x^{n-m-j} y^j z^m p(j).$$

Conversely, for a focal defector, its own strategy intrinsically changes the total defector count from j to $j + 1$, yielding

$$T_h^D = \sum_{m=h}^n \sum_{j=0}^{n-m} \frac{n!}{(n-m-j)!j!m!} x^{n-m-j} y^j z^m p(j+1).$$

Fixing m leaves $n - m$ positions occupied by ordinary cooperators and defectors. Summing the corresponding multinomial weights over these remaining positions gives

$$\sum_{j=0}^{n-m} \frac{n!}{(n-m-j)!j!m!} x^{n-m-j} y^j z^m = \binom{n}{m} z^m (1-z)^{n-m}. \quad (27)$$

If we include the failure-survival factor ρ^j in the summation instead, we obtain

$$\sum_{j=0}^{n-m} \frac{n!}{(n-m-j)!j!m!} x^{n-m-j} y^j z^m \rho^j = \binom{n}{m} z^m (x + \rho y)^{n-m}. \quad (28)$$

Because the failure probability is defined as $p(j) = 1 - \rho^j$, subtracting Eq. (28) from Eq. (27) and summing over all states satisfying the threshold $m \geq h$ explicitly yields

$$T_h(x, y, z) = \sum_{m=h}^n \binom{n}{m} z^m [(1-z)^{n-m} - (x + \rho y)^{n-m}]. \quad (29)$$

Similarly, for a focal defector, the incremented risk follows $p(j+1) = 1 - \rho^{j+1} = 1 - \rho \rho^j$. Applying the identical algebraic calculation gives

$$T_h^D(x, y, z) = \sum_{m=h}^n \binom{n}{m} z^m [(1-z)^{n-m} - \rho(x + \rho y)^{n-m}]. \quad (30)$$

Equations (29) and (30) serve as the finite-sum representations of the expectations previously defined in Eq. (3).

A.3. Unprotected failure probabilities

Setting $h = 0$ effectively removes the activation filter. By applying the binomial theorem to Eq. (29), the baseline unprotected risk simplifies to

$$\begin{aligned} T_0 &= \sum_{m=0}^n \binom{n}{m} z^m (1-z)^{n-m} - \sum_{m=0}^n \binom{n}{m} z^m (x + \rho y)^{n-m} \\ &= 1 - (z + x + \rho y)^n = 1 - q^n. \end{aligned}$$

The corresponding expression for a focal defector is straightforwardly evaluated as $T_0^D = 1 - \rho q^n$. Equivalently, by substituting $j \sim \text{Binomial}(n, y)$, we can confirm that the expectation evaluates to $\mathbb{E}[\rho^j] = (1 - y + \rho y)^n = q^n$. It naturally follows that the additional expected failure probability generated by the focal defector is exactly $T_0^D - T_0 = (1 - \rho)q^n > 0$.

A.4. Assembly of the expected payoffs

In a realized group, the expected number of contributing co-players is $\mathbb{E}[i + m] = n(x + z)$. Because a focal C - or S -player also contributes to the public good, their inclusion raises this expectation to $\mathbb{E}[i + m + 1] = 1 + n(x + z)$.

For a focal contributor subject to a generic co-player threshold h , the expected residual failure probability decomposes as

$$\begin{aligned} \mathbb{E}[p(j) (1 - \omega \mathbb{I}_{m \geq h})] &= \mathbb{E}[p(j)] - \omega \mathbb{E}[p(j) \mathbb{I}_{m \geq h}] \\ &= T_0 - \omega T_h. \end{aligned} \tag{31}$$

For a focal defector, the corresponding residual risk expression is analogous, yielding $T_0^D - \omega T_h^D$.

A focal ordinary cooperator explicitly requires $m \geq M$ for activation. Combining the expected public-good return with the residual risk derived in Eq. (31) gives

$$P_C = \frac{rc[1 + n(x + z)]}{N} - c - L[T_0 - \omega T_M].$$

Conversely, a focal defector faces the same activation condition but strictly alters both the contribution count and the failure probability:

$$P_D = \frac{rcn(x + z)}{N} - L[T_0^D - \omega T_M^D].$$

Finally, a focal protective cooperator intrinsically counts toward the required threshold, therefore needing only $M - 1$ protective co-players to trigger the mechanism:

$$P_S = \frac{rc[1 + n(x + z)]}{N} - c - k - L[T_0 - \omega T_{M-1}].$$

These derivations successfully reproduce the expected payoffs in Eq. (5) presented in the main text.

Appendix B. Pivotality identity and closed payoff differences

B.1. Risk-weighted pivotality

The terms T_{M-1} and T_M differ only in the specific configuration where $m = M - 1$. Therefore, their difference directly isolates the pivotal probability:

$$\begin{aligned} T_{M-1} - T_M &= \mathbb{E}[p(j) \mathbb{I}_{m=M-1}] \\ &= \binom{n}{M-1} z^{M-1} [(1-z)^{n-M+1} - (x + \rho y)^{n-M+1}] \\ &= \Psi_M. \end{aligned} \tag{32}$$

This formally proves the expansion of Ψ_M in Eq. (6). Furthermore, subtracting the expected payoff of ordinary cooperation from that of protective cooperation yields

$$P_S - P_C = -k + L\omega(T_{M-1} - T_M) = -k + L\omega\Psi_M, \tag{33}$$

which proves Eq. (7).

B.2. Closed payoff differences

The survival-weighted activation term R_M initially defined in Section 2.5 admits the finite-sum representation

$$R_M(x, y, z) = \sum_{m=M}^n \binom{n}{m} z^m (x + \rho y)^{n-m}.$$

Indeed, fixing m and summing the survival factor ρ^j over the remaining positions occupied by ordinary cooperators and defectors directly gives $(x + \rho y)^{n-m}$. By subtracting Eq. (29) from Eq. (30) evaluated at $h = M$, we obtain the risk difference

$$\begin{aligned} T_M^D - T_M &= (1 - \rho) \sum_{m=M}^n \binom{n}{m} z^m (x + \rho y)^{n-m} \\ &= (1 - \rho) R_M. \end{aligned}$$

For the gradient $A = P_C - P_D$, the public-good terms naturally reduce to the baseline margin $a = rc/N - c$, while the collective-loss terms expand to

$$\begin{aligned} A &= a - L [(T_0 - \omega T_M) - (T_0^D - \omega T_M^D)] \\ &= a + L(1 - \rho) [q^n - \omega R_M]. \end{aligned}$$

After explicitly substituting the simplex constraint $y = 1 - x - z$, this derivation yields Eq. (9). Finally, the gradient B is recovered via

$$\begin{aligned} B &= P_S - P_D = (P_C - P_D) + (P_S - P_C) \\ &= A - k + L\omega\Psi_M, \end{aligned}$$

which proves Eq. (10).

Appendix C. Boundary algebra and Jacobian calculations

C.1. General Jacobian

For notational readability, the explicit dependence of the functions A , B , F , and G on the protection-effectiveness parameter ω is suppressed in this section. We write the planar system in Eq. (12) as the vector field

$$\begin{aligned} F(x, z) &= x [(1 - x)A(x, z) - zB(x, z)], \\ G(x, z) &= z [(1 - z)B(x, z) - xA(x, z)]. \end{aligned}$$

Direct partial differentiation of these components yields

$$\begin{aligned} F_x &= (1 - 2x)A + x(1 - x)A_x - zB - xzB_x, \\ F_z &= x [(1 - x)A_z - B - zB_z], \\ G_x &= z [(1 - z)B_x - A - xA_x], \\ G_z &= (1 - 2z)B - xA + z(1 - z)B_z - xzA_z. \end{aligned} \tag{34}$$

Thus, the continuous-time Jacobian matrix is defined as

$$J(x, z) = \begin{pmatrix} F_x & F_z \\ G_x & G_z \end{pmatrix}. \tag{35}$$

C.2. Jacobians at the homogeneous equilibria

At the full-defection state, the reduced coordinates reside at $(x, z) = (0, 0)$, yielding the base selection gradients $A(0, 0) = A_D$ and $B(0, 0) = A_D - k$. Substitution into the general components constructs the diagonal Jacobian matrix

$$J(0, 0) = \begin{pmatrix} A_D & 0 \\ 0 & A_D - k \end{pmatrix}, \tag{36}$$

whose diagonal entries inherently provide the invasion eigenvalues reported in Eq. (14).

Similarly, at the full-cooperation state, the coordinates shift to $(x, z) = (1, 0)$ with $A(1, 0) = A_C$ and $B(1, 0) = A_C - k$. Hence, the corresponding Jacobian evaluates to

$$J(1, 0) = \begin{pmatrix} -A_C & k - A_C \\ 0 & -k \end{pmatrix}, \quad (37)$$

whose structural diagonal entries reproduce Eq. (15).

Finally, at full protective cooperation, $(x, z) = (0, 1)$. Defining $A_S = a + L(1 - \rho)(1 - \omega)$, and observing that $\Psi_M = 0$ in the total absence of defectors, yields $B(0, 1) = A_S - k$. The resultant Jacobian becomes

$$J(0, 1) = \begin{pmatrix} k & 0 \\ -A_S & k - A_S \end{pmatrix}. \quad (38)$$

Its exact eigenvalues are determined to be k and $k - a - L(1 - \rho)(1 - \omega)$, strictly aligning with Eq. (16).

C.3. Nonhyperbolic homogeneous states

The zero-eigenvalue cases at the homogeneous states fundamentally require a boundary-flow argument to rigorously ascertain their stability.

On the invariant C - D edge, the restricted gradient defined as

$$A_{CD}(x) = a + L(1 - \rho)[\rho + (1 - \rho)x]^n$$

is a strictly increasing function of x . If $A_D = A_{CD}(0) = 0$, then $A_{CD}(x) > 0$ for every sufficiently small perturbation $x > 0$. Since the local dynamic flow satisfies $\dot{x} = x(1 - x)A_{CD}(x) > 0$, nearby states strictly move away from full defection. Thus, the full-defection state is classified as nonhyperbolic and structurally unstable when $A_D = 0$. Conversely, when $A_D = k$, the complementary eigenvalue is strictly positive ($\lambda_C^D = k > 0$), rendering the state unstable regardless of the zero eigenvalue.

If $A_C = A_{CD}(1) = 0$, the strict monotonicity implies $A_{CD}(x) < 0$ for any $x < 1$ sufficiently close to unity. Consequently, $\dot{x} = x(1 - x)A_{CD}(x) < 0$, forcing nearby edge states to move away from full cooperation. The state is therefore topologically nonhyperbolic and unstable when $A_C = 0$.

At full protective cooperation, the invasion eigenvalue of an ordinary cooperator is $\lambda_C^S = k > 0$ across every maintained parameter regime. Hence, this state fundamentally remains unstable even when its second eigenvalue vanishes.

C.4. The C - D edge

On the boundary $z = 0$, we have $y = 1 - x$, and the activation terms structurally vanish because $M \geq 2$. Equation (9) therefore reduces to

$$A_{CD}(x) := A(x, 0) = a + L(1 - \rho)[\rho + (1 - \rho)x]^n.$$

Its derivative with respect to x is strictly positive, explicitly evaluating to

$$A'_{CD}(x) = nL(1 - \rho)^2[\rho + (1 - \rho)x]^{n-1} > 0.$$

Because the endpoints naturally satisfy $A_{CD}(0) = A_D$ and $A_{CD}(1) = A_C$, a unique nontrivial edge equilibrium emerges precisely when $A_D < 0 < A_C$. Analytically solving the condition $A_{CD}(x) = 0$ yields the coordinate

$$x_{CD} = \frac{[-a/(L(1 - \rho))]^{1/n} - \rho}{1 - \rho}.$$

At this equilibrium, $A(x_{CD}, 0) = 0$ and $B(x_{CD}, 0) = -k$, and the Jacobian is upper triangular:

$$J(x_{CD}, 0) = \begin{pmatrix} x_{CD}(1 - x_{CD})A_x(x_{CD}, 0) & * \\ 0 & -k \end{pmatrix}. \quad (39)$$

The tangential eigenvalue is strictly positive while the transverse eigenvalue is negative, verifying that the interior edge equilibrium E_{CD} constitutes a topological saddle.

C.5. The excluded zero-cost limit

The standard formulation of the model assumes a strictly positive operating cost ($k > 0$). In the hypothetical limit where $k = 0$, the system constrained to the C – S edge features $y = 0$, $p(0) = 0$, and $\Psi_M = 0$. Therefore, the payoff gap closes ($P_S - P_C = 0$), implying $B = A$. Given the simplex identity $x + z = 1$ along this boundary, the planar vector field completely vanishes:

$$\begin{aligned} F(x, z) &= x(1 - x - z)A = 0, \\ G(x, z) &= z(1 - x - z)A = 0. \end{aligned}$$

Hence, every point on the C – S edge is an equilibrium when $k = 0$. This gives a nonhyperbolic continuum whose transverse stability to defector invasion may vary along the edge.

C.6. The D – S edge

At any nontrivial equilibrium satisfying the root condition $B(0, z_{DS}; \omega) = 0$, the general Jacobian matrix derived in Eq. (34) simplifies to

$$J(0, z_{DS}; \omega) = \begin{pmatrix} A(0, z_{DS}; \omega) & 0 \\ * & z_{DS}(1 - z_{DS})B_z(0, z_{DS}; \omega) \end{pmatrix}. \quad (40)$$

Its diagonal entries are the transverse and tangential eigenvalues in Eq. (17).

C.7. Boundary saddle–node reduction

Holding all parameters other than ω fixed, the dynamics on the D – S edge are governed by the scalar field

$$\dot{z} = \mathcal{G}(z, \omega) := z(1 - z)B(0, z; \omega). \quad (41)$$

At a nontrivial boundary equilibrium satisfying $B(0, z; \omega) = 0$, the first derivative reduces to $\mathcal{G}_z = z(1 - z)B_z(0, z; \omega)$. Because $0 < z < 1$, the fold condition $\mathcal{G}_z = 0$ is equivalent to $B_z = 0$.

Evaluating a candidate fold that simultaneously satisfies $B = B_z = 0$, the parametric and higher-order spatial derivatives expand as

$$\begin{aligned} \mathcal{G}_\omega &= z(1 - z)B_\omega(0, z; \omega), \\ \mathcal{G}_{zz} &= z(1 - z)B_{zz}(0, z; \omega). \end{aligned}$$

Since $z(1 - z) > 0$ for $0 < z < 1$, the conditions $\mathcal{G}_\omega \neq 0$ and $\mathcal{G}_{zz} \neq 0$ are equivalent to $B_\omega(0, z; \omega) \neq 0$ and $B_{zz}(0, z; \omega) \neq 0$.

The second eigenvalue of the planar system is $\lambda_\perp = A(0, z; \omega)$. The condition $\lambda_\perp \neq 0$ ensures that the tangential zero eigenvalue is simple; if $\lambda_\perp < 0$, perturbations transverse to the invariant edge decay locally.

C.8. Interior equilibria

At a fully polymorphic interior equilibrium, the selection gradients vanish ($A = B = 0$). Substituting these conditions directly into Eq. (34) gives the specific Jacobian

$$J_* = \begin{pmatrix} x^*[(1 - x^*)A_x - z^*B_x] & x^*[(1 - x^*)A_z - z^*B_z] \\ z^*[(1 - z^*)B_x - x^*A_x] & z^*[(1 - z^*)B_z - x^*A_z] \end{pmatrix}_{E_*}. \quad (42)$$

This matrix factorizes as

$$J_* = \begin{pmatrix} x^* & 0 \\ 0 & z^* \end{pmatrix} \begin{pmatrix} 1 - x^* & -z^* \\ -x^* & 1 - z^* \end{pmatrix} \begin{pmatrix} A_x & A_z \\ B_x & B_z \end{pmatrix}_{E_*}. \quad (43)$$

The determinant of the central matrix is $(1 - x^*)(1 - z^*) - x^*z^* = 1 - x^* - z^* = y^*$. Hence,

$$\det J_* = x^*y^*z^* (A_x B_z - A_z B_x)_{E_*}, \quad (44)$$

which proves Eq. (22). The trace is

$$\begin{aligned} \text{tr } J_* &= x^* [(1 - x^*)A_x - z^*B_x]_{E_*} \\ &\quad + z^* [(1 - z^*)B_z - x^*A_z]_{E_*}. \end{aligned} \quad (45)$$

By the Routh–Hurwitz criteria, a hyperbolic planar equilibrium is locally asymptotically stable if and only if $\det J_* > 0$ and $\text{tr } J_* < 0$. If $\det J_* < 0$, the two real eigenvalues have opposite signs and the equilibrium is a saddle.

Appendix D. Numerical branch tracking and bifurcation diagnostics

All numerical calculations were performed in Python using NumPy and SciPy for root finding and linear algebra, pandas for tabular output, and Matplotlib for visualization. Equilibria were computed directly from the closed payoff differences in Eqs. (9) and (10); no Monte Carlo approximation of random group composition was used.

D.1. Parameter-stepped equilibrium searches

For each fixed value of ω , interior equilibria were obtained by solving

$$\begin{aligned} A(x, z; \omega) &= 0, \\ B(x, z; \omega) &= 0, \end{aligned}$$

subject to $x > 0$, $z > 0$, and $x + z < 1$. We used the Powell hybrid method with solver tolerance 10^{-11} and initial guesses from the admissible points of a 13×13 grid with coordinates from 0.025 to 0.925. A solution was retained when all three strategy frequencies exceeded 10^{-8} and the residual norm was below 10^{-8} . Roots separated by less than 10^{-6} in Euclidean distance were merged.

Nontrivial equilibria on the invariant D – S edge satisfy

$$B(0, z; \omega) = 0, \quad 0 < z < 1. \quad (46)$$

For the one-parameter branch diagram, the interval $(10^{-7}, 1 - 10^{-7})$ was sampled at 1800 uniformly spaced points. Each sign-changing bracket was refined with Brent's method, and roots separated by less than 10^{-5} were merged. Because a sign-change search may miss a double root at a fold, saddle–node coordinates were computed separately from the augmented equations given below.

Figure 1 uses $\omega = 0, 0.002, \dots, 1$ together with the independently refined bifurcation values. At each value of ω , we applied both the interior seed grid and the boundary scan. **The procedure relies on a parameter-stepped root search combined with numerical branch matching, rather than formal pseudo-arclength continuation. Consequently, the reported branch set remains strictly conditional on the predefined grids and tolerances; it represents a comprehensive numerical detection rather than an analytical guarantee of global nonexistence.**

The reduced planar Jacobian was evaluated by centered finite differences with step size $2 \cdot 10^{-6}$. An equilibrium was classified as locally asymptotically stable when both eigenvalues had real parts below -10^{-7} and as unstable when both real parts exceeded 10^{-7} . It was classified as a saddle when the product of the two eigenvalue real parts was below -10^{-10} . Cases with a near-zero eigenvalue were treated as nonhyperbolic and examined using the augmented bifurcation conditions.

D.2. Interior saddle–node diagnostics

At a verified interior equilibrium, the determinant factorization previously detailed in Eq. (22) provides the compact representation $\det J = xyz \det D(A, B)$, where the structural submatrix is defined as

$$D(A, B) = \begin{pmatrix} A_x & A_z \\ B_x & B_z \end{pmatrix}.$$

Because $x, y, z > 0$ at an interior equilibrium, $\det J = 0$ is equivalent to $\det D(A, B) = 0$. We therefore located SN1 by solving

$$\begin{aligned} A(x, z; \omega) &= 0, \\ B(x, z; \omega) &= 0, \\ \det D(A, B) &= 0. \end{aligned} \quad (47)$$

The derivatives in $D(A, B)$ were evaluated by centered finite differences with step size $2 \cdot 10^{-5}$.

The solution is

$$(x^*, y^*, z^*, \omega_{\text{SN1}}) = (0.2994, 0.4272, 0.2734, 0.2882).$$

At this point, the planar Jacobian has one simple zero eigenvalue and one positive nonzero eigenvalue, $\lambda_{\text{nz}} = 0.5299$.

Let v_0 and w_0 denote the right and left null vectors of the Jacobian at SN1, normalized by $w_0^\top v_0 = 1$. For the planar vector field $f(X, \omega)$ with $X = (x, z)^\top$, the saddle-node coefficients are

$$\alpha = w_0^\top f_\omega, \quad \beta = \frac{1}{2} w_0^\top D_X^2 f[v_0, v_0].$$

Centered finite differences with step size 10^{-4} give

$$\alpha_{\text{SN1}} = -0.2871, \quad \beta_{\text{SN1}} = 0.4826. \quad (48)$$

Both coefficients are nonzero, confirming that SN1 is a nondegenerate saddle-node bifurcation. Since $-\alpha_{\text{SN1}}/\beta_{\text{SN1}} > 0$, the two local branches exist for $\omega > \omega_{\text{SN1}}$. The positive nonzero eigenvalue implies that the pair consists of a saddle and an unstable node.

Repeating the calculation with finite-difference steps $4 \cdot 10^{-4}$, $2 \cdot 10^{-4}$, $5 \cdot 10^{-5}$, and $2.5 \cdot 10^{-5}$ preserved the four-decimal values in Eq. (48).

D.3. Boundary saddle-node and edge-transcritical diagnostics

The boundary saddle-node SN2 was located by solving

$$B(0, z; \omega) = 0, \quad B_z(0, z; \omega) = 0. \quad (49)$$

We approximated B_z by centered finite differences with step size $2 \cdot 10^{-5}$. The resulting fold is

$$(z^*, \omega_{\text{SN2}}) = (0.2718, 0.5254).$$

At this point, the boundary diagnostics are

$$\begin{aligned} B(0, z^*; \omega_{\text{SN2}}) &= 0, \\ B_z(0, z^*; \omega_{\text{SN2}}) &= 0, \\ A(0, z^*; \omega_{\text{SN2}}) &= -0.469, \\ B_\omega(0, z^*; \omega_{\text{SN2}}) &= 1.5511, \\ B_{zz}(0, z^*; \omega_{\text{SN2}}) &= -14.6775. \end{aligned}$$

The nonzero value of A shows that the transverse eigenvalue does not vanish at SN2, while $B_\omega \neq 0$ and $B_{zz} \neq 0$ establish the nondegeneracy of the boundary fold.

For the scalar edge field $\mathcal{G}(z, \omega) = z(1 - z)B(0, z; \omega)$, centered differences with step size 10^{-4} give

$$\alpha_{\text{SN2}} = \mathcal{G}_\omega = 0.307, \quad \beta_{\text{SN2}} = \frac{1}{2} \mathcal{G}_{zz} = -1.4525.$$

Since $-\alpha_{\text{SN2}}/\beta_{\text{SN2}} > 0$, the two boundary branches exist for $\omega > \omega_{\text{SN2}}$. Their stability follows from $\lambda_\perp = A(0, z; \omega)$ and $\lambda_\parallel = z(1 - z)B_z(0, z; \omega)$. The lower- z branch has $\lambda_\parallel > 0$ and $\lambda_\perp < 0$ and is therefore a saddle. Both eigenvalues are negative on the upper- z branch, which is locally asymptotically stable within the simplex.

The unstable nontrivial boundary branch intersects the pure- S branch at $\omega_{\text{TC}} = 0.6880820$, where

$$\begin{aligned} B(0, 1; \omega_{\text{TC}}) &= 0, \\ B_z(0, 1; \omega_{\text{TC}}) &= 2.8328, \\ B_\omega(0, 1; \omega_{\text{TC}}) &= -3.0136. \end{aligned}$$

The pure- S equilibrium exists for every ω , and the nontrivial boundary branch crosses it transversely. The collision is therefore an edge transcritical bifurcation. Because the second planar eigenvalue is $0.4000 > 0$, this bifurcation does not create a locally attracting equilibrium in the full simplex.

D.4. Discrete activation-threshold comparison

Because M is discrete, we examined each value in a separate boundary calculation. For $M = 2$, the locally asymptotically stable boundary branch is created at the saddle-node defined by Eq. (49). For $M = 3$ and $M = 4$, a nontrivial boundary branch exists before it becomes locally asymptotically stable. The corresponding stability thresholds were obtained from

$$\begin{aligned} B(0, z; \omega) &= 0, \\ A(0, z; \omega) &= 0. \end{aligned}$$

These equations locate a zero transverse ordinary-cooperator invasion eigenvalue on an existing D - S branch.

For $M = 3$, the threshold is $(z_c, \omega_c) = (0.7935, 0.6608)$ and the tangential eigenvalue is -0.1904 . For $M = 4$, the threshold is $(z_c, \omega_c) = (0.9541, 0.8321)$ and the tangential eigenvalue is -0.3502 . Only the transverse eigenvalue vanishes at these points, so we classify them as transverse stability crossings rather than boundary saddle-node bifurcations. A more specific codimension-one classification would require an additional normal-form calculation and tracking of the associated interior branch.

D.5. Pointwise classification of two-parameter maps

The two-parameter maps in Fig. 3 were constructed by classifying each sampled parameter pair. We used 401 equally spaced values of ω on $[0, 1]$ and 241 equally spaced values of the secondary parameter over $0.15 \leq k \leq 0.75$, $2.5 \leq L \leq 6$, or $0.35 \leq \gamma \leq 2.5$.

On the D - S edge, the closed payoff differences are polynomials in z and affine in ω . For each fixed value of the secondary parameter $\eta \in \{k, L, \gamma\}$, we wrote them as

$$\begin{aligned} A(0, z; \omega, \eta) &= A_0(z; \eta) + \omega A_1(z; \eta), \\ B(0, z; \omega, \eta) &= B_0(z; \eta) + \omega B_1(z; \eta). \end{aligned}$$

At each sampled parameter pair, we computed all roots of the degree- $(N - 1)$ polynomial $B(0, z; \omega, \eta)$. A root was retained when the magnitude of its imaginary part was below 10^{-8} , $10^{-9} < z < 1 - 10^{-9}$, and the scaled polynomial residual was below 10^{-7} .

For each retained boundary root, local stability in the full simplex was evaluated from $\lambda_{\perp} = A(0, z; \omega, \eta)$ and $\lambda_{\parallel} = z(1 - z)B_z(0, z; \omega, \eta)$. A parameter pair was classified as “present” when at least one nontrivial boundary root satisfied $\lambda_{\perp} < -10^{-7}$ and $\lambda_{\parallel} < -10^{-7}$; otherwise it was classified as “absent”. This pointwise procedure permits multiple stable intervals at the same secondary-parameter value and does not assume that the stable region occupies one side of a single fold.

The solid contours in Fig. 3 were extracted from changes in the binary pointwise classifications. They are grid-resolved numerical stability boundaries, not analytically continued saddle-node loci. The complete classification grids, including the locally stable roots, their eigenvalues, and disconnected contour segments, were exported as CSV files.

D.6. Basin computation and convergence checks

The basin map in Fig. 4 was computed at $\omega = 0.90$ on a uniform barycentric grid with scale factor 300, giving 45 451 initial states. The planar system was integrated by the classical fourth-order Runge-Kutta method with $\Delta t = 0.10$, 1800 steps, and $t_{\max} = 180$.

To classify the terminal states, we exclusively relied on the candidate attractors established during the preceding equilibrium and local-stability analysis. Rather than identifying new attractors independently during the grid simulation, we utilized these pre-calculated states as fixed reference points for the basin assignment. Specifically at $\omega = 0.90$, their reduced (x, z) coordinates are

$$\mathbf{D} = (0, 0), \quad \mathbf{C} = (1, 0), \quad E_{DS}(0.90) = (0, 0.5008).$$

For a terminal state (x_f, z_f) , we defined

$$d_{\min} = \min_{E \in \{\mathbf{D}, \mathbf{C}, E_{DS}\}} \|(x_f, z_f) - E\|_2.$$

The initial condition was assigned to the nearest candidate attractor only when $d_{\min} < 10^{-6}$; otherwise it was recorded as unclassified. The exact pure- S initial state was retained as unclassified because it remains at the unstable pure- S equilibrium and does not approach any of the three attractors used in the basin classification.

We checked convergence on the same grid using $\Delta t = 0.05$ with 3600 steps and $t_{\max} = 180$, and $\Delta t = 0.10$ with 3000 steps and $t_{\max} = 300$. All three integration protocols assigned the same 45 450 initial states, with only the exact pure- S state unclassified. No grid point changed basin assignment under either refinement. This finite-grid agreement supports the numerical robustness of Fig. 4 for the reported parameters but does not constitute a global basin or convergence theorem.

Acknowledgements

This research was supported by the National Philosophy and Social Science Foundation of China (Grant No. 22ZD158), the Yunnan Fundamental Research Projects (Grant Nos. 202501AT070453, 202601CI070171), the Central Guidance Fund for Local Science and Technology Development (No. 202407AB110006), the Yunnan Science and Technology Talent Program (No. 202305AC160049), the Research Fund of Yunnan University of Finance and Economics (Grant Nos. 2025D05, 2025D59), and the National Research, Development and Innovation Office (NKFIH) under Grant No. K142948.

CRedit authorship contribution statement

Shulan Li: Conceptualization, Methodology, Formal analysis, Writing – original draft. Xiaogang Li: Software, Validation, Visualization, Writing – review and editing. Lei Shi: Supervision, Project administration, Funding acquisition, Writing – review and editing. Jun He: Investigation, Formal analysis, Validation. Attila Szolnoki: Funding acquisition, Writing – review and editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data and code availability

No empirical datasets were used in this study. All reported results were obtained from analytical derivations and numerical solutions of the model. The custom code used to generate the numerical results and figures is available from the corresponding author upon reasonable request.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the authors used ChatGPT (OpenAI) and Kimi (Moonshot AI) in order to improve the readability and language of the manuscript and to assist with LaTeX formatting. After using these tools, the authors reviewed and edited the content as needed and take full responsibility for the content of the publication.

References

- [1] R. Axelrod, W. D. Hamilton, The Evolution of Cooperation, *Science* 211 (1981) 1390–1396.
- [2] J. Maynard Smith, Evolution and the theory of games, in: *Did Darwin get it right? Essays on games, sex and evolution*, Springer, 1982, pp. 202–215.
- [3] G. Szabó, G. Fáth, Evolutionary games on graphs, *Physics reports* 446 (2007) 97–216.
- [4] A. Szolnoki, M. Perc, Reward and cooperation in the spatial public goods game, *EPL (Europhysics Letters)* 92 (2010) 38003.
- [5] G. Hardin, The Tragedy of the Commons, *Science* 162 (1968) 1243–1248.
- [6] E. Ostrom, *Governing the commons: The evolution of institutions for collective action*, Cambridge Univ. Press, 1990.
- [7] M. Perc, J. Gómez-Gardeñes, A. Szolnoki, L. M. Floría and Y. Moreno, Evolutionary dynamics of group interactions on structured populations: a review, *J. R. Soc. Interface* 10 (2013) 20120997.
- [8] M. J. W. Eberhard, The evolution of social behavior by kin selection, *The Quarterly Review of Biology* 50 (1975) 1–33.
- [9] R. L. Trivers, The evolution of reciprocal altruism, *Q. Rev. Biol.* 46 (1971) 35–57.
- [10] M. A. Nowak, K. Sigmund, Evolution of indirect reciprocity, *Nature* 437 (2005) 1291–1298.
- [11] M. A. Nowak, R. M. May, Evolutionary games and spatial chaos, *Nature* 359 (1992) 826–829.
- [12] D. S. Wilson, A theory of group selection, *Proc. Nat. Acad. Sci. USA* 72 (1975) 143–146.
- [13] M. A. Nowak, Five rules for the evolution of cooperation, *Science* 314 (2006) 1560–1563.
- [14] S. He, Q. Li, M. Feng, A. Szolnoki, Reputation assimilation mechanism for sustaining cooperation, *Chaos, Solitons and Fractals* 202 (2026) 117586.
- [15] C. Wang, W. Zhu, A. Szolnoki, The conflict between self-interaction and updating passivity in the evolution of cooperation, *Chaos, Solitons and Fractals* 173 (2023) 113667.
- [16] H. Yue, X. Xiong, M. Feng, A. Szolnoki, Dynamic evolution of cooperation based on adaptive reputation threshold and game transition, *Chaos, Solit. and Fract.* 199 (2025) 116693.
- [17] E. Fehr, S. Gächter, Cooperation and punishment in public goods experiments, *American Economic Review* 90 (2000) 980–994.
- [18] R. Boyd, H. Gintis, S. Bowles, P. J. Richerson, The evolution of altruistic punishment, *Proceedings of the National Academy of Sciences* 100 (2003) 3531–3535.
- [19] D. G. Rand, A. Dreber, T. Ellingsen, D. Fudenberg, M. A. Nowak, Positive interactions promote public cooperation, *Science* 325 (2009) 1272–1275.
- [20] C. Hauert, A. Traulsen, H. Brandt, M. A. Nowak, K. Sigmund, Via freedom to coercion: the emergence of costly punishment, *Science* 316 (2007) 1905–1907.
- [21] A. Szolnoki, M. Perc, Second-order free-riding on antisocial punishment restores the effectiveness of prosocial punishment, *Physical Review X* 7 (2017) 041027.
- [22] K. Panchanathan, R. Boyd, Indirect reciprocity can stabilize cooperation without the second-order free rider problem, *Nature* 432 (2004) 499–502.
- [23] C. Hauert, S. De Monte, J. Hofbauer, K. Sigmund, Volunteering as red queen mechanism for cooperation in public goods games, *Science* 296 (2002) 1129–1132.
- [24] C. Shen, M. Jusup, L. Shi, Z. Wang, M. Perc, P. Holme, Exit rights open complex pathways to cooperation, *Journal of the Royal Society Interface* 18 (2021) 20200777.
- [25] S. Li, Z. He, D. Jia, C. Shen, L. Shi, J. Tanimoto, Granting leaders priority exit options promotes and jeopardizes cooperation in social dilemmas, *Neurocomputing* 583 (2024) 127566.
- [26] T. Sasaki, S. Uchida, The evolution of cooperation by social exclusion, *Proceedings of the Royal Society B: Biological Sciences* 280 (2013) 20122498.
- [27] K. Li, R. Cong, T. Wu, L. Wang, Social exclusion in finite populations, *Physical Review E* 91 (2015) 042810.
- [28] S. Li, C. Du, X. Li, C. Shen, L. Shi, Antisocial peer exclusion does not eliminate the effectiveness of prosocial peer exclusion in structured populations, *Journal of Theoretical Biology* 576 (2024) 111665.
- [29] M. Perc, J. J. Jordan, D. G. Rand, Z. Wang, S. Boccaletti, A. Szolnoki, Statistical physics of human cooperation, *Physics Reports* 687 (2017) 1–51.
- [30] C. Wang, W. Zhang, X. Wang, A. Szolnoki, Inter-role reciprocity in evolutionary trust game on square lattices, *Chaos* 35 (2025) 083117.
- [31] R. T. Croson, M. B. Marks, Step returns in threshold public goods: A meta-and experimental analysis, *Experimental Economics* 2 (2000) 239–259.
- [32] M. Milinski, D. Semmann, H.-J. Krambeck, J. Marotzke, Stabilizing the earths climate is not a losing game: Supporting evidence from public goods experiments, *Proceedings of the National Academy of Sciences* 103 (2006) 3994–3998.
- [33] M. Milinski, R. D. Sommerfeld, H.-J. Krambeck, F. A. Reed, J. Marotzke, The collective-risk social dilemma and the prevention of simulated dangerous climate change, *Proceedings of the National Academy of Sciences* 105 (2008) 2291–2294.
- [34] S. Barrett, A. Dannenberg, Climate negotiations under scientific uncertainty, *Proceedings of the National Academy of Sciences* 109 (2012) 17372–17376.
- [35] J. M. Pacheco, F. C. Santos, M. O. Souza, B. Skyrms, Evolutionary dynamics of collective action in n-person stag hunt dilemmas, *Proceedings of the Royal Society B: Biological Sciences* 276 (2008) 315.
- [36] J. Wang, F. Fu, T. Wu, L. Wang, Emergence of social cooperation in threshold public goods games with collective risk, *Physical Review E* 80 (2009) 016101.
- [37] F. C. Santos, J. M. Pacheco, Risk of collective failure provides an escape from the tragedy of the commons, *Proceedings of the National Academy of Sciences* 108 (2011) 10421–10425.
- [38] X. Chen, A. Szolnoki, M. Perc, Risk-driven migration and the collective-risk social dilemma, *Physical Review E* 86 (2012) 036101.
- [39] X. Chen, A. Szolnoki, M. Perc, Averting group failures in collective-risk social dilemmas, *Europhysics Letters* 99 (2012) 68003.

- [40] A. Tavoni, A. Dannenberg, G. Kallis, A. Löschel, Inequality, communication, and the avoidance of disastrous climate change in a public goods game, *Proceedings of the National Academy of Sciences* 108 (2011) 11825–11829.
- [41] V. V. Vasconcelos, F. C. Santos, J. M. Pacheco, S. A. Levin, Climate policies under wealth inequality, *Proceedings of the National Academy of Sciences* 111 (2014) 2212–2216.
- [42] O. P. Hauser, D. G. Rand, A. Peysakhovich, M. A. Nowak, Cooperating with the future, *Nature* 511 (2014) 220–223.
- [43] J. Zhang, C. Zhang, M. Cao, How insurance affects altruistic provision in threshold public goods games, *Scientific Reports* 5 (2015) 9098.
- [44] W. Barfuss, J. F. Donges, V. V. Vasconcelos, J. Kurths, S. A. Levin, Caring for the future can turn tragedy into comedy for long-term collective action under risk of collapse, *Proceedings of the National Academy of Sciences* 117 (2020) 12915–12922.
- [45] L. Liu, X. Chen, A. Szolnoki, Coevolutionary dynamics via adaptive feedback in collective-risk social dilemma game, *Elife* 12 (2023) e82954.
- [46] S. Hua, M. Xu, L. Liu, X. Chen, Coevolutionary dynamics of collective cooperation and dilemma strength in a collective-risk game, *Physical Review Research* 6 (2024) 023313.
- [47] Z. Wang, M. Jusup, H. Guo, L. Shi, S. Geček, M. Anand, M. Perc, C. T. Bauch, J. Kurths, S. Boccaletti, et al., Communicating sentiment and outlook reverses inaction against collective risks, *Proceedings of the National Academy of Sciences* 117 (2020) 17650–17655.
- [48] N. He, X. Chen, A. Szolnoki, Central governance based on monitoring and reporting solves the collective-risk social dilemma, *Applied Mathematics and Computation* 347 (2019) 334–341.
- [49] L. Wang, S. Hua, Y. Liu, Z. Lu, L. Zhang, L. Liu, A. Szolnoki, Strategic competition in informal risk-sharing mechanisms versus collective index insurance, *Journal of the Royal Society Interface* 22 (2025) 20250163.
- [50] S. Hua, L. Liu, A. Szolnoki, Adaptive target regulation for multi-agent cooperation in collective-risk dilemmas, *Expert Systems with Applications* 324 (2026) 132471.
- [51] M. Xu, S. Hua, L. Liu, A. Szolnoki, Adaptive governance in collective risk social dilemmas, *Communications Physics* 9 (2026) 177.
- [52] L. Wang, S. Hua, Y. Liu, L. Zhang, L. Liu, A. Szolnoki, Coevolutionary dynamics of cooperation, risk, and cost in collective risk games, *PLoS Computational Biology* 22 (2026) e1013512.
- [53] P. D. Taylor, L. B. Jonker, Evolutionary stable strategies and game dynamics, *Mathematical Biosciences* 40 (1978) 145–156.
- [54] J. Hofbauer, K. Sigmund, *Evolutionary games and population dynamics*, Cambridge University Press, 1998.
- [55] T. A. Han, Z. Song, T. Cimpéanu, M. H. Duong, M. Krellner, V. Capraro, M. Perc, Cooperation versus social welfare, *Physics of life reviews* (2025).