RESEARCH ARTICLE

The strategy dynamics of collective systems: Underlying hindrances beyond two-actor coordination

Ambrosio Valencia-Romero[1*] and Paul T. Grogan[2]

1 The Roux Institute at Northeastern University, Portland, ME 04101, United States
2 Arizona State University, School of Computing and Augmented Intelligence, Tempe, AZ, 85281, U.S.A.

* a.valenciaromero@northeastern.edu

Abstract

Engineering systems, characterized by their high technical complexity and societal intricacies, require a strategic design approach to navigate multifaceted challenges. Understanding the circumstances that affect strategic action in these systems is crucial for managing complex real-world challenges. These challenges go beyond localized coordination issues and encompass intricate dynamics, requiring a deep understanding of the underlying structures impacting strategic behaviors, the interactions between subsystems, and the conflicting needs and expectations of diverse actors. Traditional optimization and game-theoretic approaches to guide individual and collective decisions need adaptation to capture the complexities of these design ecosystems, particularly in the face of increasing numbers of decision-makers and various interconnections between them. This paper presents a framework for studying strategic decision-making processes in collective systems. It tackles the combinatorial complexity and interdependencies inherent in large-scale systems by representing strategic decision-making processes as binary normal-form games, then dissects and reinterprets them in terms of multiple compact games characterized by two real-numbered structural factors and classifies them across four strategy dynamical domains associated with different stability conditions. We provide a mathematical characterization and visual representation of emergent strategy dynamics in games with three or more actors intended to facilitate its implementation by researchers and practitioners and elicit new perspectives on design and management for optimizing systems-of-systems performance. We conclude this paper with a discussion of the opportunities and challenges of adopting this framework within and beyond the context of engineering systems.

1 Introduction

Understanding emergence in large-scale systems-of-systems has become a prominent feature in addressing complex real-world challenges [1]. Such collective systems entail the integration of interconnected entities, each with a level of operational and managerial autonomy from the others, while collectively contributing to the overarching system's objectives. Collaborative decision-making plays a pivotal role in facilitating information exchange and coordinating the activities of individual actors to guide the system toward optimized performance. However, despite the potential benefits of coordinated collective action, its seamless execution faces structural challenges related to the combinatorial complexity of aligning individual objectives and incentives [2]. These hindrances go beyond localized actor-to-actor coordination challenges and encompass intricate dynamics that result from the interplay of many subsystems, direct and

indirect stakeholders, and decision-making agents. Addressing these challenges necessitates a deep understanding of the underlying structure of the collective decision-making process that impacts the strategic behaviors and interactions within the system.

Engineering systems are a subset of collective systems marked by high technical complexity, social intricacy, and elaborate processes geared toward fulfilling significant societal functions [3]. While general collective systems refer to groups of agents collaborating to produce a collective behavior that transcends the sum of individual contributions [4], the nuanced nature of engineering systems implies the need for a strategic design approach to navigate their multifaceted challenges. The complexity within engineering systems becomes evident in endeavors such as aircraft design, where societal needs and expectations, communicated through regulations, can directly impact lower-level technical decisions [5]. The evolving concept of complexity within engineering systems necessitates ongoing efforts to define, measure, and comprehend its intricate nature. Challenges arise in understanding and modeling emergence, a fundamental characteristic of complex collective systems, which manifests in various forms within engineering contexts, such as non-linear emergent properties and those arising from interdependent choices in multi-agent systems [6].

The strategic design of engineering systems seeks to tackle the complex interplay between technical and social dimensions. In the technical realm, the challenge lies in comprehending interactions between subsystems while maintaining their levels of operational and managerial independence [7]. Simultaneously, the social dimension of engineering systems involves diverse actors, categorized into users, stakeholders, and societal entities. The conflicting needs and expectations of these actors amplify the complexity of the system design [3]. The strategic challenges embedded in engineering systems design, particularly concerning the emergence of unfavorable dynamics and the modeling of interactive decisions among multiple actors, highlight the need for advanced methodologies. Traditional game-theoretic approaches, often designed for two-player scenarios, must be revised to capture the intricacies of complex design ecosystems [5]. The classification of emergence based on non-linearities and multi-agent systems guides the selection of appropriate methods for system-level governance, necessitating shifts in design perspectives [6]. As engineering systems become more complex, a comprehensive and adaptable framework is increasingly necessary to address strategic challenges in their intricate design landscapes.

Game theory is a valuable framework to address the strategic challenges embedded in engineering systems. As the mathematical study of strategic interactions among independent actors, the last four decades have seen an increasing number of researchers employing game-theoretical methods to model decision dynamics in multidisciplinary engineering systems, treating different systems actors as distinct players engaged in a sequence of games throughout the design process [8–10]. Recent literature shows applications of game theory to study modularity as a mechanism for facilitating cooperation in systems-of-systems [11], develop computational models to assist engineering systems design under competition [12], and investigate the link between risk attitudes and strategic decisions in the context of collective systems [13]. Existing formulations emphasize dynamics among two actors, for which extensive work exists centered on Nash equilibria in social dilemma games. However, as design ecosystems become more intricate, existing game-theoretic approaches need to be revised in modeling the complex interactions among an increasing number of design actors [5]. Broader classes of engineered systems — including inter-agency, international, or public-private partnerships — exhibit strategy dynamics among more than two actors. More existing work is needed to formulate and understand the emergent dynamics in strategic scenarios with three or more actors due partly to a combinatorial growth of possible outcomes.

We turn to non-cooperative game theory to comprehend the nature of these strategy dynamics in collective systems. By adopting this analytical lens, we gain valuable insights into the strategic components associated with the stability of collective action and the risk of coordination failure among multiple decision-makers pursuing their self-interests. The

application of game theory in this context helps us analyze how individual motivations and incentives can influence collaborative decision-making processes and how this, in turn, impacts the system's overall performance and collective efficiency. Understanding and describing the interdependence of decisions among the actors in the system is essential to producing useful models of emergent behavior and anticipating potential outcomes that cannot be solely attributed to individual components. By investigating these dynamics, we can shed light on how collective systems adapt and respond to various stimuli, offering new perspectives on system design and management.

1.1 Motivating challenges

Understanding the emergence of strategies within complex collective systems poses significant challenges for engineers and system designers when evaluating a system's ability to integrate with others and align with common goals. This property, referred to as directional integrability [14], when extended to larger systems-of-systems, is linked to each subsystem's level of managerial control and collaboration [7]. Individual actors within the system face critical choices regarding the openness of their subsystems, and these choices have far-reaching implications for the system's overall performance. The decision to pursue an open strategy, facilitating full integration, or to maintain a closed system introduces complex trade-offs. Actors must weigh the potential benefits of collective integration against the costs and risks associated with opening their own systems. Although an open strategy might be optimal on a collective level, it might not be the most advantageous path for the individual actor.

The complexity escalates dramatically as the number of actors within the system increases. Decision-making becomes a tangled web of anticipating other actors' choices, potential coalitions, and the risks associated with incomplete information. Understanding the social dynamics at play becomes essential, as factors like communication, the way options are framed [15, 16], and actors' varying motivations and incentives [17] significantly influence the strategic landscape. A particularly worrying outcome in multi-actor systems is the potential for integration to halt. Imagine a scenario where a single actor initially takes the risk of opening their system, hoping others will follow. However, if others exploit this openness, keeping their systems closed and reaping the benefits without contributing, this can discourage further integration efforts. The system can easily become stuck in a situation where the benefits of wider integration seem out of reach, as no individual actor wants to be the only one taking on the burden of an open system.

To overcome these challenges, there's a critical need for new modeling approaches and analytical tools explicitly designed to handle the emergent dynamics of multi-actor scenarios. These models must move beyond traditional two-player game-theoretic frameworks to robustly capture complex interactions and coalition formation. Holistic decision-making approaches considering the technical, social, and strategic interdependencies are needed. Alongside sophisticated modeling, a focus on developing mechanisms to incentivize integration is paramount. Finding ways to align individual actor benefits with the goals of the collective system is essential to ensure sustained progress and avoid stagnation. Addressing these challenges is crucial for successfully engineering and managing complex systems-of-systems. A deep understanding of these emergent strategic dynamics and incentive structures will empower us to build systems that foster coordination, reduce incompatibilities, and ultimately facilitate the harmonious integration of interdependent subsystems for the greater good.

1.2 Our research endeavor

Our work focuses on studying complex strategy dynamics in collective systems, with an aim to comprehend the factors that influence the decisions and interactions of system actors. We acknowledge the complexity of these systems and strive to design mechanisms that promote

collaboration in large-scale systems-of-systems. The objective of this paper is to provide a general mathematical characterization and visual representation of emergent strategy dynamics in binary games that can be leveraged in the strategic design of engineering systems. It extends the existing concept of structural fear and greed in two-player games to any finite number of actors. These structural factors are presented on a Cartesian coordinate plane, with each quadrant defining a dynamical domain with specific strategic stability conditions.

The remainder of this paper is organized as follows. Section 2 recaps the mainstream treatment of strategy dynamics centered around normal-form games with two players and two available strategies each. Section 3 presents our extended characterization of strategy dynamics for binary strategic-form games with any number of players, with more details provided in the Appendix. Here we introduce the concept of *strategic hindrance*, with Section 3.3 applying the proposed framework to assess the evolution of strategy dynamics in the design for integrability of an urban transit system stylized as a volunteer's dilemma, and Section A.4 extending this framework to visually assess additional generic strategic settings (viz. public goods game, fated truel game, majority game, and others). Then, in Section 4, we address the assumptions and limitations of our proposed approach. We conclude with Section 5, summarizing contributions and our vision for future work.

2 Strategy dynamics in 2×2 social dilemmas

The concept of fear and greed strategy dynamics is based on the study of social dilemmas using game theory. A "game" is any decision-making process involving multiple independent actors with specific preferences over the decisions any of them could make and their consequences. We focus on normal-form games, the most common representations used to model social dilemmas.

2.1 Normal-form games

A normal-form game consists of

- A player set $\mathcal{N}$ consisting of $n \geq 1$ player names or other identifiers (ids);
  - for the sake of simplicity, we use $\mathcal{N} = \{1, \ldots, n\}$ and dummy ids $i, j, k \in \mathcal{N}$;
  - a game with $n = 1$ is referred to as a "trivial" decision problem.
- A finite set of *pure* individual strategies $\mathcal{S}_i = \{s_i^{(1)}, \ldots, s_i^{(m_i)}\} \ni s_i$, with $m_i \geq 2 \forall i \in \mathcal{N}$;
  - a pure strategy is a complete contingency plan that a player can execute in anticipation of the actions that the other players could take;
  - the product $\mathcal{S} = \mathcal{S}_1 \times \cdots \times \mathcal{S}_n$ is the set of all possible pure collective strategies $s = \langle s_1, \ldots, s_n \rangle$, with $|\mathcal{S}| = \prod_i^n m_i$;
  - an $n \times 2$ normal-form game (also known as *binary* game) is one in which $m_i = 2$ for every $i \in \mathcal{N}$, resulting in a total of 2^n pure collective strategies.
- A *payoff* function $U_i : \mathcal{S} \mapsto \mathbb{R}^n$ valuing the preference of player i for each $s \in \mathcal{S}$;
  - the value of $U_i(s)$ is assumed to be measured on a cardinal utility scale;
  - $\exists s \in \mathcal{S}$: $U_i(s) \neq 0$, or equivalently, $\max_s |U_i(s)| \neq 0$ for every $i \in \mathcal{N}$.

The central objective of game theory is to understand what will, could, or should happen in a given game. One of the main solution concepts in game theory is the Nash equilibrium, which unified several earlier ideas related to the stability of collective action in non-cooperative games [18]. The ideas we develop in this work are relevant to studying pure-strategy Nash

equilibria (PNE), in which every player is expected to commit fully to one and only one pure individual strategy (see Definition A.1 in the Appendix).

Figs 1(a) and (b) show examples of individual payoff matrices in 2×2 social dilemmas. We can also write all payoffs in a 2×2 game as a simplified bimatrix:

$$[\langle U_i(s_i,s_j)\,,\,U_j(s_j,s_i)\rangle]_{2\times 2} = \begin{bmatrix} \langle U_i(\varphi_i,\varphi_j)\,,\,U_j(\varphi_j,\varphi_i)\rangle & \langle U_i(\varphi_i,\psi_j)\,,\,U_j(\psi_j,\varphi_i)\rangle \\ \langle U_i(\psi_i,\varphi_j)\,,\,U_j(\varphi_j,\psi_i)\rangle & \langle U_i(\psi_i,\psi_j)\,,\,U_j(\psi_j,\psi_i)\rangle \end{bmatrix}, \quad (1)$$

where $\mathcal{S}_k = \{\varphi_k, \psi_k\}$ for every player $k \in \mathcal{N} = \{i, j\}$. The collective strategy set is $\mathcal{S} = \mathcal{S}_i \times \mathcal{S}_j = \{\langle\varphi_i,\varphi_j\rangle, \langle\varphi_i,\psi_j\rangle, \langle\psi_i,\varphi_j\rangle, \langle\psi_i,\psi_j\rangle\}$. Strategies $\varphi = \langle\varphi_i,\varphi_j\rangle$ and $\psi = \langle\psi_i,\psi_j\rangle$ are called *diagonal*; the other collective strategies, namely those in set $\mathcal{S}\backslash\{\varphi,\psi\}$, are called *off-diagonal*. Labeling a combination of individual strategies as diagonal or off-diagonal is left to the reader's discretion, and it does not affect the nature of the game. In this work, φ is also referred to as the *status quo* strategy — or starting point — and ψ is treated as the alternative or *desired* outcome (from the perspective of the incentive designer or researcher) consistent with economics literature [19].

2.2 Structural fear and greed in 2×2 social dilemmas

Classical 2×2 social dilemmas name each strategy after the description that human players would use to describe their behavior [15] and are often symmetric — meaning their payoff functions do not depend on the players' identities, i.e. $U_i(s'_i, s_j) = U_j(s'_j, s_i)$. For instance, let us assume that φ_k is a "defective" individual strategy and that ψ_k is a "cooperative" one. Coordinating on φ and ψ can thus be called mutual (or unanimous) defection and cooperation, respectively. These are the usual strategy labels in a classical social dilemma. Fig 1(a) shows a generic payoff matrix used to characterize these games. Cooperation yields a "reward", $U_i(\psi_i,\psi_j) = R$, which would be preferable to mutual defection, which results in a "punishment", $U_i(\varphi_i,\varphi_j) = P < R$ [20]. The values $U_i(\psi_i,\varphi_j) = S$ and $U_i(\varphi_i,\psi_j) = T$ are called "sucker's" and "temptation" payoffs, respectively, terms borrowed from the *prisoner's dilemma* (PD) game [21]. A sample individual payoff matrix of a PD game is shown in Fig 1(b). A PD game is characterized by the relationships $T > R > P > S$. The *stag hunt* game, in contrast, of which a sample individual payoff matrix is presented in Fig 1(c), satisfies $R > T > P > S$ [22] — a "periodic table" of 2×2 games can be derived from different orderings of P, R, S, and T as described in Ref. [23].

The S payoff is associated with a player's "fear" of pursuing cooperation when others might defect. Meanwhile, the T payoff is a player's incentive to defect away from mutual cooperation driven by the "greed" for higher returns. Ahn et al. [24] measure structural fear and greed, respectively, as

$$F_i(\varphi,\psi) = \frac{P-S}{\max\{P,R,S,T\} - \min\{P,R,S,T\}} \in [-1,+1] \quad (2)$$

and

$$G_i(\varphi,\psi) = \frac{T-R}{\max\{P,R,S,T\} - \min\{P,R,S,T\}} \in [-1,+1]\,. \quad (3)$$

The expressions "$P-S$" and "$T-R$" in Eqs. (2) and (3) are the deviation losses measured with respect to the status quo strategy φ_i. In the context of the 2×2 PD game, the structural fear ($F_i > 0$) is the deficit of unilaterally cooperating when the other defects and the structural greed ($G_i > 0$) is the benefit of taking advantage of a cooperating partner by choosing to defect instead [25]. Either factor favors defection if positive. The higher the values of F_i and G_i, the greater the incentive that player i has to defect instead of playing a cooperative strategy.

Previous works have identified four 2×2 strategy dynamical domains exhibiting different payoff and risk dominance conditions based on the values of S and T, while setting $P = 0$ and

(a) generic social dilemma

$U_i(s_i, s_j)$		s_j	
		φ_j	ψ_j
s_i	φ_i	P	T
	ψ_i	S	R

φ: defect; ψ: cooperate.

(b) A prisoner's dilemma

$U_i(s_i, s_j)$		s_j	
		φ_j	ψ_j
s_i	φ_i	0	2
	ψ_i	-1	1

φ: confess; ψ: stay silent.

(c) A stag hunt

$U_i(s_i, s_j)$		s_j	
		φ_j	ψ_j
s_i	φ_i	0	$1/2$
	ψ_i	$-1/2$	1

φ: chase hares; ψ: hunt the stag.

Fig 1. Generic individual payoff matrix in a two-player two-strategy social dilemma game in normal form. The strategy labels φ and ψ represent "defection" and "cooperation", respectively. The value of P represents the punishment resulting from mutual defection; $R > P$ is the reward of mutual cooperation; S stands for sucker's payoff (what a player gets from cooperating if the other defects); and T stands for temptation to defect (the incentive to deviate away from the cooperative strategy in the expectation of higher returns).

$R = 1$ [26–29]. These values map to one combination of positive (high) or negative (low) structural fear and greed, as provided below and in Fig 1, per Eqs. (2) and (3):

Harmony / `HA` (or cooperation):

$$S > 0 \quad \wedge \quad T < 1 \qquad \Leftrightarrow \qquad F_i < 0 \quad \wedge \quad G_i < 0$$

Coexistence / `CX` (or anti-coordination):

$$S > 0 \quad \wedge \quad T > 1 \qquad \Leftrightarrow \qquad F_i < 0 \quad \wedge \quad G_i > 0$$

Bistability / `BI` (or coordination):

$$S < 0 \quad \wedge \quad T < 1 \qquad \Leftrightarrow \qquad F_i > 0 \quad \wedge \quad G_i < 0$$

Defection / `DE`:

$$S < 0 \quad \wedge \quad T > 1 \qquad \Leftrightarrow \qquad F_i > 0 \quad \wedge \quad G_i > 0$$

We introduce a fifth dynamical domain to represent player i's indifference about the strategy set:

Indifference / `ZZ`:

$$S = 0 \quad \wedge \quad T = 1 \qquad \Leftrightarrow \qquad F_i = 0 \quad \wedge \quad G_i = 0$$

The 2×2 harmony and defection dynamical domains are also known as (type ψ or type φ) *dominance* dynamics [30, 31]; while the coexistence and bistability domains are sometimes referred to as *negative* and *positive frequency dependence dynamics*, respectively [32, 33]. The latter are also known as *bipolar* [34] — note that any 2×2 coexistence game can be transformed into a bistability/bipolar game by switching the strategy labels of one of the two players.

Fig 2 shows how the values of S and T map to the factors F_i and G_i per Eqs. (2) and (3), assuming $P = 0$ and $R = 1$. The vast majority of the literature on social dilemma games focuses on the payoff region defined by $S \in [-1, 1]$ and $T \in [0, 2]$. These games cover the entire harmony domain, $3/8$ of the bistability and coexistence domains, and only $1/6$ of the defection domain — that is, $23/48$ of the whole fear and greed strategy dynamics space. The structural fear and greed values for the PD game in Fig 1(a) are $F_i = G_i = 1/3$, which fall on defection dynamics, while the values for the SH game in Fig 1(b) are $F_i = 1/3$ and $G_i = -1/3$, falling on bistability dynamics. Examples of harmony and coexistence games respectively include the *concord* game ($S = 1/3$ and $T = 2/3 \Rightarrow F_i = G_i = -1/3$) and the *chicken* game ($S = 1/2$ and $T = 3/2 \Rightarrow F_i = -1/3$ and $G_i = 1/3$).

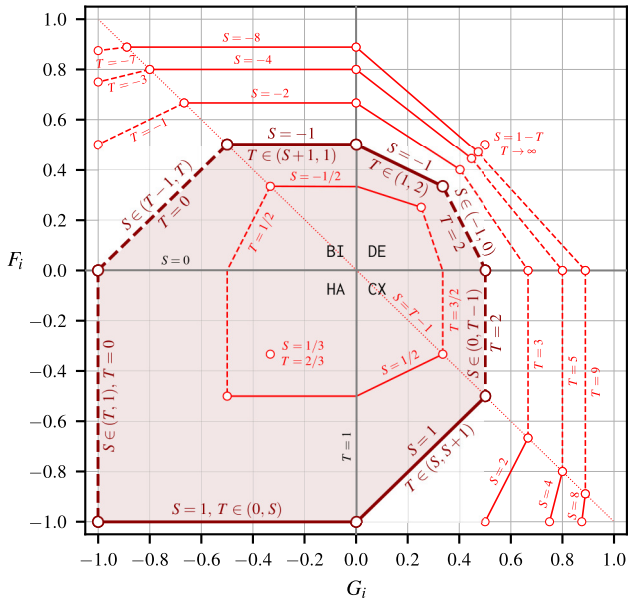


Fig 2. Mapping between the sucker's (S) and the temptation (T) payoffs and the structural fear (F_i) and structural greed (G_i) factors per Eqs. (2) and (3), assuming $P = 0$ and $R = 1$. Each quadrant is labeled after the 2×2 strategy dynamics resulting from different combinations of low and high levels of structural fear and greed: HA / harmony ($F_i < 0,\ G_i < 0$); CX / coexistence ($F_i < 0,\ G_i > 0$); BI / bistability ($F_i > 0,\ G_i < 0$); and DE / defection ($F_i < 0,\ G_i > 0$). The vast majority of the literature on social dilemma games focuses on the shaded region defined by $S \in [-1, 1]$ and $T \in [0, 2]$.

2.3 Deviation losses and strategic stability

For general $n\times2$ games, let us define the *rescaled deviation loss* ℓ_i in terms of the difference in player i's payoff when deviating from s_i^* to $s_i \in \mathcal{S}_i$ assuming the strategies of the other players remain fixed:

Definition 1 *Let $i \in \mathcal{N}$ be a* standpoint *player in the n-player normal-form game*

$$\mathcal{G}_{\mathcal{N}} = \Big(\mathcal{N}, (\mathcal{S}_i)_{i\in\mathcal{N}}, (U_i)_{i\in\mathcal{N}}\Big),$$

where $\mathcal{N}$ is the set of n players, $\mathcal{S}_i$ is the set of individual strategies, $\mathcal{S} = \mathcal{S}_i^n$ are the collective strategies, and $U_i : \mathcal{S} \mapsto \mathbb{R}$ is player i's payoff function. The rescaled individual payoff loss of deviating from s_i^* to $s_i \in \mathcal{S}_i$ contingent on a collective strategy s_{-i} played by all other players ($-i = \mathcal{N}\setminus\{i\}$) is

$$\ell_i(s_i^*, s_{-i}) = \frac{1}{A_i}\Big[U_i(s_i^*, s_{-i}) - U_i(s_i, s_{-i})\Big] \in [-1, +1], \tag{4}$$

where $A_i = \max_s U_i(s) - \min_s U_i(s) > 0$ is the peak-to-peak amplitude of i's payoffs.

If $\mathcal{G}_{\mathcal{N}}$ is an $n\times2$ game and $\mathcal{S}_i = \{\varphi_i, \psi_i\}$, we have that $\ell_i(\varphi_i, s_{-i}) = -\ell_i(\psi_i, s_{-i})$; for the sake of simplicity, we write $\ell_i^*(s_{-i}) = \ell_i(\varphi_i, s_{-i})$ moving forward. Thus, the maximum number of unique values of $\ell_i^*(s_j)$ per player in a binary normal-form game is 2^{n-1}. For $n = 2$, we write

$$\ell_i^*(s_j) = \frac{1}{A_i}\Big[U_i(\varphi_i, s_j) - U_i(\psi_i, s_j)\Big] \in [-1, +1], \tag{5}$$

We now rewrite Eqs. (2) and (3) in terms of Eqs. (4) and (5) as

$$F_i(\varphi,\psi) = \frac{1}{A_i}\Big[U_i(\varphi_i,\varphi_j) - U_i(\psi_i,\varphi_j)\Big] = -\ell_i(\psi_i,\varphi_j) = \ell_i^*(\varphi_j), \tag{6}$$

and

$$G_i(\varphi,\psi) = \frac{1}{A_i}\Big[U_i(\varphi_i,\psi_j) - U_i(\psi_i,\psi_j)\Big] = -\ell_i(\psi_i,\psi_j) = \ell_i^*(\psi_j). \tag{7}$$

If the labels φ and ψ are swapped above, $F_i(\psi,\varphi) = -G_i(\varphi,\psi)$ and $G_i(\psi,\varphi) = -F_i(\varphi,\psi)$.

In binary games, equilibrium conditions depend only on the signa of each player's loss of deviating away from a collective strategy (see Definitions A.2 and A.2a in the Appendix). More precisely, $s^\star = \langle s_i^\star, s_{-i}^\star\rangle \in \mathcal{S}$ is a PNE if

$$\min_{i\in\mathcal{N}} \ell_i(s_i^\star, s_{-i}^\star) \geq 0.$$

In a 2×2 game, if both players' structural fear or greed values fall in the same region (as is the case for symmetric games), then

$$\operatorname{sgn}\ell_i(s_i,s_j) = \operatorname{sgn}\ell_j(s_j,s_i), \tag{8}$$

for some $s_i \in \mathcal{S}_i = \{\varphi_i,\psi_i\}$ and some $s_j \in \mathcal{S}_j = \{\varphi_j,\psi_j\}$, where $\operatorname{sgn}:\mathbb{R}\mapsto\{-1,\,0,\,+1\}$ is the signum function. If both sides of Eq. (8) are equal to $+1$, then $s = \langle s_i, s_j\rangle$ must be a strict PNE. We denote the set of PNE as $\mathcal{S}^\star$. We use the bimatrix notation in Eq. (1) to compare cardinal payoffs with the signa of the rescaled deviation losses and evaluate $\mathcal{S}^\star$ in a 2×2 game:

$$\begin{aligned}[\langle U_i, U_j\rangle]_{2\times 2} &\sim \begin{bmatrix} \langle\ \ \operatorname{sgn}\ell_i^*(\varphi_j),\ \operatorname{sgn}\ell_j^*(\varphi_i)\ \rangle & \langle\ \ \operatorname{sgn}\ell_i^*(\psi_j), -\operatorname{sgn}\ell_j^*(\varphi_i)\ \rangle \\ \langle -\operatorname{sgn}\ell_i^*(\varphi_j),\ \operatorname{sgn}\ell_j^*(\psi_i)\ \rangle & \langle -\operatorname{sgn}\ell_i^*(\psi_j), -\operatorname{sgn}\ell_j^*(\psi_i)\ \rangle \end{bmatrix} \\ &\sim \begin{bmatrix} \langle\ \ \operatorname{sgn}F_i,\ \operatorname{sgn}F_j\ \rangle & \langle\ \ \operatorname{sgn}G_i, -\operatorname{sgn}F_j\ \rangle \\ \langle -\operatorname{sgn}F_i,\ \operatorname{sgn}G_j\ \rangle & \langle -\operatorname{sgn}G_i, -\operatorname{sgn}G_j\ \rangle \end{bmatrix}.\end{aligned} \tag{9}$$

Mapping a 2×2 game to the bimatrix form in Eq. (9) allows us to identify any PNE in it more quickly and how they relate to the structural fear and greed factors. Fig 3 reintroduces the main 2×2 strategy dynamical domains using the aforementioned notation. As an example, the payoffs in a symmetric game with harmony dynamics can be written as

$$[\langle U_i, U_j\rangle]_{2\times 2} = \begin{bmatrix} \langle 1,1\rangle & \langle 3,2\rangle \\ \langle 2,3\rangle & \underline{\langle 4,4\rangle} \end{bmatrix} \sim \begin{bmatrix} \langle -,-\rangle & \langle -,+\rangle \\ \langle +,-\rangle & \underline{\langle +,+\rangle} \end{bmatrix};$$

we underline the cells where each of the two signa are greater than or equal to zero. Thus in the game above, there is one PNE, ψ, since $F_i, F_j < 0$ and $G_i, G_j < 0$, and

$$\operatorname{sgn}\ell_i(\varphi_i,\cdot) = -1 = \operatorname{sgn}\ell_j(\varphi_j,\cdot),$$

implies that both players would always prefer ψ_i (second row) and ψ_j (second column) over φ_i and φ_j. Similarly, in a symmetric game with coexistence dynamics, for example

$$[\langle U_i, U_j\rangle]_{2\times 2} = \begin{bmatrix} \langle 1,1\rangle & \underline{\langle 4,2\rangle} \\ \underline{\langle 2,4\rangle} & \langle 3,3\rangle \end{bmatrix} \sim \begin{bmatrix} \langle -,-\rangle & \underline{\langle +,+\rangle} \\ \underline{\langle +,+\rangle} & \langle -,-\rangle \end{bmatrix},$$

there are two PNE, $\mathcal{S}^\star = \{\langle\varphi_i,\psi_j\rangle, \langle\psi_i,\varphi_j\rangle\}$, the off-diagonal strategies, since $F_i < 0$ and $G_j > 0$ — likewise, $F_j < 0$ and $G_i > 0$ — imply

$$\begin{aligned} -\operatorname{sgn}F_i &= +1 = \operatorname{sgn}G_j \\ -\operatorname{sgn}\ell_i^*(\varphi_j) &= +1 = \operatorname{sgn}\ell_j^*(\psi_i) \\ \operatorname{sgn}\ell_i(\psi_i,\varphi_j) &= +1 = \operatorname{sgn}\ell_j(\varphi_j,\psi_i), \end{aligned}$$

$F_i = \ell_i^*(\varphi_j)$	$G_i = \ell_i^*(\psi_j)$

U_i		s_j	
		φ_j	ψ_j
s_i	φ_i	$\ell_i^*(\varphi_j)$	$\ell_i^*(\psi_j)$
	ψ_i	$-\ell_i^*(\varphi_j)$	$-\ell_i^*(\psi_j)$

⇨

$F_i < 0$	$G_i < 0$

sgn U_i		s_j	
		φ_j	ψ_j
s_i	φ_i	−	−
	ψ_i	+	+

e.g. concord game, peace game

(a) Harmony / HA

$F_i < 0$	$G_i > 0$

sgn U_i		s_j	
		φ_j	ψ_j
s_i	φ_i	−	+
	ψ_i	+	−

e.g. chicken, Bach or Stravinsky

(b) Coexistence / CX

$F_i > 0$	$G_i < 0$

sgn U_i		s_j	
		φ_j	ψ_j
s_i	φ_i	+	−
	ψ_i	−	+

e.g. stag hunt, coordination games

(c) Bistability / BI

$F_i > 0$	$G_i > 0$

sgn U_i		s_j	
		φ_j	ψ_j
s_i	φ_i	+	+
	ψ_i	−	−

e.g. prisoner's dilemma

(d) Defection / DE

Fig 3. Classification of strategy dynamics in 2×2 social dilemma games in terms of the individual payoffs $U_i(s_i, s_j)$ and their relative differences (+ versus −). An economically rational player would presumably prefer those collective strategies marked with a plus sign (+) over those marked with a minus sign (−). In the case of indifference dynamics, notice that both s_i rows are equal, implying that player i's payoff only depends on player j's selection.

indicating that not taking the same action would be the most economically rational outcome.

A game with one or more PNE can also be formed by two individual payoff matrices exhibiting different strategy dynamics. We provide two examples. One: we can model U_i after the harmony payoffs in Fig 3(a) and U_j after the (transposed) defection payoffs in Fig 3(d); using the bimatrix game notation in Eq. (9), we can write

$$[\langle U_i, U_j \rangle]_{2\times 2} = \begin{bmatrix} \langle\ 0,\ \ 0\rangle & \langle\ 0, -1\rangle \\ \underline{\langle\ 1,\ \ 0\rangle} & \langle\ 1, -1\rangle \end{bmatrix} \sim \begin{bmatrix} \langle -, + \rangle & \langle -, - \rangle \\ \underline{\langle +, + \rangle} & \langle +, - \rangle \end{bmatrix}, \tag{10}$$

where the collective strategy $\langle \psi_i, \varphi_j \rangle$ is the only equilibrium in this game. The individual strategies in Eq. (10) are also *perfectly limited*, meaning they yield an individual payoff that depends only on each player's strategy [35]. Using Eqs. (6) and (7), we calculate the structural fear and greed for players i and j as $F_i = G_i = -1 < 0$ (both factors support ψ_i) and $F_j = G_j = +1 > 0$ (both factors support φ_j), respectively. Two: let us reuse the coexistence payoffs to model U_i and (transposed) bistability payoffs per Fig 3(c) to model U_j and write

$$[\langle U_i, U_j \rangle]_{2\times 2} = \begin{bmatrix} \langle\ 0,\ \ 0\rangle & \langle\ 0, -1\rangle \\ \langle\ 1,\ \ 0\rangle & \langle -1,\ \ 1\rangle \end{bmatrix} \sim \begin{bmatrix} \langle -, + \rangle & \langle +, - \rangle \\ \langle +, - \rangle & \langle -, + \rangle \end{bmatrix}; \tag{11}$$

this time, however, there are no PNE. Also, each player only has one perfectly limited strategy: φ_i and φ_j always yield a zero payoff of zero to players i and j regardless of the other's choice of individual strategy. The values of structural fear and greed for players i and j are $\langle F_i, G_i \rangle = \langle -1/2, +1/2 \rangle$ and $\langle F_j, G_j \rangle = \langle +1/2, -1/2 \rangle$, respectively. If only the signa of each player's structural fear and greed factors are considered, the bimatrix game in Eq. (11) is qualitatively equivalent to the 2×2 zero-sum game

$$[\langle U_i, U_j \rangle]_{2\times 2} = \begin{bmatrix} \langle -1,\ \ 1\rangle & \langle\ 1, -1\rangle \\ \langle\ 1, -1\rangle & \langle -1,\ \ 1\rangle \end{bmatrix}. \tag{12}$$

In the following section, we extend the definitions of ℓ_i^*, F_i, and G_i to normal-form games with any number of players $n \geq 2$ and binary strategy sets $\mathcal{S}_i = \{\varphi_i, \psi_i\}$ for every $i \in \mathcal{N}$.

3 Strategy dynamics in n×2 games

In this section, we reformulate the rescaled deviation losses $\ell_i^*(s_{-i}) = \ell_i(\varphi_i, s_{-i})$ for a player $i \in \mathcal{N}$ in an n×2 game in terms of the possible pure individual strategies played by subsets of

$\mathcal{N}\backslash\{i\}$. We focus on the scenarios in which standpoint player i breaks their game into smaller 2×2 games by making two types of conjectures about the possible individual actions of others: 1) some players will coordinate their strategies and act as one, and 2) some other players have already made their choice of individual strategy and will not change their minds. Later, we apply the proposed framework to describe individual and collective strategy dynamics in 3×2 games and characterize social dilemmas with three players.

3.1 Dissecting fear and greed in $n\times2$ games: strategic hindrance

We start by defining *player-reduced binary normal-form games*:

Definition 2 *Let $\mathcal{G}_\mathcal{N} = (\mathcal{N}, (\mathcal{S}_i)_{i\in\mathcal{N}}, (U_i)_{i\in\mathcal{N}})$ be a normal-form game; let $i \in \mathcal{N}$ be a standpoint player in game $\mathcal{G}_\mathcal{N}$; let $\mathcal{K} \subset \mathcal{N}\setminus\{i\}$ be a proper subset of players in $\mathcal{N}$ excluding i; and let $\mathcal{J} = \mathcal{N}\backslash(\mathcal{K}\cup\{i\})$ be the non-empty set of players different from i that are not in $\mathcal{K}$. Assume, without loss of generality, that player i anticipates a collective action $s_\mathcal{K}$ by the players in $\mathcal{K}$. From player i's standpoint, the normal-form game $\mathcal{G}_\mathcal{N}$ reduces to*

$$\mathcal{G}_{\mathcal{N}\backslash\mathcal{K}} = \Big(\mathcal{N}\backslash\mathcal{K}, (\mathcal{S}_i)_{i\in\mathcal{N}\backslash\mathcal{K}}, (U_i(\cdot, s_\mathcal{K}))_{i\in\mathcal{N}\backslash\mathcal{K}}\Big). \tag{13}$$

That is, $\mathcal{G}_{\mathcal{N}\backslash\mathcal{K}}$ is a player-reduced game observed by player i within $\mathcal{G}_\mathcal{N}$.

We use Definition 2 to reformulate Eq. (4) in terms of $s_i \in \{\varphi_i, \psi_i\}$ and $s_{-i} = \langle s_\mathcal{J}, s_\mathcal{K}\rangle$ as

$$\ell_i^*(s_\mathcal{J}, s_\mathcal{K}) = A_i^{-1}\cdot\Big[U_i(\varphi_i, s_\mathcal{J}, s_\mathcal{K}) - U_i(\psi_i, s_\mathcal{J}, s_\mathcal{K})\Big] \in [-1, +1]; \tag{14}$$

where $s_\mathcal{J} \in \mathcal{S}_i^{|\mathcal{J}|}$, $|\mathcal{J}| \in [1\,..\,n-1]$; and $s_\mathcal{K} \in \mathcal{S}_i^{|\mathcal{K}|}$, $|\mathcal{K}| \in [0\,..\,n-2]$.

Next, we characterize player i's fear and greed in game $\mathcal{G}_{\mathcal{N}\backslash\mathcal{K}}$ assuming a coordinated action $s_\mathcal{J} \in \{\varphi_\mathcal{J}, \psi_\mathcal{J}\}$ by player-set $\mathcal{J}$ and an unmovable action $s_\mathcal{K} = \sigma_\mathcal{K}$ by player-set $\mathcal{J}$:

Definition 3 *Let $\mathcal{G}_{\mathcal{N}\backslash\mathcal{K}}$ be a player-reduced game observed by player i, where $s_\mathcal{K} = \sigma_\mathcal{K}$ is the collective strategy of the (excluded) players in $\mathcal{K}$. Assume that, in the face of uncertainty, player i conjectures that all players in $\mathcal{J}$ act as one player with individual strategy $s_\mathcal{J} \in \{\varphi_\mathcal{J}, \psi_\mathcal{J}\}$; and that the players in $\mathcal{K}$ have adopted a presumably immovable collective strategy $s_\mathcal{K} = \sigma_\mathcal{K}$ of which player i is "certain." The values of structural fear and greed specific to player i concerning a unanimous course of action by the players in $\mathcal{J}$ are*

$$F_i^{(\sigma_\mathcal{K})}(\varphi, \psi) = \ell_i^*(\varphi_\mathcal{J}, \sigma_\mathcal{K}) = A_i^{-1}\cdot\Big[U_i(\varphi_i, \varphi_\mathcal{J}, \sigma_\mathcal{K}) - U_i(\psi_i, \varphi_\mathcal{J}, \sigma_\mathcal{K})\Big] \tag{15}$$

and

$$G_i^{(\sigma_\mathcal{K})}(\varphi, \psi) = \ell_i^*(\psi_\mathcal{J}, \sigma_\mathcal{K}) = A_i^{-1}\cdot\Big[U_i(\varphi_i, \psi_\mathcal{J}, \sigma_\mathcal{K}) - U_i(\psi_i, \psi_\mathcal{J}, \sigma_\mathcal{K})\Big], \tag{16}$$

where $A_i = \max_s U_i(s) - \min_s U_i(s) > 0$ and $s_\mathcal{K} \in \mathcal{S}_i^{|\mathcal{K}|}$.

A few of notes on Definition 3:

- Setting $\mathcal{N} = \{i, j\}$ — which implies $\mathcal{J} = \{j\}$ and $\mathcal{K} = \emptyset$ — turns Eqs. (15) and (16) into Eqs. (6) and (7), respectively.
- We intend the assumption of a fixed $s_\mathcal{K} = \sigma_\mathcal{K}$ to represent the scenarios in which each player $k \in \mathcal{K}$ are committed to playing a particular individual strategy $\sigma_k \in \mathcal{S}_k$ and player i both know of their intentions and expects them to be fully realized; we call this a *type I player-reduced binary game: stationary $\mathcal{K}$-players*.

- A second type of player-reduced binary game concerns the scenario in which one or more players in $\mathcal{K}$ are not expected to commit to $\sigma_{\mathcal{K}}$; we call this a *type II player-reduced binary game: reversing $\mathcal{K}$-players* (see Section A.2 in the Appendix for more details).

We now define the *strategic hindrance* of standpoint player i as the space of structural fear and greed values for all possible type I player-reduced games where no more than $n-2$ players are fixed.

Definition 4 *Given a binary n-player normal-form game $\mathcal{G}_{\mathcal{N}}$ with individual strategy set $\mathcal{S}_i = \{\varphi_i, \psi_i\}$, we define the space of structural fear and greed values, or strategic hindrance, of a player $i \in \mathcal{N}$ as*

$$\mathbf{H}_i^{(n)}(\varphi,\psi) = \begin{bmatrix} F_i^{(\sigma_{\mathcal{K}})}(\varphi,\psi) \\ G_i^{(\sigma_{\mathcal{K}})}(\varphi,\psi) \end{bmatrix}^T_{\times g(n)} = \begin{bmatrix} \ell_i^*(\varphi_{\mathcal{J}},\sigma_{\mathcal{K}}) \\ \ell_i^*(\psi_{\mathcal{J}},\sigma_{\mathcal{K}}) \end{bmatrix}^T_{\times g(n)}, \quad \forall \mathcal{J}\cup\mathcal{K} = \mathcal{N}\setminus\{i\},\ \mathcal{J}\neq\emptyset. \tag{17}$$

where $g(n)$ is the number of possible player-reduced binary games $\mathcal{G}_{\mathcal{N}\setminus\mathcal{K}}$ with stationary $\mathcal{K}$-players (type I).

The total number of possible type I player-reduced binary games with fixed $s_{\mathcal{K}} = \sigma_{\mathcal{K}}$ is (see Section A.2 in the Appendix):

$$g(n) = 3^{n-1} - 2^{n-1}, \qquad \forall n \geq 1. \tag{18}$$

There are only 2^{n-1} possible $\ell_i^*(s_{-i})$ that can be input as one of the $2\cdot g(n)$ entries of the strategic hindrance space in Eq. (4). For instance, if $n=3$, $g(n)=5$, and each of the $2\times 5 = 10$ entries in player i's strategic hindrance space takes one of the $2^{3-1}=4$ values of $\ell_i^*(s_{-i})$. This implies that some deviation losses appear more than once and thus impact the emergent strategy dynamics more than others. We quantify such an impact using the function

$$\gamma(n_{\varphi\setminus i}, n_{\psi\setminus i}) = 2^{n_{\varphi\setminus i}} + 2^{n_{\psi\setminus i}} - 2, \tag{19}$$

which counts how many entries of each player's strategic hindrance space are equal to $\ell_i^*(s_{-i})$ based solely on the number of players in $\mathcal{N}\setminus\{i\}$ that play either φ_j and ψ_j, $n_{\varphi\setminus i}$ and $n_{\psi\setminus i}$, satisfying $n_{\varphi\setminus i} + n_{\psi\setminus i} = n-1$. The value of γ in Eq. (19) is associated to the number of type I player-reduced binary games in Eq. (18) via the following alternative formula for $g(n)$ (see Section A.3 in the Appendix):

$$g(n) = \frac{1}{2}\left[\sum_{k=0}^{n-1} \gamma(k, n-1-k)\cdot C(n-1,k)\right]; \tag{20}$$

where $C(n-1,k) = \binom{n-1}{k}$ is the binomial coefficient and k counts the entries in s_{-1} that are equal to either φ_j or ψ_j — note that Eqs. (19) and (20) are symmetric with respect to strategy, that is $\gamma(n_{\varphi\setminus i}, n_{\psi\setminus i}) = \gamma(n_{\psi\setminus i}, n_{\varphi\setminus i})$. For instance, if $n=3$ and $\mathcal{N}=\{1,2,3\}$, the values of γ associated with each possible s_{-1} are

$$\begin{aligned} \text{for } s_{-1} = \langle\varphi_2,\varphi_3\rangle,\ n_{\varphi\setminus i} = 2 \text{ and } n_{\psi\setminus i} = 0 \quad &\Rightarrow \quad \gamma(2,0) = 3; \\ \text{for } s_{-1} = \langle\varphi_2,\psi_3\rangle,\ n_{\varphi\setminus i} = 1 \text{ and } n_{\psi\setminus i} = 1 \quad &\Rightarrow \quad \gamma(1,1) = 2; \\ \text{for } s_{-1} = \langle\psi_2,\varphi_3\rangle,\ n_{\varphi\setminus i} = 1 \text{ and } n_{\psi\setminus i} = 1 \quad &\Rightarrow \quad \gamma(1,1) = 2; \\ \text{and for } s_{-1} = \langle\psi_2,\psi_3\rangle,\ n_{\varphi\setminus i} = 0 \text{ and } n_{\psi\setminus i} = 2 \quad &\Rightarrow \quad \gamma(0,2) = 3; \end{aligned}$$

and, per Eq. (20), $g(3) = (1/2)\cdot[\gamma(0,2)\cdot(1) + \gamma(1,1)\cdot(2) + \gamma(2,0)\cdot(1)] = (1/2)\cdot 10 = 5$.

Table 1. Strategic hindrance space $\mathbf{H}_i^{(3)}(\varphi,\psi)$ for player i in a 3×2 game.

			$\mathbf{H}_i^{(3)}(\varphi,\psi)$				
$\mathcal{J}$	$\mathcal{K}$	$\sigma_{\mathcal{K}}$	$F_i^{(\sigma_{\mathcal{K}})}$		$G_i^{(\sigma_{\mathcal{K}})}$		
$\{j\}$	$\{k\}$	φ_k	$\ell_i^*(\varphi_j,\varphi_k)$				$\ell_i^*(\psi_j,\varphi_k)$
		ψ_k		$\ell_i^*(\varphi_j,\psi_k)$		$\ell_i^*(\psi_j,\psi_k)$	
$\{k\}$	$\{j\}$	φ_j	$\ell_i^*(\varphi_j,\varphi_k)$			$\ell_i^*(\varphi_j,\psi_k)$	
		ψ_j			$\ell_i^*(\psi_j,\varphi_k)$	$\ell_i^*(\psi_j,\psi_k)$	
$\{j,k\}$	$\emptyset$	n/a	$\ell_i^*(\varphi_j,\varphi_k)$			$\ell_i^*(\psi_j,\psi_k)$	

3.2 Strategic hindrance in 3×2 games

This section inspects the information that can be extracted from a strategic hindrance of a player in a 3×2 game when represented in a Euclidean plane. Solving Eq. (17) for $n=3$, player i's strategic hindrance can be expressed as:

$$\mathbf{H}_i^{(3)}(\varphi,\psi)=\frac{1}{A_i}\begin{bmatrix}\ell_i^*(\varphi_j,\varphi_k) & \ell_i^*(\varphi_j,\psi_k) & \ell_i^*(\varphi_j,\varphi_k) & \ell_i^*(\psi_j,\varphi_k) & \ell_i^*(\varphi_j,\varphi_k)\\ \ell_i^*(\psi_j,\varphi_k) & \ell_i^*(\psi_j,\psi_k) & \ell_i^*(\varphi_j,\psi_k) & \ell_i^*(\psi_j,\psi_k) & \ell_i^*(\psi_j,\psi_k)\end{bmatrix}^{\mathrm{T}}. \tag{21}$$

Table 1 lists each of the corresponding $g(3)=5$ structural fear and greed value pairs for $n=3$ in terms of the rescaled deviation losses $\ell_i^*(s_j,s_k)$. We use rectangular coordinates to describe a generic individual strategic hindrance space in Fig 4. In Fig 6, we add strategic hindrance spaces for the remaining two players and discuss how the visualization is useful for understanding the stability of collective actions in the game. We repeat this assessment after modifying the payoffs of one of the players, namely by multiplying by -1, which results in the strategic hindrance spaces in Fig 9.

Generic individual 3×2 strategic hindrance space

A player's strategic hindrance highlights the multiple modes of biases that are baked into the payoff structure. In a one-shot game, a player may rely on simple arithmetic to play the pure strategy that maximizes their expected value. This may be the case in games with more regular payoff structures that produce fewer unique fear and greed value pairs per player, usually falling entirely within the same strategy dynamical domain — e.g., the social dilemmas analyzed in Section A.4. All structural fear and greed value pairs that form a strategic hindrance space in a 3×2 game can be contained within a convex polygon with at most four sides — we conjecture that the number of sides of the convex polygon containing all fear and greed value pairs in an individual strategic hindrance space is at most $4(n-2)$ for $n>2$.

The payoffs and absolute deviation losses that produce the strategic hindrance space in Fig 4 are provided in Fig 5. Using Eqs. (14) and (15)–(17) we obtain player i's strategic hindrance space:

$$\mathbf{H}_i^{(3)}(\varphi,\psi)=\begin{bmatrix}F_i^{(\sigma_{\mathcal{K}})}\\ G_i^{(\sigma_{\mathcal{K}})}\end{bmatrix}^{\mathrm{T}}_{\times 5}=\frac{1}{5}\begin{bmatrix}+2 & +3 & +2 & -3 & +2\\ -3 & -1 & +3 & -1 & -1\end{bmatrix}^{\mathrm{T}}.$$

The five fear and greed value pairs in the above strategic hindrance space are all unique and fall across three 2×2 strategy dynamical domains — harmony (1), defection (1), and bistability (3). The center of mass is located at the average structural fear and greed coordinates, $\bar{F}_i=6/25$ and

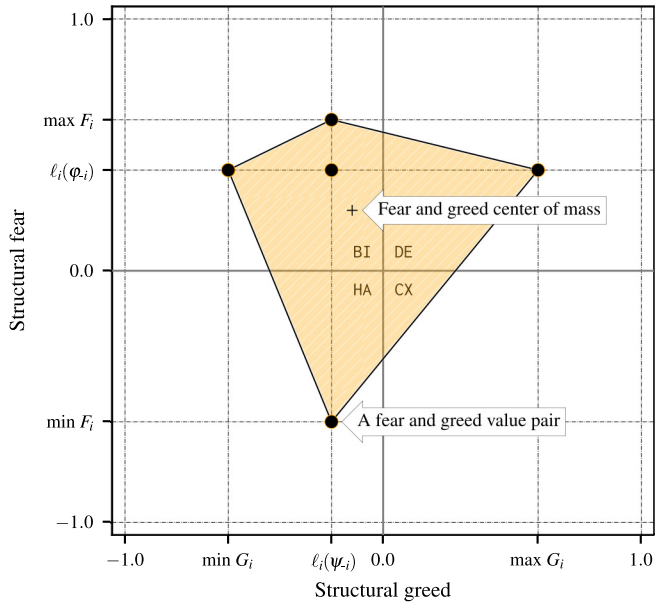


Fig 4. Sample strategic hindrance space $\mathbf{H}_i^{(3)}(\varphi, \psi)$ in a 3×2 game with $g(3) = 5$ unique type I player-reduced games observed by standpoint player i. The payoff structure is shown in Fig 5.

$U_i(s_i, s_j, s_k)$	$s_k = \varphi_k$		$s_k = \psi_k$	
	$s_j = \varphi_j$	$s_j = \psi_j$	$s_j = \varphi_j$	$s_j = \psi_j$
$s_i = \varphi_i$	1	1	3	1
$s_i = \psi_i$	−1	4	0	2
$U_i(\varphi_i, s_j, s_k) - U_i(\psi_i, s_j, s_k)$	2	−3	3	−1

Fig 5. Sample individual payoffs and absolute deviation losses in a 3×2 game. The resulting strategic hindrance space is shown in Fig 4; the peak-to-peak payoff amplitude is $A_i = 5$.

$\bar{G}_i = -3/25$. There are always $\gamma(n-1, 0)$ pairs along the ordinate $\ell_i^*(\varphi_{-i})$ and $\gamma(0, n-1)$ along the abscissa $\ell_i^*(\psi_{-i})$, with their intersection representing the player-reduced binary game where all other players play as one.

In games where players' incentives are sparser, pre-existing beliefs of what others would do could make a player dismiss the likelihood of specific outcomes. The strategic hindrance space represents the spectrum of such beliefs, each of which a standpoint player would weigh according to a perceived chance, guiding their presumptions about the course of the game. If the standpoint player in our hypothetical 3×2 game example were to assign the same weight to all instances in Fig 4, the most prevalent strategy dynamics affecting their decision-making process would be structural bistability. This would involve the anticipation of aligning their strategies with at least one of the two other players. Since the remaining emergent strategy dynamics are harmony and defection, the standpoint player would prefer everyone to pursue a unanimous course of action, fully embracing either ψ or φ.

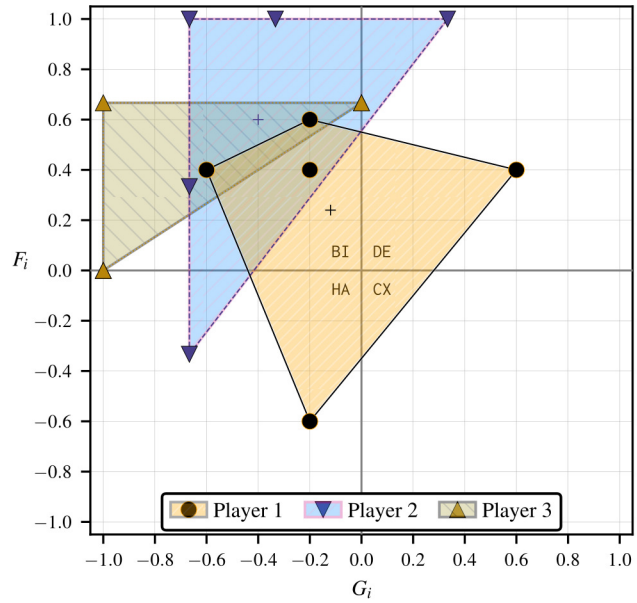


Fig 6. Strategic hindrance spaces in the sample 3×2 game in Fig 7. Player 1's payoff structure is taken from Fig 5, so their strategic hindrance space is the same as in Fig 4. Bistability dynamics prevail across the strategic hindrance spaces of all players, while the emergence of harmony and defection may add further appeal to the diagonal collective strategies ψ and φ.

$\langle U_1, U_2, U_3 \rangle$	$s_3 = \varphi_3$		$s_3 = \psi_3$	
	$s_2 = \varphi_2$	$s_2 = \psi_2$	$s_2 = \varphi_2$	$s_2 = \psi_2$
$s_1 = \varphi_1$	$\langle 1, 3, 2 \rangle$	$\langle 1, 0, 0 \rangle$	$\langle 3, 1, 0 \rangle$	$\langle 1, 0, 0 \rangle$
$s_1 = \psi_1$	$\langle -1, 0, 0 \rangle$	$\langle 4, 1, 0 \rangle$	$\langle 0, 0, 0 \rangle$	$\langle 2, 2, 3 \rangle$

Fig 7. Sample 3×2 game used to generate the strategic hindrance spaces in Fig 6. Player 1's incentives are modeled after the payoff structure in Fig 5. This game has two PNE: φ and ψ.

Inferring collective stability from multiple strategic hindrance spaces

The strategic hindrance space in Fig 4 is a prognostic of a single player's stability of rational strategic action based on their payoff structure. Suppose all other players shared similar payoff structures; in this scenario, we can anticipate the emergence of PNE coherent with the individual rational preferences coinciding in the same strategy dynamical domains. In some circumstances, we can also repeal PNE by inverting some players' strategy dynamics, breaking compatibilities apart. For the grand game to exhibit PNE consistent with player 1's preferences, the payoff structures of the other players must also result in a combination of harmony, bistability, and defection dynamics. Fig 6 highlights such a scenario with similar strategic hindrances, with player 2's strategic hindrance space spanning the same three strategy dynamical domains as player 1's, and all player 3's fear and greed value pairs (weakly) falling under bistability. The baseline 3×2 game is presented in Fig 7. The PNE set in this game, $\mathcal{S}^\star = \{\varphi, \psi\}$, results from the intersection between the rational preferences of all three players under bistability dynamics (which already favor φ and ψ) and the strict stability under harmony and defection dynamics for players 1 and 2.

In contrast with the strategic hindrance spaces in Fig 6, we can create a scenario where there

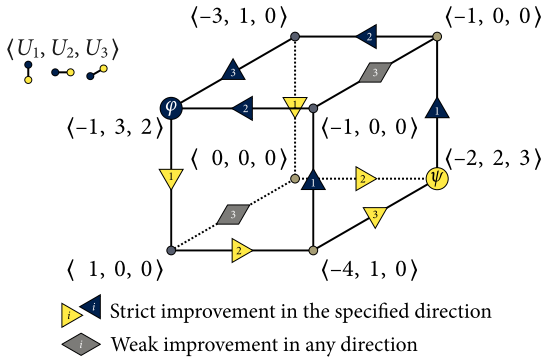


Fig 8. Modified 3×2 game after negating player 1's payoffs in Fig 7, resulting in no PNE. This 3-cube representation highlights the anticipated lack of convergence toward a collective strategy.

Table 2. Deviation losses and strategic hindrance spaces for the 3×2 game in Fig 8. All spaces are visualized in Fig 9. The results for players 2 and 3 are also shared in Fig 6.

	Player 1	Player 2	Player 3
ℓ_i^*	$\ell_1^*(\varphi_2,\varphi_3)=\frac{1}{5}(-1-1)=-\frac{2}{5}$	$\ell_2^*(\varphi_3,\varphi_1)=\frac{1}{3}(3-0)=+1$	$\ell_3^*(\varphi_1,\varphi_2)=\frac{1}{3}(2-0)=+\frac{2}{3}$
	$\ell_1^*(\varphi_2,\psi_3)=\frac{1}{5}(-3-0)=-\frac{3}{5}$	$\ell_2^*(\varphi_3,\psi_1)=\frac{1}{3}(1-0)=+\frac{1}{3}$	$\ell_3^*(\varphi_1,\psi_2)=\frac{1}{3}(0-0)=\ 0$
	$\ell_1^*(\psi_2,\varphi_3)=\frac{1}{5}(-1+4)=+\frac{3}{5}$	$\ell_2^*(\psi_3,\varphi_1)=\frac{1}{3}(0-1)=-\frac{1}{3}$	$\ell_3^*(\psi_1,\varphi_2)=\frac{1}{3}(0-0)=\ 0$
	$\ell_1^*(\psi_2,\psi_3)=\frac{1}{5}(-1+2)=+\frac{1}{5}$	$\ell_2^*(\psi_3,\psi_1)=\frac{1}{3}(0-2)=-\frac{2}{3}$	$\ell_3^*(\psi_1,\psi_2)=\frac{1}{3}(0-3)=-1$
$\mathbf{H}_i^{(3)}$	$\frac{1}{5}\begin{bmatrix}-2 & -3 & -2 & +3 & -2\\ +3 & +1 & -3 & +1 & +1\end{bmatrix}$	$\frac{1}{3}\begin{bmatrix}+3 & +1 & +3 & -1 & +3\\ -1 & -2 & +1 & -2 & -2\end{bmatrix}$	$\frac{1}{3}\begin{bmatrix}0 & -3 & 0 & -3 & -3\\ +2 & 0 & +2 & 0 & +2\end{bmatrix}$

is no set of type I player-reduced game observed by all players that strictly falls under the same 2×2 strategy dynamical domain. We can achieve this by negating player 1's payoffs in Fig 7. The resulting game is shared in Fig 8 using an alternative 3-cub representation to highlight the directions of strict and weak payoff improvement to identify PNE, with the vertical, horizontal, and depth axes listing the individual strategies for players 1, 2, and 3, respectively. Each vertex represents a collective strategy outcome. The direction of each orthogonal edge connecting two vertices is determined by the sign of the individual deviation loss associated with it. A vertex is a PNE if it is a sink [36], meaning that all arcs connected to it are incoming [36]. There are no sinks in Fig 6; so, there are no PNE. The deviation losses and strategic hindrance spaces in this game are provided in Table 2 and visualized in Fig 9.

Negating an individual payoff structure causes observed harmony dynamics to turn into defection; it also causes bistability to turn into coexistence (and vice-versa). Harmony and defection are both dominance dynamics and thus are compatible, as described for 2×2 games in Section 2.3 and Eq. (10) — in Section A.4, we discuss *public goods games*, which have a single PNE and where the aspiration level of each player can be set to result in either pure defection or pure harmony dynamics. The same cannot be said about coexistence and bistability. In the 2×2 games in Eq. (11) and Eq. (12), we observe that combining positive and negative frequency dependence may result in the absence of PNE. Coexistence dynamics do not favor a unanimous course of action. Player 1's coexistence-based preferences in Fig 9 cancel out the intersecting harmony-bistability-defection preferences of players 2 and 3 that favor φ and ψ.

It is still possible to create a 3×2 game where the incidence of coexistence and bistability results in the emergence of PNE. This can be done by mitigating harmony and defection. For instance, in Fig 8, making $U_2(s)=0$ at $s=\langle\varphi_1,\varphi_2,\psi_3\rangle$ and $s=\langle\psi_1,\psi_2,\varphi_3\rangle$ makes the rescaled

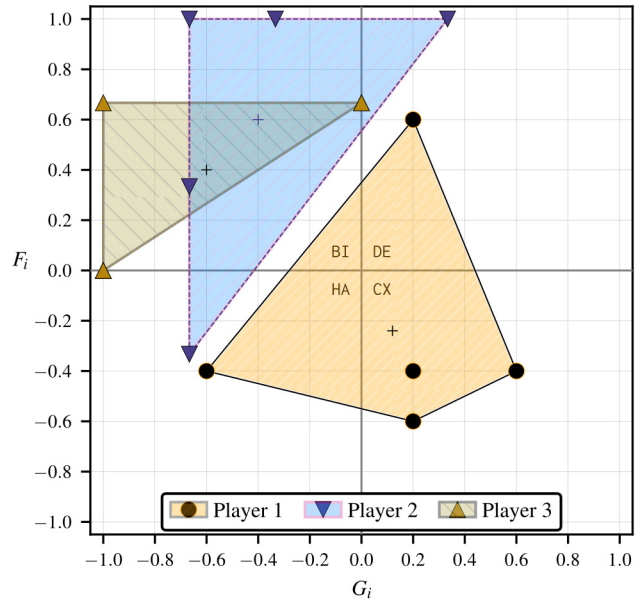


Fig 9. Strategic hindrance spaces in the 3×2 game in Fig 8. All strategic hindrance spaces but player 1's are the same as in Fig 6 — the payoff structures of players 2 and 3 are taken from Fig 7. The coexistence dynamics that emerge for player 1 indicate a greater incentive for refusing to align with the strategies of players 2 and 3.

deviation losses $\ell_2^*(\varphi_3, \psi_1)$ and $\ell_2^*(\psi_3, \varphi_1)$ in Table 2 become zero, turning all but one of the $g(3) = 5$ type I player-reduced games into borderline bistability games with either $F_2 = 0$ or $G_2 = 0$. Player 2's modified strategic hindrance is similar to that of player 3, as shown in Fig 10 — *unanimity* games, in the following section, are characterized by the emergence of analogous borderline bistability dynamics for all players. The lower left and upper right vertices in Fig 8 — respectively $\{\langle \varphi_1, \psi_2, \psi_3 \rangle$ and $\langle \psi_1, \varphi_2, \varphi_3 \rangle\}$ — weakly become sinks. These weak PNE result from the intersection of player 1's coexistence-driven disfavoring of unanimity and players 2 and 3's bistability-driven inclination to coordinate with other players. Here the incidence of player 1's observed harmony and defection dynamics is negligible.

3.3 Application case: integrability of an urban transit system

Here, we analyze the emergence of strategy dynamics in the strategic design for the integrability of a fictional urban transit system, highlighting the complexities that arise as the number of key design actors increases. In this example, we explore the concept of directional integrability [14], briefly discussed in Section 1.1, to assess how well the system-of-systems aligns with shared goals, considering the levels of managerial control and collaboration among its components.

The baseline 2×2 scenario

Let us consider two managerially and operationally independent system design actors:

- Irene, a software developer who was contracted by the public transportation authority to implement a passenger information system; and

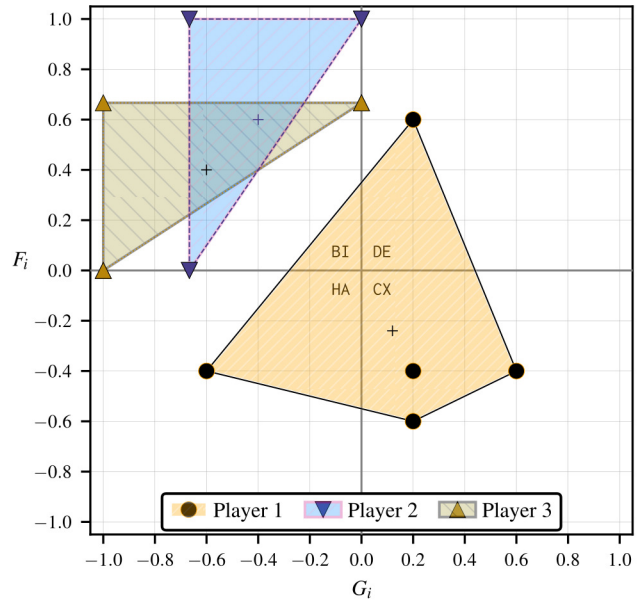


Fig 10. Altered strategy dynamics derived from the 3×2 game in Fig 8 after making $U_2(s) = 0$ at $s = \langle \varphi_1, \varphi_2, \psi_3 \rangle$ and $s = \langle \psi_1, \psi_2, \varphi_3 \rangle$, resulting in $\ell_2^*(\varphi_3, \psi_1) = \ell_2^*(\psi_3, \varphi_1) = 0$ in Table 2. The strategic hindrance spaces of players 2 and 3 are akin to those in a unanimity game (Fig A.5) characterized in Section A.4. The PNE set is $\mathcal{S}^\star = \{\langle \varphi_1, \psi_2, \psi_3 \rangle, \langle \psi_1, \varphi_2, \varphi_3 \rangle\}$.

- Jamie, an infrastructure contractor whom the regional transportation policy agency tasked to design and maintain the traffic management system.

Irene's primary role is to help commuters plan their routes and access real-time information on buses, trams, and subways, while Jamie's contribution to developing and managing road infrastructure and traffic critically impacts the flow of vehicles and the safety of passengers.

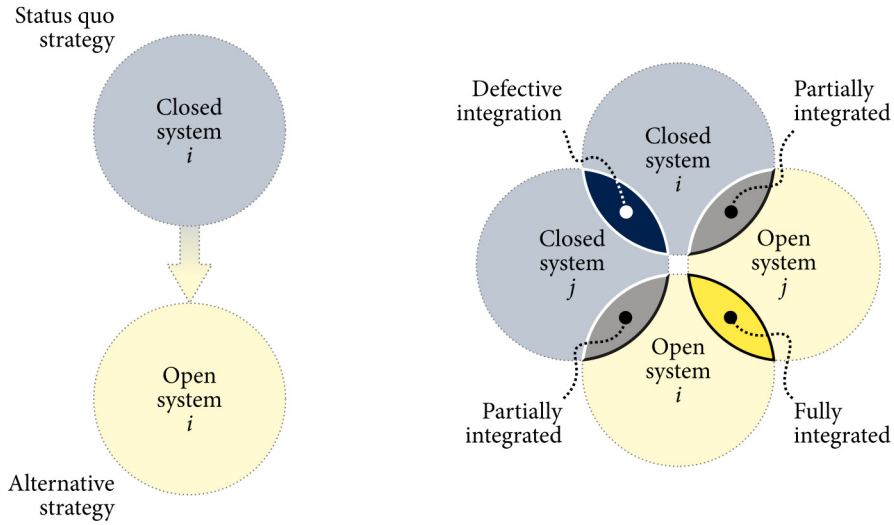
Each actor can independently choose an open or closed system strategy:

- φ_i — closed system to leverage partial integrability: the system will take advantage of the capabilities of other open systems, but the latter cannot exploit the former.
- ψ_i — open system to pursue synergistic integrability: the system's capabilities can be fully leveraged by everyone.

We consider keeping the system closed as the status quo strategy since it would not require an actor to de-constrain and expand the capabilities of their system beyond its boundaries to interact with others. In this context, pursuing synergistic integrability would be an alternative course of action — see Fig 11(a). We justify these assumptions by arguing that strategic decisions have an intended direction and are irreversible [37]. We also disregard potential halfway or "mixed" contingency plans, assuming that any decision to deviate from the status quo can be construed as a technical variation of the same alternative strategy. Thus, we focus on the incentives that affect a system actor's decision to abandon an already in-place or taken-for-granted strategy in favor of a presumably riskier technological alternative whose potential depends on whether other actors adopt it.

The strategic tradeoffs: A volunteer's dilemma

We are interested in assessing the value utility of each system actor based on how their strategic decisions would intersect. The possible strategic outcomes considered are described in Fig 11(b).



(a) Individual path to increased integrability

(b) Possible collective strategy outcomes

Actor i's expected outcomes		Actor j's strategy	
		Closed system	Open system
Actor i's strategy	Closed system	peril	opportunity
	Open system	compromise	compromise

peril $\prec$ compromise $\prec$ opportunity

(c) Actor j's presumed impact on actor i's outcomes

Fig 11. Increasing integrability strategically: (a) Each actor (say, actor i) decides whether to keep their system closed (status quo) or to allocate resources to making their system fully integrable for the benefit of other systems (alternative strategy); (b) The expected outcomes resulting from the interaction between two systems, i and j, as a tradeoff between the value of synergistic integrability versus the risky prospect of taking advantage of the other system's increased openness; and (c) hypothetical assessment of system actor i's expected outcomes from choosing to open or close their system (rows) contingent on actor j's strategy (columns).

We estimate each actor's expected value using the following three rules:

1. If no system actors open their systems ($s = \varphi$), the defective integration will result in disappointing long-term outcomes for everyone (a "peril").
2. No actor implementing an open system ($s_i = \psi_i$) benefits significantly from increased integrability when others do the same (a "compromise").
3. An actor who maintains a closed system ($s_i = \varphi_i$) can potentially take advantage of the partial integrability resulting from others opening their systems (an "opportunity").

Under these circumstances, and without any other actors involved, Irene and Jamie cannot take advantage of increased integrability at the same time. Focusing on Irene's point of view, maximizing value would require them to keep the system closed while Jamie opens theirs. If Jamie refuses, the best alternative for Irene is to adopt an open approach. It is worth noting that the chance of peril may also result from a coordination failure between subsystems due to technical or operational reasons.

We summarize the expectations of any of the two actors in Fig 11(c) and assume they are (ordinally) symmetric, as their preferences depend only on the combination of strategies, not on who makes each decision. Assigning values to each outcome turns this problem into a normal-form game. The compromise of opening the system is more preferred than the peril of a failed integration but less preferred than the opportunity of taking advantage of other open systems; thus, we set "peril" $= -1$, "compromise" $= 0$, and "opportunity" $= +1$. This is equivalent to a *volunteer's dilemma*, where "volunteering" and "free-riding" are the alternative and the status quo strategies, respectively. We can model the payoff function for this game using Eq. (A.6) in Section A.4 in the Appendix:

$$U_i(s_i, s_{\mathcal{J}}, s_{\mathcal{K}}) = \begin{cases} 0 & \text{if } s_i = \psi_i \\ +1 & \text{if } s_i = \varphi_i \wedge s_{\mathcal{J}} = \psi_{\mathcal{J}} \\ -1 & \text{otherwise,} \end{cases} \tag{22}$$

With two actors, there is only one value of structural fear and one value of structural greed, the former is negative and congruent with the difference between the value of "peril" and the value of "compromise", and the latter is positive and proportional to the value of "opportunity" minus the value of "compromise." So, it exhibits pure coexistence dynamics.

Potential impact of increasing the number of actors

Assume the involvement of a third actor named Kumar, who has been tasked to oversee safety and communications in the larger urban transit ecosystem and whose actions are critical to achieving synergistic integrability. Like Irene and Jamie, Kumar is presumed to be operationally and managerially independent in implementing an open or a closed system; however, Irene and Jamie may not be aware of Kumar's exact level of autonomy. For example, Irene may believe that a governmental agency supersedes Kumar's authority in whole or part and that whatever Kumar can do has already been defined. From Irene's standpoint:

(a) Kumar may *never* open their system — a plausible assumption considering the prioritization of security, reliability, data privacy, and regulatory compliance.

(b) Kumar may *surely* open the safety and communications system for scalability, adaptiveness, and transparency.

In the context of the volunteer's dilemma, it can be inferred that if Irene decides to maintain a closed system, it will likely result in favorable outcomes in scenario (b). However, if Kumar chooses not to pursue integrability, the risk of peril is greater in the event of (a). Notice that, in scenarios (a) and (b), it could be Jamie, not Kumar, the one presumed to have fixed their strategy. Also, it is possible that

(c) Jamie and Kumar align their strategies, simultaneously opening or closing their systems,

which, as in scenario (b), would make the closed system strategy riskier. If we assign the indices i, j, and k to the three actors, we can list the following five player-reduced binary games from the standpoint of actor i:

(1) actor j's strategy may be uncertain, but actor k will surely maintain the status quo;

(2) actor j's strategy may be uncertain, but actor k will surely choose the alternative;

(3) actor k's strategy may be uncertain, but actor j will surely maintain the status quo;

(4) actor k's strategy may be uncertain, but actor j will surely choose the alternative;

(5) both j's and k's strategies may be uncertain, but they will surely align their actions.

U_i		$\langle s_j, s_k \rangle$	
		φ_k	
		φ_j	ψ_j
s_i	φ_i	-1	$+1$
	ψ_i	0	0

(a) Coexistence / CX

U_i		$\langle s_j, s_k \rangle$	
		ψ_k	
		φ_j	ψ_j
s_i	φ_i	$+1$	$+1$
	ψ_i	0	0

(b) Defection / DE

U_i		$\langle s_j, s_k \rangle$	
		φ_k	ψ_k
		φ_j	
s_i	φ_i	-1	$+1$
	ψ_i	0	0

(c) Coexistence / CX

U_i		$\langle s_j, s_k \rangle$	
		φ_k	ψ_k
		ψ_j	
s_i	φ_i	$+1$	$+1$
	ψ_i	0	0

(d) Defection / DE

U_i		$\langle s_j, s_k \rangle$	
		φ_k	ψ_k
		φ_j	ψ_j
s_i	φ_i	-1	$+1$
	ψ_i	0	0

(a) Coexistence / CX

Fig 12. Type I player-reduced binary games that would be observed by standpoint player i in a 3×2 volunteer's dilemma per Eq. (22)

Fig 12 lists each of these games. Three of them exhibit coexistence dynamics and the remaining two exhibit defection dynamics. Specifically, using Eq. (A.7) in Section A.4 in the Appendix, we obtain the strategic hindrance

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{2}\begin{bmatrix} -1 & +1 & -1 & +1 & -1 \\ +1 & +1 & +1 & +1 & +1 \end{bmatrix}^{\mathrm{T}}.$$

If we add a fourth actor, the ratio of coexistence to defection becomes 7 to 12. With five actors, it becomes 15 to 50. The incidence of defection dynamics on each player's strategic hindrance grows at a higher rate than the incidence of coexistence dynamics as n grows; more specifically,

$$\frac{\text{Number of CX pairs}}{\text{Number of DE pairs}} = \frac{2^{n-1}-1}{3^{n-1}-2^n+1}.$$

We plot the evolution of coexistence and defection dynamics in the n-player volunteer's dilemma in Fig 13 as the share of the total number of type I player-reduced binary games, $g(n)$, that fall into each domain as the number of actors increases. In the context of our urban transit system example, we can expect the progress of integrability across the ecosystem to become stagnant after one actor (if any) has broken away from the status quo and implemented an open system (one of n possible PNE), assuming the actors are "economically rational." However, if more subsystems were to be integrated, the increasing predominance of defection dynamics would lead to a decreasing likelihood of any system actor volunteering and opening their subsystems. This quantitative insight into the evolution of defection in the volunteer's dilemma game is reminiscent of the prevalence of the diffusion of responsibility and bystander effect phenomena observed in real-world settings [38].

A key takeaway of this example is how crucial the efficient coordination of actions among stakeholders and systems is for optimizing resource allocation and reaching greater collective performance. For instance, if we wanted to incentivize systems-of-systems integrability in this hypothetical setting, one potential approach is to organize interconnected subsystems into small clusters to minimize the risk of coordination failure. In real-world systems-of-systems, we expect actors' strategic tradeoffs and information to be asymmetrical and evolving, requiring the design of mechanisms to align incentives and optimize the allocation of resources. As engineering systems become more complex, managing integration and reducing incompatibility among diverse subsystems becomes increasingly challenging. Tackling these challenges will enable the management of interdependent subsystems in a harmonious way throughout each stage of the design process, ultimately leading to the creation of efficient and effective systems-of-systems that benefit society as a whole.

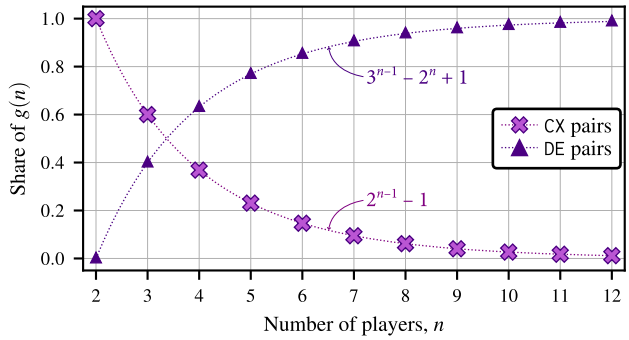


Fig 13. Evolution of the strategy dynamics in n-player volunteer's dilemma as a proportion of the number of $\langle F_i, G_i \rangle$ pairs in an individual strategic hindrance space that fall on a specific domain to the total number of type I player-reduced games, $g(n)$. The incidence of defection dynamics in the volunteer's dilemma becomes more significant as n grows, while the proportion of type I player-reduced games with coexistence dynamics decreases.

4 Discussion of assumptions and limitations

The formulation proposed in this paper takes inspiration from existing subjectivistic and epistemic approaches to game theory [39, 40], focusing less on the players' *knowledge* and more on their *beliefs*. This worldview is present, in particular, in the computation of values of structural fear and greed for every player-reduced normal-form game — grouping one or more actors as one player (set $\mathcal{J}$) and setting the strategies of any remaining player (set $\mathcal{K}$) as "fixed". We can use the latter to represent actors in the context of engineering systems whose actions have presumably been "optimized". For example, in the example of an urban transit system in Section 3.3, a designer who believes that at least one other system actor is bound to invest in integrability might decide to take advantage of the latter. Here, the actor attempts to complete information about the game despite lacking "common" or "objective" knowledge about the likelihood of a specific combination of other players' strategies. While the strategic hindrance space allows a standpoint player to become better informed of such combinations and understand how they are distributed across the four strategy dynamical domains, the weight that the actor assigns to each player-reduced game will still depend on their understanding of the other players and the context in which individual and collective decisions are being made.

Another instance of incompleteness of information that could be modeled as a player-reduced normal-form game is when a player is uncertain about whether the rest of the players are acting as individual agents or as a collective [41]. For example, in the El Farol bar game in Eq. (A.12), there are players who might never go to the bar because they believe everyone else is going (by coincidence or by collusion), so it would already be 2/3 full by the time they arrive — in a dynamic context, this could lead to a "belief-distorted" Nash equilibria, as discussed by Wiszniewska-Matyszkiel [42]. For such player, going to El Farol bar is like playing a chicken game (low fear, high greed); but for those who are not so pessimistic about the bar being crowded, the game is called harmony (low fear, low greed).

Studying strategy selection in n-player binary normal-form games in terms of individual deviations losses is not original of this work. Examples of similar approaches can be found in Güth and Kalfoken's [43] theory of resistance avoidance and Selten's [34] theory of risk dominance in bipolar games. These approaches focus on measuring the likeliness that an individual remains at (or deviates towards) an equilibrium point; both works then prescribe objective ways in which such "resistances" or "diagonal probabilities" could be aggregated. Studying equilibria in player-reduced normal-form games from a subjectivistic point of view is not a new concept either. For instance, in Kalai and Lehrer's [44] model of subjective games,

players in a large game environment would restrict themselves to assessing only the portion of the game they know (viz. an "environment response function") and approaching a subjective Nash equilibrium that might coincide with an objective equilibrium regardless of how optimistic or pessimistic their beliefs are about the "outsiders" to their subjective games. Similarly, Battigalli and Guaitoli [45] drop the "common prior assumption" (the premise of the existence of a collectively known probability distribution across the players' strategies [46]) and favor a model of "steady-state equilibria" where an individual bases their actions on conjectures about others' strategies and is unable (or even apathetic) to fact check.

Despite its subjectivistic nature, the proposed framework shares some similarities with existing objective and normative approaches to characterizing n-player games. In yet another work on equilibria selection, Güth [47] assesses the applicability of requiring consistency and non-emptiness of equilibria when choosing a solution concept. The analysis, which builds on previous work by Peleg and Tijs [48] and Norde et al. [49], questions the robustness of a solution concept (whether Nash equilibrium or not) when all players are considered and when some of the players are "gone" and their inclinations might not be revealed. On one hand, compared to Güth's analysis, the structural fear and greed spaces provide a visual means of understanding how robust a collective strategy is to unilateral deviations, namely, by checking if those fear and greed spaces fall inside regions of single PNE like the harmony and defection strategy dynamical domains. On the other hand, the proposed framework also provides insights into how consistent a collective strategy outcome remains across all possible player-reduced games by estimating how much the players' structural fear and greed spaces overlap. Both framework qualities do not constitute a solution concept themselves but could prove helpful when assessing incentive mechanisms to guide strategic behaviors toward desired outcomes.

A significant limitation of our framework is its focus on one-shot normal-form games where players are limited to only two pure individual strategies, and these actions are assumed to be irreversible once taken. It does not account for the possibility of mixed strategies, where players can adopt a probabilistic combination of pure strategies, which would add a layer of versatility to the model. Moreover, the assumption of irreversible actions may only sometimes be accurate in real-world engineering systems. Going back to the urban transit system example, the strategic decisions involved, while initially appearing irreversible, may be subject to change as the systems development process evolves. Feedback from the public, changing urban dynamics, unforeseen route modifications, and integration of new transportation modes, could lead to adjustments to the original "game". Despite these limitations, our framework can serve as a supplementary early approach to evaluate the long-term implications of strategic decisions in engineering systems. It also allows researchers and practitioners to revise and reevaluate the landscape of strategic decisions and tradeoffs throughout the design process, accommodating the complexities and evolving nature of such engineering systems projects. Including mixed strategies and the possibility of reversible actions in future iterations of this framework could enhance its applicability and accuracy in representing the dynamics of strategic decision-making in complex engineering systems.

5 Conclusion and opportunities for future work

This paper formalizes the concept of fear and greed to describe two dimensions of strategic hindrances in strategic games with more than two players. Coming up from an engineering systems background, we discuss a motivating example highlighting the importance of strategic decisions in our field before introducing our proposed approach to studying strategic decision-making processes while capturing the complexities of incentives and relationships between actors. Building on non-cooperative game theory literature, we present a method to dissect large normal-form games, repackage them into player-reduced instances from each player's standpoint, and embrace a subjectivistic lens to scan each player's overall strategic

stability. Our framework helps us visualize the shape and scope of the strategy dynamics that influence the players' strategic behavior. This framework can help engineering systems researchers and practitioners identify and mitigate the structural forces that hinder the alignment of individual actions toward reaching and sustaining collective harmony.

The proposed framework can be extended to applications in cognitive engineering, particularly in enhancing the understanding and management of strategic decision-making in engineering systems design. By modeling strategic interactions as simplified games, this framework allows for a more explicit conceptualization of the human and socio-technical elements that underlie strategic decision-making processes. In cognitive engineering, where the emphasis is on aligning human cognitive capabilities with the design and operation of complex systems, our approach seeks to simplify the intricate decision-making landscape and make it more accessible and understandable from a human cognitive perspective. The mathematical rigor of the framework provides a structured way to analyze and interpret strategic interactions and reconcile human intuition with quantitative analysis. This aspect is particularly beneficial in enhancing strategic decision-support systems to augment human cognition. Our framework can be integrated into such tools and facilitate the presentation of complex strategic scenarios in a more digestible format, allowing for a more comprehensive evaluation of options and their potential consequences.

In exploring the intricacies of strategic decision-making in the context of collective systems, we encounter several challenges that require further investigation. These challenges present us with valuable opportunities to enrich our understanding and develop frameworks that help sustain collaboration among operationally and managerially independent actors in the design of engineering systems. This framework could serve as a foundation for developing practical incentive mechanisms and network interventions in future work. This direction aims to facilitate the emergence of more favorable strategy dynamics within engineering systems. By understanding the nuances of individual and collective strategic behaviors, researchers and practitioners can identify opportunities to guide actors toward more mutually beneficial and sustainable strategies subtly. For instance, tailored incentives can help align individual motivations with broader system goals, thus encouraging cooperative behaviors that positively contribute to the overall health and efficiency of the system.

Network interventions can also help rectify interactions and dependencies between blocs and coalitions of actors whose strategies significantly influence the strategic hindrance of a specific decision-maker. These interventions could include adjusting communication channels and redefining linked incentives to enhance collaboration and reduce friction in decision-making processes. Applying and refining these approaches within the framework's structure can help steer engineering systems toward more efficient and harmonious operations. Extending this framework in such practical directions can unlock a more grounded and realistic understanding of how strategic decisions play out in complex engineering environments. Ultimately, this could lead to more resilient and adaptive engineering systems capable of navigating the challenges of evolving socio-technical landscapes. We hope this line of work helps unlock new strategies and frameworks that propel engineering systems design to tackle increasingly complex socio-technical challenges.

Acknowledgement

The authors would like to acknowledge Steven Hoffenson, Jordan Suchow, and Alparslan Emrah Bayrak, for their invaluable guidance, feedback, and support as advisors for Ambrosio Valencia-Romero's doctoral dissertation [50], upon which this work builds upon.

Appendix

A.1 Pure-strategy Nash equilibria in normal-form games

Definition A.1 *Let $\mathcal{G}_{\mathcal{N}} = (\mathcal{N}, (\mathcal{S}_i)_{i\in\mathcal{N}}, (U_i)_{i\in\mathcal{N}})$ be a normal-form game. A collective strategy $s^\star = \langle s_i^\star, s_{-i}^\star \rangle \in \mathcal{S}$ is a pure-strategy Nash equilibrium (PNE) if for each player $i \in \mathcal{N}$ and each $s_i \in \mathcal{S}_i$ the following inequality is satisfied:*

$$U_i(s_i^\star, s_{-i}^\star) - U_i(s_i, s_{-i}^\star) \geq 0.$$

We now provide an equivalent definition of PNE for binary normal-form games.

Definition A.2 *Let $\mathcal{G}_{\mathcal{N}}$ be a $n\times 2$ game, and let $\mathcal{S}_i = \{s_i, s_i^\star\}$. A collective strategy $s^\star = \langle s_i^\star, s_{-i}^\star \rangle \in \mathcal{S}$ is a PNE in $\mathcal{G}_{\mathcal{N}}$ if*

$$\min_{i\in\mathcal{N}} \Big[U_i(s_i^\star, s_{-i}) - U_i(s_i, s_{-i}) \Big] \geq 0.$$

We can also characterize PNE in $n\times 2$ games in terms of the rescaled deviation losses ℓ_i in Definition 1. From Eq. (4), we know that

$$U_i(s_i^\star, s_{-i}) - U_i(s_i, s_{-i}^\star) \propto \ell_i(s_i^\star, s_{-i}^\star).$$

We can use this relationship to restate Definition A.2 as follows

Definition A.2a *Let $\mathcal{G}_{\mathcal{N}}$ be a $n\times 2$ game, and let $s_i, s_i^\star \in \mathcal{S}_i = \{\varphi_i, \psi_i\}$, where $s_i \neq s_i^\star$. A collective strategy $s^\star = \langle s_i^\star, s_{-i}^\star \rangle \in \mathcal{S}$ is a PNE in $\mathcal{G}_{\mathcal{N}}$ if*

$$\min_{i\in\mathcal{N}} \ell_i(s_i^\star, s_{-i}^\star) \geq 0.$$

A.2 Types and number of player-reduced binary games

Definition 2 in Section 3 describes the notion of a player-reduced binary game as an event in which a standpoint player i in the normal-form game $\mathcal{G}_{\mathcal{N}}$ makes early conjectures about the strategic behavior of a subset of players $\mathcal{K} \subset \mathcal{N} \setminus \{i\}$. We express this as

$$\mathcal{G}_{\mathcal{N}\setminus\mathcal{K}} = \Big(\mathcal{N}\setminus\mathcal{K}, (\mathcal{S}_i)_{i\in\mathcal{N}\setminus\mathcal{K}}, (U_i(\cdot, s_{\mathcal{K}}))_{i\in\mathcal{N}\setminus\mathcal{K}} \Big). \tag{13}$$

where $s_{\mathcal{K}}$ is the collective action by the players in $\mathcal{K}$. Player i would move on to making conjectures about the strategic behavior of the remaining players in the non-empty subset $\mathcal{J} = \mathcal{N}\setminus(\mathcal{K}\cup\{i\})$. We then characterize structural fear and greed in Definition 3 under two premises: 1) that the players in $\mathcal{J}$ would align their strategies, moving collectively from a status quo $s_{\mathcal{J}} = \varphi_{\mathcal{J}}$ towards the alternative $s_{\mathcal{J}} = \psi_{\mathcal{J}}$; and 2) that the actions of those in $\mathcal{K}$ would remain fixed at $s_{\mathcal{K}} = \sigma_{\mathcal{K}}$.

We refer to the first and second premises as *aligned $\mathcal{J}$-players* and *stationary $\mathcal{K}$-players*, respectively, and we label the player-reduced binary games that meet both conditions as *type I*. Relaxing the second premise allows us to elicit other types of player-reduced binary games that could be observed by player i in $\mathcal{G}_{\mathcal{N}}$. In particular, for $n > 2$, there are games Type II $\mathcal{G}_{\mathcal{N}\setminus\mathcal{K}}$ where one or more players $k \in \mathcal{K}$ whom player i expects will not commit to individual actions $s_k = \sigma_k \in \sigma_{\mathcal{K}}$. We call these player-reduced binary games *type II*, and we refer to the alternative premise associated with it as *reversing $\mathcal{K}$-players*. We respectively denote by $g(n)$ and $g^*(n)$ the numbers of possible type I and type II player-reduced games observed by a player i in $\mathcal{G}_{\mathcal{N}}$, explain their derivation, and provide multiple analytic expressions to calculate them.

Number of player-reduced binary games of type I: stationary K-players

To determine the expression for $g(n)$, we note that

- The value of $g(n)$ is also the number of ways in which we can partition set $\mathcal{N}\backslash\{i\}$ into one non-proper subset ($\mathcal{J}$) and one proper subset ($\mathcal{K}$).
- We must count all possible $\sigma_{\mathcal{K}}$ in Eq. (17) for every $\mathcal{K}\subset\mathcal{N}\backslash\{i\}$. Let $\mathcal{Q}$ be a subset of $\mathcal{K}$ and let its complement be $\mathcal{Q}^c=\mathcal{K}\backslash\mathcal{Q}$. We can use $\mathcal{Q}$ and $\mathcal{Q}^c$ to identify the players in $\mathcal{K}$ who presumably committed to φ_k and ψ_k, respectively — that is, $\sigma_{\mathcal{K}}=\langle\varphi_{\mathcal{Q}},\psi_{\mathcal{Q}^c}\rangle$. It follows that, since either $\mathcal{Q}$ or $\mathcal{Q}^c$ can be empty, the set of possible $\sigma_{\mathcal{K}}$ has the same cardinality as $\mathcal{P}(\mathcal{K})$, the set of all subsets (or power set) of $\mathcal{K}$: $2^{|\mathcal{K}|}$.
- Every possible $\mathcal{K}$ contains between zero and $n-2$ players taken from $\mathcal{N}\backslash\{i\}$. Also, the number of $\mathcal{K}$ of the same size is the number of $|\mathcal{K}|$-combinations of elements in $\mathcal{N}\backslash\{i\}$. Then, the value of $g(n)$ can be expressed as the sum of all the products between $C(n-1,k)$ — the binomial coefficient — and the total number of possible $\sigma_{\mathcal{K}}$, which we know is $2^{|\mathcal{K}|}$.

These give us our first formula for the number of player-reduced games:

$$g(n)=|\{\mathcal{G}_{\mathcal{N}\backslash\mathcal{K}}\}|=\sum_{\mathcal{K}\subset\mathcal{N}\backslash\{i\}}|\mathcal{P}(\mathcal{K})|=\sum_{|\mathcal{K}|=0}^{n-2}2^{|\mathcal{K}|}\cdot\binom{n-1}{|\mathcal{K}|};\tag{A.1}$$

adding and subtracting the $(n-1)$-th term gives:

$$g(n)=-2^{n-1}\cdot\cancel{\binom{n-1}{n-1}}^{1}+\sum_{|\mathcal{K}|=0}^{n-1}2^{|\mathcal{K}|}\cdot(1^{|\mathcal{K}|})\cdot\binom{n-1}{|\mathcal{K}|};$$

which by means of the binomial identity simplifies to

$$g(n)=3^{n-1}-2^{n-1},\qquad\forall n\geq 1.\tag{18}$$

Starting with $g(1)=0$ (no game), Eq (18) returns the Lucas sequence

$$0,1,5,19,65,211,665,2059,6305,19171,\ldots$$

listed as sequence `A001047` in Sloane's *Encyclopedia* [51].

The value of $g(n)$ can also be expressed in terms of Stirling numbers of the second kind, $\left\{{n \atop k}\right\}$, which count the number of ways to partition a set of n elements into k non-empty subsets, as follows [52]:

$$g(n)=2\left\{{n \atop 3}\right\}+\left\{{n \atop 2}\right\};$$

where $\left\{{n \atop 3}\right\}=\frac{1}{2}(1+3^{n-1})-2^{n-1}$ and $\left\{{n \atop 2}\right\}=2^{n-1}-1$. For instance, if $\mathcal{N}=\{i,j,k\}$,

$$\begin{aligned}&\left\{{3 \atop 3}\right\}=1 &&\because\mathcal{N}=\{i\}\cup\{j\}\cup\{k\};\\ \text{and}\quad&\left\{{3 \atop 2}\right\}=3 &&\because\mathcal{N}=\{i\}\cup\{j,k\}=\{i,j\}\cup\{k\}=\{i,k\}\cup\{j\}.\end{aligned}$$

So $g(3)=2(1)+3=5$. In the context of the static games with $n>2$ treated in this work, the term $\left\{{n \atop 3}\right\}$ is associated with the number of player-reduced games where the individual strategies of at least two players other than i differ, i.e. $s_j\neq s_k$, where $j,k\in\mathcal{N}\backslash\{i\}$; and the term $\left\{{n \atop 2}\right\}$ is associated with the player-reduced games $\mathcal{G}_{\mathcal{N}\backslash\mathcal{K}}$ whose strategy profiles include the diagonal collective strategy $s_{\text{-}i}\in\{\varphi_{\text{-}i},\psi_{\text{-}i}\}$, i.e. $s_1=\cdots=s_{i-1}=s_{i+1}=\cdots=s_n$.

The directed graphs in Fig A.1 describe the player-reduced binary games with fixed $s_{\mathcal{K}}=\sigma_{\mathcal{K}}$, $\mathcal{K}\subset\mathcal{N}\backslash\{i\}$, for $2\leq n\leq 4$. The vertices represent the possible collective strategies $s_{\text{-}i}=\in\mathcal{S}_i^{n-1}$ that the standpoint player i faces. Each arc represents one player-reduced game; the source and target vertices are associated with structural fear and structural greed, respectively.

n	$\mathcal{J}$	$\mathcal{K}$	$\sigma_{\mathcal{K}}$	$\mathcal{G}_{\mathcal{N}\setminus\mathcal{K}}$
2	2	$\emptyset$	n/a	A
3	2	3	φ_3 ψ_3	B C
	3	2	φ_2 ψ_2	D E
	-1	$\emptyset$	n/a	F
4	2	{3,4}	$\langle\varphi_3,\varphi_4\rangle$ $\langle\psi_3,\varphi_4\rangle$ $\langle\varphi_3,\psi_4\rangle$ $\langle\psi_3,\psi_4\rangle$	G H I J
	3	{2,4}	$\langle\varphi_2,\varphi_4\rangle$ $\langle\psi_2,\varphi_4\rangle$ $\langle\varphi_2,\psi_4\rangle$ $\langle\psi_2,\psi_4\rangle$	K L M N
	4	{2,3}	$\langle\varphi_2,\varphi_3\rangle$ $\langle\psi_2,\varphi_3\rangle$ $\langle\varphi_2,\psi_3\rangle$ $\langle\psi_2,\psi_3\rangle$	O P Q R
	{2,3}	4	φ_4 ψ_4	S T
	{2,4}	3	φ_3 ψ_3	U V
	{3,4}	2	φ_2 ψ_2	W X
	-1	$\emptyset$	n/a	Y

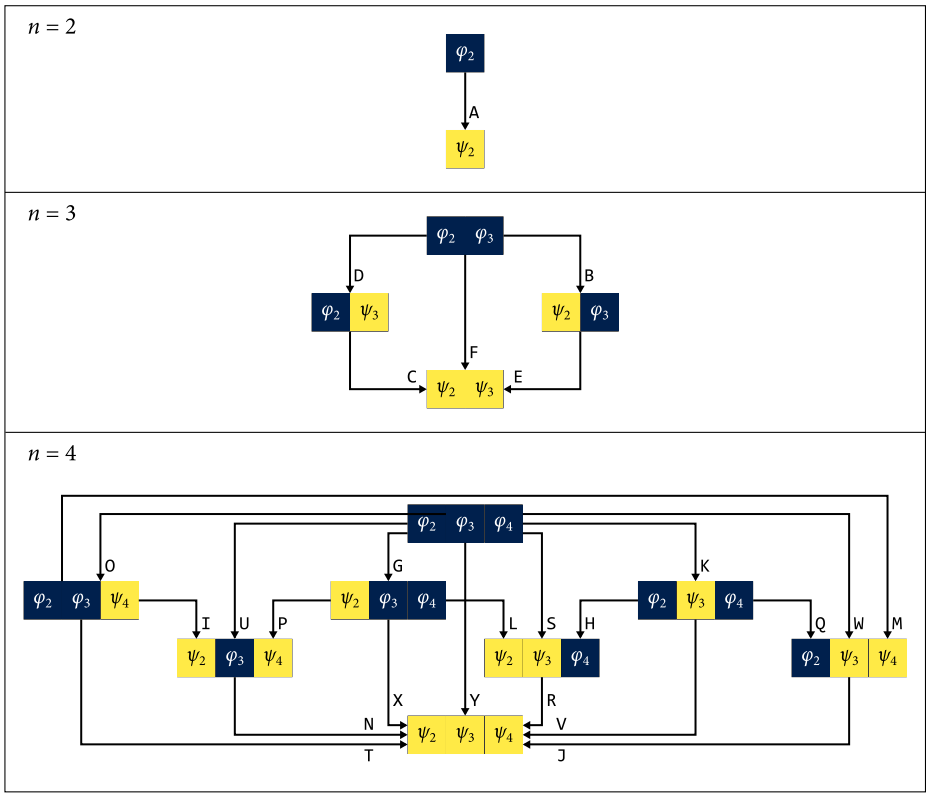


Fig A.1. Type I player-reduced normal-form games $\mathcal{G}_{\mathcal{N}\setminus\mathcal{K}}$ observed by each player in $n\times 2$ games for $2<n<4$. Every player-reduced game is formed by the actions of standpoint player $i=1$, the coordinated actions of the players in set $\mathcal{J}=\mathcal{N}\setminus(\mathcal{K}\cup\{i\})$, and the presumably set in stone action $\sigma_{\mathcal{K}}$ by all players in $\mathcal{K}$. On the right, we create directed graphs made of 2^{n-1} vertices, one for every collective strategy $s_{-i}=\langle s_{\mathcal{J}}, s_{\mathcal{K}}\rangle$, and connect them according to the possible deviations under the premises of aligned $\mathcal{J}$-players and stationary $\mathcal{K}$-players. Then, each edge represents a possible player-reduced game (listed from A to Y); and the source and target vertices respectively represent the components of s_{-i} in $\ell_i(s_{-i})$ needed to compute the values of F_i and G_i per Eqs. (15) and (16).

Number of player-reduced binary games of type II: reversing K-players

We derive $g^*(n)$ by analyzing the discrete geometry of all possible transitions between two different collective strategies s_{-i} and s^*_{-i} that can be observed by player i. Either action can be equal to any possible combinations of $n-1$ individual strategies $s_j\in\mathcal{S}_j=\{\varphi_j,\psi_j\}$. So, the total number of possible transitions is

$$|\mathcal{S}_j^{n-1}\times\mathcal{S}_j^{n-1}|-|\mathcal{S}_j^{n-1}|=(2^{n-1})\cdot(2^{n-1})-2^{n-1}=4^{n-1}-2^{n-1}.$$

In calculating the number of type I player-reduced binary games $g(n)=3^{n-1}-2^{n-1}$, we assumed the collective action $s_{\mathcal{J}}$ by the aligned $\mathcal{J}$-players moved in the direction $\varphi_{\mathcal{J}}\to\psi_{\mathcal{J}}$. Asserting the premise of reversing $\mathcal{K}$-players, we note type I player-reduced binary games also extend to $s_{\mathcal{J}}$ moving in the opposite direction, namely $\psi_{\mathcal{J}}\to\varphi_{\mathcal{J}}$. This means that the total number of type II player-reduced binary games must be equal to

$$\begin{aligned} g^*(n) &= 4^{n-1}-2^{n-1}-2\cdot g(n) \\ &= 4^{n-1}-2(3^{n-1})+2^{n-1}. \end{aligned} \tag{A.2}$$

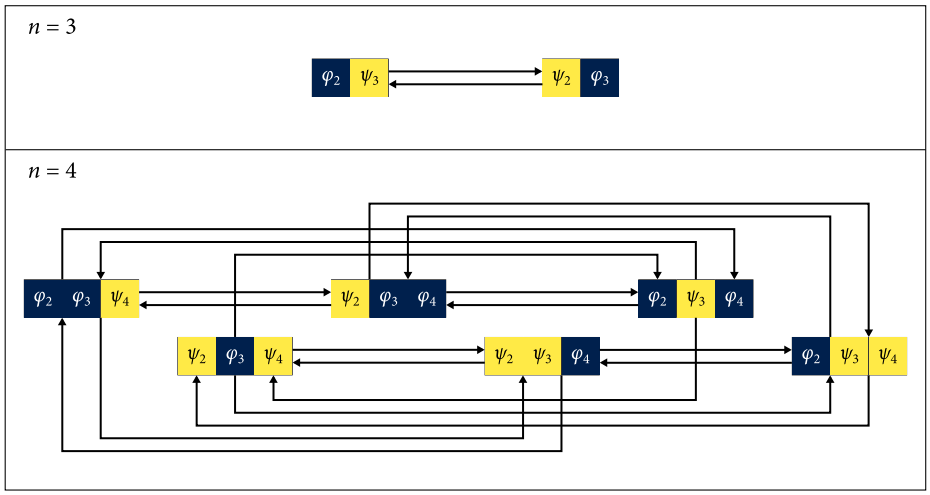


Fig A.2. Transitions between non-diagonal collective strategies representing the number of player-reduced games with anti-coordination for $n = 3$ and $n = 4$. Let $\mathcal{J} \subset \mathcal{N} \setminus \{i\}$ and $\mathcal{K} = \mathcal{N} \setminus (\mathcal{J} \cup \{i\})$; and let Q be a subset of $\mathcal{K}$, with complement $Q^c = \mathcal{K} \setminus Q$. As all players in $\mathcal{J}$ deviate from $\varphi_{\mathcal{J}} \to \psi_{\mathcal{J}}$, all players in Q deviate from $\psi_Q \to \varphi_Q$ while the players in Q^c, if any, keep their strategies fixed. The number of player-reduced games following the aforementioned description of anti-coordination is $4^{n-1} - 2(3^{n-1}) - 2^{n-1}$. This figure complements Fig A.1.

Fig A.2 shows the directed graphs describing the associations between collective strategies that result in type II player-reduced games and that complement the type I games presented in Fig A.1 for $n = 3$ and $n = 4$ — note that $g^*(2) = 0$ consistent with $\mathcal{K} = \emptyset$. Combined, the total number of player-reduced binary games with stationary and reversing $\mathcal{K}$-players assuming the aligned $\mathcal{J}$-players move from $\varphi_{\mathcal{J}}$ towards $\to \psi_{\mathcal{J}}$ is

$$g(n) + g^*(n) = 4^{n-1} - 3^{n-1}.$$

We can also count $g(n)$ and $g^*(n)$ by breaking apart the transition between any two collective strategies into the moves by each player and labeling the number of existing intersecting collective actions transitioning in the same direction. This approach is demonstrated using square grids in Fig A.3 for $n \in [2 \,..\, 4]$. The result shows the number $g(n)$ emerge as the number of states in the Hanoi graph analog of the Sierpiński triangle — from the Tower of Hanoi puzzle with 3 towers and $n-1$ disks — minus the 2^{n-1} states along the diagonal $s_j = s_j^*$; while the number $g^*(n)$, counting all type II player-reduced binary games, emerge as two times the number of even values in Pascal's triangle up to row 2^{n-1}.

A.3 Distribution of deviation losses in strategic hindrance spaces

We characterize the incidence of the rescaled deviation losses $\ell_i^*(s_{-i})$ on $\mathbf{H}_i^{(n)}(\varphi, \psi)$, player i's strategic hindrance in Definition 4, across observed type I player-reduced games. Let us set

- $\mathcal{N}_\varphi$ and $\mathcal{N}_\psi$ as the subsets of players in $\mathcal{N}$ that play φ and ψ, respectively;
- $n_\varphi = |\mathcal{N}_\varphi|$ and $n_\psi = |\mathcal{N}_\psi|$ as the numbers of players that adopt such strategies; and
- $n_{\varphi \setminus i} = |\mathcal{N}_\varphi \setminus \{i\}|$ and $n_{\psi \setminus i} = |\mathcal{N}_\psi \setminus \{i\}|$, so $n_{\varphi \setminus i} + n_{\psi \setminus i} = n - 1$.

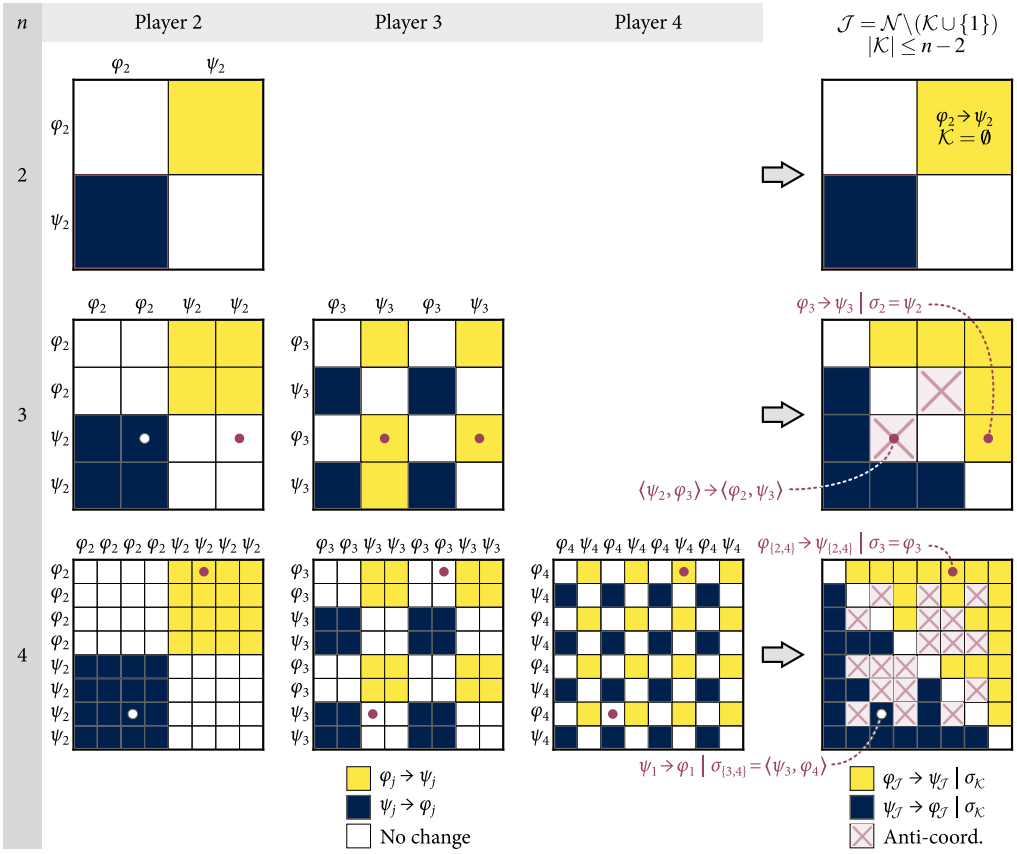


Fig A.3. Alternative approach to counting the number of player-reduced binary games observed by one player (viz. $i = 1$) in an $n \times 2$ game, for $n \in [2 \,..\, 4]$. Each grid on the rightmost column shows all possible transitions from s_{-i} and s^*_{-i}; these grids are broken apart into individual transitions from s_j and $s^*_j \in \mathcal{S}_j$ — $\varphi_j \to \psi_j$ in yellow/light gray and $\psi_j \to \varphi_j$ in dark blue/gray, going from row to column; and either null transitions of stationary $\mathcal{K}$-players — blank cells, if any, with "no changes" — or reversed transition of reversing $\mathcal{K}$-players — marked with an $\times$.

Recall that $\mathcal{J} = \mathcal{N} \setminus (\mathcal{K} \cup \{i\})$ and $|\mathcal{J}| \in [1 \,..\, n-1]$. Replacing $k = |\mathcal{K}| = n - 1 - |\mathcal{J}|$ in Eq. (A.1), gives us an expression for $g(n)$ in terms of $\mathcal{J}$:

$$g(n) = \sum_{|\mathcal{J}|=1}^{n-1} 2^{n-1-|\mathcal{J}|} \cdot \binom{n-1}{n-1-|\mathcal{J}|} = \sum_{|\mathcal{J}|=1}^{n-1} (2^{|\mathcal{J}|} - 1) \cdot \binom{n-1}{|\mathcal{J}|} = \sum_{\mathcal{J} \subseteq \mathcal{N} \setminus \{i\}} |\mathcal{P}^+(\mathcal{J})|, \tag{A.3}$$

where $\mathcal{P}^+(\mathcal{J})$ is the set of all non-empty subsets of $\mathcal{J}$. From the definition of strategic hindrance in Eq. (17), we know $\dim_{\mathbb{R}}(\langle F_i \rangle) = \dim_{\mathbb{R}}(\langle G_i \rangle) = g(n)$ and all players in $\mathcal{J}$ would coordinate on either $\varphi_{\mathcal{J}}$ or $\psi_{\mathcal{J}}$. Let $|\mathcal{J}| = n_{\varphi \setminus i}$ if $s_{\mathcal{J}} = \varphi_{\mathcal{J}}$ and $|\mathcal{J}| = n_{\psi \setminus i}$ if $s_{\mathcal{J}} = \psi_{\mathcal{J}}$. We can write

$$\begin{aligned} g(n) &= \frac{1}{2}\left[\dim_{\mathbb{R}}(\langle F_i \rangle) + \dim_{\mathbb{R}}(\langle G_i \rangle)\right] \\ &= \frac{1}{2}\left[\sum_{n_{\varphi \setminus i}=0}^{n-1} (2^{n_{\varphi \setminus i}} - 1) \cdot \binom{n-1}{n_{\varphi \setminus i}}\right] + \frac{1}{2}\left[\sum_{n_{\psi \setminus i}=0}^{n-1} (2^{n_{\psi \setminus i}} - 1) \cdot \binom{n-1}{n_{\psi \setminus i}}\right] \end{aligned} \tag{A.4}$$

after adding and subtracting the 0-th term to the summation (without affecting the interpretation of $g(n)$ using Eq. (A.3), as $2^0 - 1 = 0$).

$\gamma_{\times C(n-1,k)}$		k											
n	$g(n)$	0	1	2	3	4	5	6	7	8	9	10	11
1	0	$0_{\times 1}$											
2	1	$1_{\times 1}$	$1_{\times 1}$										
3	5	$3_{\times 1}$	$2_{\times 2}$	$3_{\times 1}$									
4	19	$7_{\times 1}$	$4_{\times 3}$	$4_{\times 3}$	$7_{\times 1}$								
5	65	$15_{\times 1}$	$8_{\times 4}$	$6_{\times 6}$	$8_{\times 4}$	$15_{\times 1}$							
6	211	$31_{\times 1}$	$16_{\times 5}$	$10_{\times 10}$	$10_{\times 10}$	$16_{\times 5}$	$31_{\times 1}$						
7	665	$63_{\times 1}$	$32_{\times 6}$	$18_{\times 15}$	$14_{\times 20}$	$18_{\times 15}$	$32_{\times 6}$	$63_{\times 1}$					
8	2,059	$127_{\times 1}$	$64_{\times 7}$	$34_{\times 21}$	$22_{\times 35}$	$22_{\times 35}$	$34_{\times 21}$	$64_{\times 7}$	$127_{\times 1}$				
9	6,305	$255_{\times 1}$	$128_{\times 8}$	$66_{\times 28}$	$38_{\times 56}$	$30_{\times 70}$	$38_{\times 56}$	$66_{\times 28}$	$128_{\times 8}$	$255_{\times 1}$			
10	19,171	$511_{\times 1}$	$256_{\times 9}$	$130_{\times 36}$	$70_{\times 84}$	$46_{\times 126}$	$46_{\times 126}$	$70_{\times 84}$	$130_{\times 36}$	$256_{\times 9}$	$511_{\times 1}$		
11	58,025	$1{,}023_{\times 1}$	$512_{\times 10}$	$258_{\times 45}$	$134_{\times 120}$	$78_{\times 210}$	$62_{\times 252}$	$78_{\times 210}$	$134_{\times 120}$	$258_{\times 45}$	$512_{\times 10}$	$1{,}023_{\times 1}$	
12	175,099	$2{,}047_{\times 1}$	$1{,}024_{\times 11}$	$514_{\times 55}$	$262_{\times 165}$	$142_{\times 330}$	$94_{\times 462}$	$94_{\times 462}$	$142_{\times 330}$	$262_{\times 165}$	$514_{\times 55}$	$1{,}024_{\times 11}$	$2{,}047_{\times 1}$

Fig A.4. Number of reduced-player games, $g(n)$ from Eq. (A.1), and values of $\gamma(k, n-1-k)$, $k \in \{n_{\varphi\setminus i}, n_{\psi\setminus i}\}$, from Eq. (A.5) observed by each player in games with up to $n = 12$ players. The subscripts denote the occurrences of each γ and are equal to $C(n-1,k)$. For every integer $n \geq 1$, the sum of the $\gamma(k, n-1-k) \cdot C(n-1,k)$ products equal $2g(n)$ per Eq. (20). For $n = 3$, notice that the deviation losses $\ell_i^*(\varphi_2, \psi_3)$ and $\ell_i^*(\psi_2, \varphi_3)$ would be weighted twice ($k = 1$); this is consistent with the directed graph representation of the player-reduced games for $n = 3$, in Fig A.1, where both vertices $\langle \varphi_2, \psi_3 \rangle$ and $\langle \psi_2, \varphi_3 \rangle$ are once a source and once a target.

Redefining $g(n)$ in terms of $n_{\varphi\setminus i}$ and $n_{\psi\setminus i}$ allows us to count the number of player-reduced games where a specific number of players in $\mathcal{N} \setminus \{i\}$ select one strategy or the other. Adding one to $n_{\varphi\setminus i}$ means subtracting one from $n_{\psi\setminus i}$, and vice versa. Let $n_{\varphi\setminus i} = k$ and $n_{\psi\setminus i} = n-1-k$; we can rewrite Eq. (A.4) as

$$g(n) = \frac{1}{2}\left[\sum_{k=0}^{n-1}(2^k - 1)\cdot\binom{n-1}{k}\right] + \frac{1}{2}\left[\sum_{k=0}^{n-1}(2^{n-1-k} - 1)\cdot\binom{n-1}{n-1-k}\right]$$

$$g(n) = \frac{1}{2}\left[\sum_{k=0}^{n-1}\gamma(k, n-1-k)\cdot C(n-1,k)\right] \tag{20}$$

where $C(n-1,k) = \binom{n-1}{k} = \binom{n-1}{n-1-k}$ and the function

$$\gamma(k, n-1-k) = 2^k + 2^{n-1-k} - 2, \tag{A.5}$$

or

$$\gamma(n_{\varphi\setminus i}, n_{\psi\setminus i}) = 2^{n_{\varphi\setminus i}} + 2^{n_{\psi\setminus i}} - 2, \tag{19}$$

counts how many entries of each individual strategic hindrance space of size $g(n) \times 2$ are equal to $\ell_i^*(s_{-i})$ based solely on the number of players in $\mathcal{N} \setminus \{i\}$ that play φ and those that play ψ. Values of $\gamma(k, n-1-k)$ and their occurrences for games with up to $n = 12$ players are provided in Fig A.4. The triangular arrays arising from the computation of $\gamma(n-1-k, k)$ and the product $\gamma(n-1-k, k) \cdot C(n-1,k)$ can be found in Sloane's *Encyclopedia* [51] under the catalog numbers `A350770` and `A350771`, respectively.

A.4 Strategy dynamics of 3×2 social dilemma games

This section calculates the strategic hindrance spaces of nine social dilemmas in normal form and $n = 3$. The strategic hindrance of every player in each game are the same; thus, their strategy dynamics can be considered symmetrical. All of the strategic hindrance spaces are compiled in Fig A.5.

Volunteer's dilemma

Each player can either volunteer for the common good (ψ) or do nothing and possibly benefit from the contributions of others (φ) — helping maintain the status quo. Player i's payoff in this game is defined as

$$U_i(s_i, s_{\mathcal{J}}, s_{\mathcal{K}}) = \begin{cases} 0 & \text{if } s_i = \psi_i \\ a & \text{if } s_i = \varphi_i \wedge s_{\mathcal{J}} = \psi_{\mathcal{J}} \\ -b & \text{otherwise,} \end{cases} \tag{A.6}$$

where a and b are both positive real numbers. Each player in the volunteer's dilemma in Eq. (A.6) has one *perfectly limited strategy*, viz. ψ_i, which yields constant $U_i(\psi_i, s_{-i})$ for any $s_{-i} \in \mathcal{S}_i^{n-1}$ (i.e. the payoff of volunteering does not depend on whether others also volunteer or not) restricting interactive effects [35]. There are exactly n PNE in this game, each of them equal to $s^\star = \langle \psi_i, \varphi_{-i} \rangle$ for every $i \in \mathcal{N}$: every time one and only one player has volunteered.

Incentives modeled after a volunteer's dilemma can lead to maintaining the status quo becoming the dominant strategy. After one actor has chosen a cooperative strategy, such as investing in integrability in the context of the urban transit system, others, acting economically rational, may decide to seize such opportunity and take advantage of the volunteer's contribution without reciprocating. As all actors would prefer to keep their systems closed to avoid the disadvantages of being the only one to open their system, integration could fail due to collective inaction. Defection dynamics are anticipated to emerge and dominate the conflict as the number of players increases.

Using Eqs. (14)–(16) to obtain the rescaled deviation losses in this game for every player $i \in \mathcal{N}$ — with the help of Table 1:

$$\begin{aligned}
\ell_i^*(\varphi_j, \varphi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \varphi_j, \varphi_k) - U_i(\psi_i, \varphi_j, \varphi_k) \Big) = \big(-b - 0\big)/A_i = -b/A_i, \\
\ell_i^*(\varphi_j, \psi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \varphi_j, \psi_k) - U_i(\psi_i, \varphi_j, \psi_k) \Big) = \big(\ \ a - 0\big)/A_i = \ \ a/A_i, \\
\ell_i^*(\psi_j, \varphi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \psi_j, \varphi_k) - U_i(\psi_i, \psi_j, \varphi_k) \Big) = \big(\ \ a - 0\big)/A_i = \ \ a/A_i, \\
\ell_i^*(\psi_j, \psi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \psi_j, \psi_k) - U_i(\psi_i, \psi_j, \psi_k) \Big) = \big(\ \ a - 0\big)/A_i = \ \ a/A_i.
\end{aligned}$$

where $A_i = \max U_i - \min U_i = a + b$. Using Eq. (21), the strategic hindrance is:

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{a+b} \begin{bmatrix} -b & +a & -b & +a & -b \\ +a & +a & +a & +a & +a \end{bmatrix}^{\mathrm{T}}. \tag{A.7}$$

Three of the $g(3) = 5$ player-reduced binary games exhibit coexistence while the remaining two exhibit defection. Fig A.5 shows the strategic hindrance space of the 3×2 volunteer's dilemma for $a = 1$ and $b = 2$.

Diner's dilemma

Three graduate students go for lunch and agree to split the bill equally before ordering. When the menu arrives, they choose individually between ordering the combo plate with extra protein (φ) or just a fresh empanada (ψ). The difference in cost between dishes, c (in utility units), is

believed to be greater than the difference in satisfaction, b; that is $0 < b < c$. However, they might be tempted to order the more satisfying combo plate if they suspect everyone else will order empanadas, implying that $b > c/n$ — which guarantees $U_i(\varphi_i, s_{-i}) > U_i(\psi_i, s_{-i})$, signaling pure defection dynamics. Let $\mathcal{N}_\varphi \subseteq \mathcal{N}$ be the subset of diners ordering the combo plate, $n_\varphi = |\mathcal{N}_\varphi|$, and let $n_{\varphi\setminus i} = |\mathcal{N}_\varphi \setminus \{i\}|$; player i's payoff is modeled as

$$U_i(s_i, s_{-i}) = \begin{cases} b - n_\varphi \left(\dfrac{c}{n}\right) & \text{if } s_i = \varphi_i \\ -n_{\varphi\setminus i} \left(\dfrac{c}{n}\right) & \text{otherwise.} \end{cases} \tag{A.8}$$

From Eq. (A.8), we notice that the difference in individual payoff between choosing $s_i = \varphi_i$ and choosing $s_i = \psi_i$ for a given $s_{-i} \in \mathcal{S}_i^{n-1}$ always equals $\max U_i = b - c/n > 0$. This game has a single PNE (viz. φ) for any $n \geq 2$. The 3-player diner's dilemma in Fig A.5 uses $b = 1$ and $c = 2$. Using Eqs. (14)–(16) and Table 1:

$$\begin{aligned}
\ell_i^*(\varphi_j, \varphi_k) &= A_i^{-1} \cdot \left[\left(1 - 3 \cdot \tfrac{2}{3}\right) - \left(-2 \cdot \tfrac{2}{3}\right) \right] = +(1/3)/A_i, \\
\ell_i^*(\varphi_j, \psi_k) &= A_i^{-1} \cdot \left[\left(1 - 2 \cdot \tfrac{2}{3}\right) - \left(-1 \cdot \tfrac{2}{3}\right) \right] = +(1/3)/A_i, \\
\ell_i^*(\psi_j, \varphi_k) &= A_i^{-1} \cdot \left[\left(1 - 2 \cdot \tfrac{2}{3}\right) - \left(-1 \cdot \tfrac{2}{3}\right) \right] = +(1/3)/A_i, \\
\ell_i^*(\psi_j, \psi_k) &= A_i^{-1} \cdot \left[\left(1 - 1 \cdot \tfrac{2}{3}\right) - \left(\ \ 0 \cdot \tfrac{2}{3}\right) \right] = +(1/3)/A_i;
\end{aligned}$$

where $A_i = \max U_i - \min U_i = b + c - 2 \cdot c/n = 5/3$. All rescaled deviation losses are signed positive; ordering the combo plate instead of just a pastry is always the dominant strategy. The strategic hindrance is, using Eq. (21):

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{5/3} \begin{bmatrix} +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} \\ +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} \end{bmatrix}^{\mathrm{T}} = \frac{1}{5} \begin{bmatrix} +1 \\ +1 \end{bmatrix}_{\times 5}^{\mathrm{T}}.$$

Glance and Huberman [53] introduced the diner's dilemma as a potential extension of the PD game to study the emergence of cooperation in settings with more than two players. A player-reduced binary game is a classical 2×2 PD game if

$$U_i(\varphi_i, \psi_{\mathcal{J}}, \sigma_{\mathcal{K}}) > U_i(\psi_i, \psi_{\mathcal{J}}, \sigma_{\mathcal{K}}) > U_i(\varphi_i, \varphi_{\mathcal{J}}, \sigma_{\mathcal{K}}) > U_i(\psi_i, \varphi_{\mathcal{J}}, \sigma_{\mathcal{K}}), \tag{A.9}$$

which is equivalent to the relationship $T > R > P > S$ from Section 2.2 and Fig 1. Setting

$$\frac{1}{n} < \frac{b}{c} < \frac{n+1}{2 \cdot n}$$

guarantees that all player-reduced games in a diner's dilemma are classical 2×2 PD games.

Public goods game (PGG)

The public goods game (PGG) with linear variable contributions is defined by

$$U_i(s_i, s_{-i}) = \begin{cases} 1 + n_{\psi\setminus i} \left(\dfrac{a_i}{n}\right) & \text{if } s_i = \varphi_i \\ n_\psi \left(\dfrac{a_i}{n}\right) & \text{otherwise,} \end{cases} \tag{A.10}$$

where $n_\psi = |\mathcal{N}_\psi|$, $n_{\psi\setminus i} = |\mathcal{N}_\psi \setminus \{i\}|$, and $\mathcal{N}_\psi \subseteq \mathcal{N}$ is the subset of players that contribute one token to the public pool (ψ). Not contributing to the public pool is represented by strategy φ_i. Factor $a_i > 0$ multiplies the tokens in the pool before dividing them evenly among all players. Similar to Eq. (A.8), the image of Eq. (A.10) contains only two values per player, meaning the absolute individual deviation losses are constant (and equal to $1 - a_i/n$), and each player

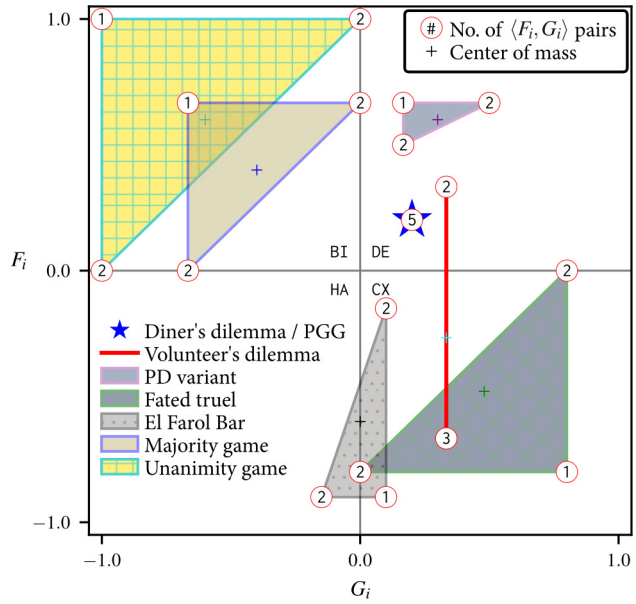


Fig A.5. A player fear and greed value space in seven 3×2 social dilemmas normal-form games. Some fear and greed value pairs are repeated.

observes dominance strategy dynamics depending on the value of a_i. For $n = 3$ and setting $a_i = 2$, Eq. (21) gives the strategic hindrance space

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{5/3} \begin{bmatrix} +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} \\ +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} & +\frac{1}{3} \end{bmatrix}^{\mathrm{T}} = \frac{1}{5} \begin{bmatrix} +1 \\ +1 \end{bmatrix}^{\mathrm{T}}_{\times 5},$$

where $A_i = \max U_i - \min U_i = 1 + a_i - 2 \cdot a_i/n = 5/3$, which exhibits the same pure defection dynamics as those of the diner's dilemma example above and whose strategic hindrance is depicted in Fig A.5 using Eq. (A.8).

Similar to the diner's dilemma, we can turn every player-reduced game in a PGG into a classical 2×2 PD game if we set, $\forall i \in \mathcal{N}$:

$$\frac{n}{2} < a_i < n,$$

which ensures that every level of mutual cooperation between two or more players will always yield a greater utility than mutual defection, yet guarantee that the temptation to defect unilaterally remains higher, satisfying Eq. (A.9). Comparing Eqs. (A.8) and (A.10), we can define any diner's dilemma into a PGG by calculating

$$a_i = \frac{n-1}{2 \cdot (b/c) + 1 - (3/n)}.$$

For values of a_i greater than n, the PGG yields harmony dynamics. Whenever $a_i \neq n$ for every $i \in \mathcal{N}$ and $n \geq 2$, the PGG has only one PNE, with each $s_i^\star$ equal to either φ_i if $\mathrm{sgn}(a-n) = -1$ or ψ_i if $\mathrm{sgn}(a-n) = +1$ — where $\mathrm{sgn} : \mathbb{R} \mapsto \mathbb{Z}$ is the signum function which returns 1 if $x > 0$, -1 if $x < 0$, and 0 if $x = 0$. If $a_i = n$, the game is that of indifference (as $n_\psi = 1 + n_{\psi \setminus i}$), and $\mathcal{S}^\star$ is trivially equal to 2^n.

A prisoner's dilemma variant

Here, we introduce an alternative version of a PD game with increased incentives for defection. Prosecutors separately promise three alleged co-conspirators in a crime a lighter treatment if they snitch on their associates (φ) instead of remaining silent (ψ). All three suspects will face a minimum sentence if they keep quiet. If some snitch while others refuse to talk, the tipsters will be acquitted, while the quiet ones will face a combined sentence of c times the minimum. But if they all confess, they will all face b times the minimum sentence. Let $1 < b < c$; player i's payoff in this game is defined as

$$U_i(s_i, s_{\mathcal{J}}, s_{\mathcal{K}}) = \begin{cases} 0 & \text{if } s_i = \varphi_i \wedge s_{\mathcal{J}} = \psi_{\mathcal{J}} \\ -1 & \text{if } \langle s_i, s_{\mathcal{J}}, s_{\mathcal{K}} \rangle = \psi \\ -b & \text{if } \langle s_i, s_{\mathcal{J}}, s_{\mathcal{K}} \rangle = \varphi \\ -c/n_\psi & \text{otherwise.} \end{cases}$$

The rescaled deviation losses in the 3×2 version of this game are

$$\begin{aligned}
\ell_i^*(\varphi_j, \varphi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \varphi_j, \varphi_k) - U_i(\psi_i, \varphi_j, \varphi_k) \Big) = \big(-b + c \quad \big)/A_i = \ (c-b)/A_i, \\
\ell_i^*(\varphi_j, \psi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \varphi_j, \psi_k) - U_i(\psi_i, \varphi_j, \psi_k) \Big) = \big(\quad 0 + c/2 \big)/A_i = +(c/2)/A_i, \\
\ell_i^*(\psi_j, \varphi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \psi_j, \varphi_k) - U_i(\psi_i, \psi_j, \varphi_k) \Big) = \big(\quad 0 + c/2 \big)/A_i = +(c/2)/A_i, \\
\ell_i^*(\psi_j, \psi_k) &= A_i^{-1} \cdot \Big(U_i(\varphi_i, \psi_j, \psi_k) - U_i(\psi_i, \psi_j, \psi_k) \Big) = \big(\quad 0 + 1 \quad \big)/A_i = \qquad +1/A_i.
\end{aligned}$$

where $A_i = c$. Fig A.5 shows the individual strategic hindrance in a three-person prisoner's dilemma, with $b = 2$ and $c = 6$, calculated via Eq. (21):

$$\begin{aligned}
\mathbf{H}_i^{(3)}(\varphi, \psi) &= \frac{1}{c} \begin{bmatrix} c-b & c/2 & c-b & c/2 & c-b \\ c/2 & +1 & c/2 & +1 & +1 \end{bmatrix}^{\mathrm{T}} \\
&= \frac{1}{6} \begin{bmatrix} +4 & +3 & +4 & +3 & +4 \\ +3 & +1 & +3 & +1 & +1 \end{bmatrix}^{\mathrm{T}}.
\end{aligned}$$

This take on the n-player prisoner's dilemma game is a simplification of Weil's [54] description with a stronger condition on the payoff dominance of defection modeled after the original, classical interpretation for two actors by Merrill Flood and Melvin Dresher, and contextualized by Albert W. Tucker [55]. Condition $b < c$ guarantees that unanimous defection (φ) is always the only PNE and that it is suboptimal regardless of the number of players. And setting $b > 1$ guarantees that unanimous cooperation (ψ) is always preferred over φ. However, in contrast with the diner's dilemma and the PGG — in which every player-reduced game can be turned into a classical 2×2 PD game — the reward of staying silent in this model is lower than the punishment from defecting any time at least one suspect confesses. These characteristics also differ from alternative generalizations of the prisoner's dilemma such as those in the works by Goehring and Kahan [56] and Diekmann [57] that reduce the problem to an n-player volunteer's dilemma, which has n PNE rather than only one.

Fated truel

Consider three mutual paintball adversaries, Blondie, Angel Eyes, and Tuco, each with one paintball left, holding their markers at point-blank. Each of them must decide which of the other two rivals to fire at, the one on their left (φ_i) or the one on their right (ψ_i). To make it out clean, a standpoint player must first coincide with a second player on targetting their remaining rival and pray for the latter to target the second player. When two players target the same rival, they

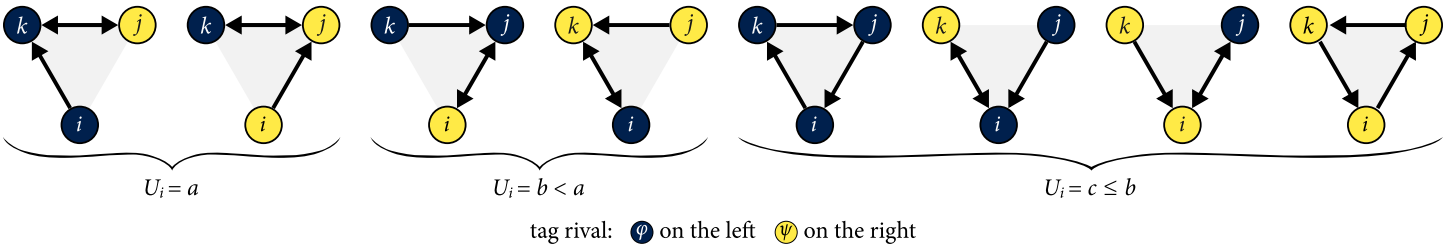


Fig A.6. Possible outcomes and individual payoffs in a fated truel game (player i's standpoint) per Eq. (A.11). Each arrow points at the players' targets.

are anti-coordinating their strategies in opposite directions to shooting each other. Specifically, two consecutive players Tuco (i) and Blondie ($i+1$), the former on the left and the latter on the right, must play φ_i and ψ_{i+1} to shoot Angel Eyes ($i-1$). Angel Eyes might still manage to mark one of them, albeit with less effectiveness. For Tuco to come out of the standoff clean, Angel Eyes must target Blondie by playing (φ_{i-1}). Let a be the expected payoff for making it out clean, b the expected payoff for firing at the same target as another player with the risk of being hit by that target, and c the expected payoff of being shot at; it follows that $a > b \geq c \in \mathbb{R}$. Player i's payoff contingent on actions by the players on their left ($i-1$) and right (i+1) is defined as

$$U_i(s_i, s_{i+1}, s_{i-1}) = \begin{cases} a & \text{if } s_{i+1} = \psi_{i+1} \wedge s_{i-1} = \varphi_{i-1} \\ b & \text{if } s_i \neq s_{i+1} = s_{i-1} \\ c & \text{otherwise.} \end{cases} \tag{A.11}$$

The eight possible outcomes in this game are depicted in Fig A.6. Two possible outcomes could see all players defeated in the standoff: φ (everyone targets their rival on their right) and ψ (their rival on their left). The generic individual strategic hindrance in a fated truel game is provided in Fig A.5 using $a = 2$, $b = 1$, and $c = -3$. The deviation losses in this game are:

$$\begin{aligned}
\ell_i^*(\varphi_{i+1}, \varphi_{i-1}) &= A_i^{-1} \cdot \left[\, c - b \,\right] = A_i^{-1} \cdot \left[-3 - \quad 1 \,\right] = -4/A_i \\
\ell_i^*(\varphi_{i+1}, \psi_{i-1}) &= A_i^{-1} \cdot \left[\, c - c \,\right] = A_i^{-1} \cdot \left[-3 - (-3)\right] = \quad 0 \\
\ell_i^*(\psi_{i+1}, \varphi_{i-1}) &= A_i^{-1} \cdot \left[\, a - a \,\right] = A_i^{-1} \cdot \left[\; 2 - \quad 2 \,\right] = \quad 0 \\
\ell_i^*(\psi_{i+1}, \psi_{i-1}) &= A_i^{-1} \cdot \left[\, b - c \,\right] = A_i^{-1} \cdot \left[\; 1 - (-3)\right] = +4/A_i
\end{aligned}$$

While $A_i = a - c = 2 - (-3) = 5$. The space of fear and greed values is:

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{5} \begin{bmatrix} -4 & 0 & -4 & 0 & -4 \\ 0 & +4 & 0 & +4 & +4 \end{bmatrix}^{\mathrm{T}}.$$

Since the payoff that player i would obtain from choosing φ_i or ψ_i depends not only on the strategies of the other players but also their identities (i.e. indices or positions), the fated truel cannot be considered a symmetric game in the classical sense, even though the resulting strategy dynamics are symmetrical. Also, notice that the fated truel game exhibits 2×2 indifference dynamics anytime $b = c$ (all values of structural fear and greed are zero) resulting in $\mathcal{S}^\star = \mathcal{S}$. When $a > b > c$, this game exhibits pure 2×2 coexistence and $\mathcal{S}^\star = \mathcal{S} \setminus \{\varphi, \psi\}$.

El Farol bar

Three groups of friends decide independently whether to have fun at the small and only bar in Santa Fé (ψ), unaware of how crowded it is. Neither group wants to be at a bar packed beyond

its limits; if they find the bar to be more than 2/3 full, they will regret not having stayed (at someone's) home (φ). So, although they would love to meet another group there, they prefer not all groups to show up, which requires anti-coordination. For any $a > b \geq 0$, player i's payoff in this game is defined as

$$U_i(s_i, s_j, s_k) = \begin{cases} 0 & \text{if } \langle s_i, s_j, s_k \rangle = \psi \\ a & \text{if } s_i = \psi_i \wedge s_{-i} = \varphi_{-i} \\ b & \text{if } s_i = \psi_i \wedge s_j \neq s_k \\ c & \text{otherwise.} \end{cases} \tag{A.12}$$

For $n = 3$, this game has 3 PNE each of them equal to $s^\star = \langle \varphi_i, \psi_{-i} \rangle$ for every $i \in \mathcal{N}$. In Fig A.5, $a = 100\%$ fun, $b = 25\%$ fun, and $c = 10\%$ fun. The individual strategic hindrance is:

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{100} \begin{bmatrix} -90 & -15 & -90 & -15 & -90 \\ -15 & +10 & -15 & +10 & +10 \end{bmatrix}^{\mathrm{T}}.$$

Simple majority game

A player only benefits when they are one of the $\lceil n/2 \rceil$ or more members of $\mathcal{N}$ who choose the same strategy (φ or ψ). Those who are part of a minority receive nothing. When n is even, a stalemate could result in penalties for all players. They payoff function is given by

$$U_i(s_i, s_j, s_k) = \begin{cases} 1/n_\varphi & \text{if } s_i = \varphi_i \wedge n_\varphi > n/2 \\ 1/n_\psi & \text{if } s_i = \psi_i \wedge n_\psi > n/2 \\ -c & \text{if } n_\varphi = n_\psi = n/2 \\ 0 & \text{otherwise,} \end{cases} \tag{A.13}$$

where $c \geq 0$ is the associated penalty for failing to form a simple majority. The set of PNE in this game is $\mathcal{S}^\star = \{\varphi, \psi\}$. For $n = 3$ and $c = 0$, the individual strategic hindrance is:

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{1/2} \begin{bmatrix} +\frac{1}{3} & 0 & +\frac{1}{3} & 0 & +\frac{1}{3} \\ 0 & -\frac{1}{3} & 0 & -\frac{1}{3} & -\frac{1}{3} \end{bmatrix}^{\mathrm{T}}.$$

Matching pennies with more than 2 players

Matching pennies are zero-sum variants of simple majority games. In these games, the sum of the benefits received by the players who form a majority must equal the sum of the absolute losses that minority players incur. The payoff function in a symmetric version of these games is

$$U_i(s_i, s_j, s_k) = \begin{cases} 0 & \text{if } n_\varphi \in \{0, n/2, n\} \\ -n_\psi/n_\varphi & \text{if } s_i = \varphi_i \wedge n_\varphi \in [1 \,..\, n/2) \\ -n_\varphi/n_\psi & \text{if } s_i = \psi_i \wedge n_\psi \in [1 \,..\, n/2) \\ 1 & \text{otherwise,} \end{cases} \tag{A.14}$$

The individual payoff matrix for a three-player version of this game is provided in Table A.7. The corresponding strategic hindrance space is

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{3} \begin{bmatrix} +2 & 0 & +2 & 0 & +2 \\ 0 & -2 & 0 & -2 & -2 \end{bmatrix}^{\mathrm{T}},$$

same as that of the three-player simple majority game in Fig A.5 using Eq. (A.13) with $c = 0$. While $\mathcal{S}^\star = \{\varphi, \psi\}$ for $n > 2$, the classical 2-player game version of this game in Eq. (12), characterized by $\operatorname{sgn}(U_i(s_i, s_j)) = -\operatorname{sgn}(U_j(s_j, s_i)) \neq 0$, has no PNE.

$U_i(s_i, s_j, s_k)$	$s_k = \varphi_k$		$s_k = \psi_k$	
	$s_j = \varphi_j$	$s_j = \psi_j$	$s_j = \varphi_j$	$s_j = \psi_j$
$s_i = \varphi_i$	0	1	1	−2
$s_i = \psi_i$	−2	1	1	0
$U_i(\varphi_i, s_j, s_k) - U_i(\psi_i, s_j, s_k)$	2	0	0	−2

Fig A.7. Individual payoff matrix in a three-player matching pennies game, a zero-sum variant of a majority game. The peak-to-peak payoff amplitude is $A_i = 3$. The assigning of φ or ψ as "heads" or "tails" can be done arbitrarily. The resulting structural fear and greed value space is the same as the one for the majority game in Fig A.5.

Unanimity game

All players in $\mathcal{N}$ must choose the same strategy (φ or ψ) to reap a benefit $b > 0$:

$$U_i(s_i, s_{-i}) = \begin{cases} b & \text{if } s_1 = \cdots = s_{i-1} = s_i = s_{i+1} = \cdots = s_n \\ 0 & \text{otherwise.} \end{cases}$$

For $\mathcal{N} = \{i, j, k\}$, the deviation losses in this game are: $\ell_i^*(\varphi_{-i}) = -\ell_i^*(\psi_{-i}) = b/A_i = 1$ and $\ell_i^*(\varphi_j, \psi_k) = \ell_i^*(\psi_j, \varphi_k) = 0$. The individual strategic hindrance is:

$$\mathbf{H}_i^{(3)}(\varphi, \psi) = \frac{1}{b} \begin{bmatrix} +b & 0 & +b & 0 & +b \\ 0 & -b & 0 & -b & -b \end{bmatrix}^{\mathrm{T}}.$$

All type I player-reduced binary games in a $n \times 2$ unanimity game are 2×2 bistability games. Under these dynamics, stable rational collective action requires that every player belongs to a (sub-) coalition with at least one other player willing to choose the same strategy. With $n = 3$, only two of such coalitions can be formed: $\langle \varphi_1, \varphi_2, \varphi_3 \rangle$ and $\langle \psi_1, \psi_2, \psi_3 \rangle$. With $n = 4$, there are eight ways in which rational players would align their strategies: two instances in which all players choose the same strategy (i.e. φ and ψ) plus six instances in which two of them play φ_i and the other two play ψ_i. We can obtain the number of PNE as $|\mathcal{S}^\star| = \max\{2, 2^n - 2n\}$ for $n \geq 2$, which adds the two collective diagonal strategies, φ and ψ, and every possible binary sequence of length $n = n_\varphi + n_\psi$ where $n_\varphi > 1$ and $n_\psi > 1$ (sequence `A052515` in Sloane's *Encyclopedia* [51]).

A $n \times 2$ game with no PNE

After characterizing the relationship between deviation losses, the emergence of strategy dynamics, and equilibria conditions in several classical $n \times 2$ social dilemmas, we conclude with a formula that recursively constructs a binary game with no PNE for $n > 2$. Building on the bimatrix game in Eq. (11) and Definition A.2a, the following formula allocates a balanced ternary payoff vector $U \in \{-1, 0, +1\}^n$ to every $s \in \mathcal{S}$ such that $\min_{i \in \mathcal{N}} \ell_i(s_i, s_{-i}) < 0$; by assigning a strict cyclic order to $\mathcal{N}$, e.g. $[1, 2, \ldots, n-1, n] \Rightarrow [i, i+1, \ldots, i-2, i-1]$, the perfectly limited individual payoff function

$$U_i(s_i, s_{\mathcal{J}}, s_{\mathcal{K}}) = \begin{cases} 0 & \text{if } s_i = \varphi_i \\ +1 & \text{if } s_{\mathcal{J}} = \langle \varphi_{i+1}, \varphi_{i-1} \rangle \vee s_{\mathcal{J}} = \langle \psi_{i+1}, \varphi_{i+2} \rangle \\ -1 & \text{otherwise,} \end{cases} \tag{A.15}$$

where $i \in [1 \,..\, n]$, guarantees that there is at least one player in $\mathcal{N}$ and no more than $n-1$ that would be better off deviating away from any $s \in \mathcal{S}$. For instance, if $n = 3$, the rescaled deviation

losses for player $i = 1$ are

$$\begin{aligned}
\ell_i^*(\varphi_{i+1},\varphi_{i-1}) &= A_i^{-1}\cdot\Big(U_i(\varphi_i,\varphi_{i+1},\varphi_{i-1}) - U_i(\psi_i,\varphi_{i+1},\varphi_{i-1})\Big) = \big(\ \ 0-1\big)/A_i = -b/A_i,\\
\ell_i^*(\varphi_{i+1},\psi_{i-1}) &= A_i^{-1}\cdot\Big(U_i(\varphi_i,\varphi_{i+1},\psi_{i-1}) - U_i(\psi_i,\varphi_{i+1},\psi_{i-1})\Big) = \big(\ \ 0+1\big)/A_i = +b/A_i,\\
\ell_i^*(\psi_{i+1},\varphi_{i-1}) &= A_i^{-1}\cdot\Big(U_i(\varphi_i,\psi_{i+1},\varphi_{i-1}) - U_i(\psi_i,\psi_{i+1},\varphi_{i-1})\Big) = \big(\ \ 0-1\big)/A_i = -b/A_i,\\
\ell_i^*(\psi_{i+1},\psi_{i-1}) &= A_i^{-1}\cdot\Big(U_i(\varphi_i,\psi_{i+1},\psi_{i-1}) - U_i(\psi_i,\psi_{i+1},\psi_{i-1})\Big) = \big(\ \ 0+1\big)/A_i = +b/A_i;
\end{aligned}$$

where $A_i = 2$. The individual strategic hindrance is:

$$\begin{aligned}
\mathbf{H}_i^{(3)}(\varphi,\psi) &= \frac{1}{A_i}\begin{bmatrix} \ell_i^*(\varphi_{i+1},\varphi_{i-1}) & \ell_i^*(\varphi_{i+1},\psi_{i-1}) & \ell_i^*(\varphi_{i+1},\varphi_{i-1}) & \ell_i^*(\psi_{i+1},\varphi_{i-1}) & \ell_i^*(\varphi_{i+1},\varphi_{i-1}) \\ \ell_i^*(\psi_{i+1},\varphi_{i-1}) & \ell_i^*(\psi_{i+1},\psi_{i-1}) & \ell_i^*(\varphi_{i+1},\psi_{i-1}) & \ell_i^*(\psi_{i+1},\psi_{i-1}) & \ell_i^*(\psi_{i+1},\psi_{i-1}) \end{bmatrix}^{\mathrm{T}} \\
&= \frac{1}{2}\begin{bmatrix} -1 & +1 & -1 & -1 & -1 \\ -1 & +1 & +1 & +1 & +1 \end{bmatrix}^{\mathrm{T}}.
\end{aligned}$$

References

1. Haugen RA, Skeie NO, Muller G, Syverud E. Detecting Emergence in Engineered Systems: A Literature Review and Synthesis Approach. Systems Eng. 2023;26(4):463–481. doi:10.1002/sys.21660.

2. Lluch I, Golkar A. Architecting Federations of Systems: A Framework for Capturing Synergy. Systems Eng. 2019;22(4):295–312. doi:10.1002/sys.21482.

3. de Weck OL, Roos D, Magee CL. Engineering Systems: Meeting Human Needs in a Complex Technological World. Cambridge, MA, United States: MIT Press; 2011.

4. Robinett RD, Wilson DG. Collective Systems: Physical and Information Exergies. SAND2007-2327, Sandia National Laboratories; 2007. doi:10.2172/909392.

5. Chen W, Ahmed F, Cui Y, Sha Z, Contractor N. Data-Driven Preference Modelling in Engineering Systems Design. In: Maier A, Oehmen J, Vermaas PE, editors. Handbook of Engineering Systems Design. Springer; 2022. p. 407–440. doi:10.1007/978-3-030-46054-9_15-2.

6. Heydari B, Herder P. Technical and Social Complexity. In: Maier A, Oehmen J, Vermaas PE, editors. Handbook of Engineering Systems Design. Springer; 2022. p. 221–250. doi:10.1007/978-3-030-46054-9_9-2.

7. Maier MW. Architecting Principles for Systems-of-systems. Systems Eng. 1998;1(4):267–284. doi:10.1002/(SICI)1520-6858(1998)1:4<267::AID-SYS3>3.0.CO;2-D.

8. Vincent TL. Game Theory as a Design Tool. ASME J Mech Des. 1983;105(2):165–170. doi:10.1115/1.3258503.

9. Lewis K, Mistree F. Modeling Interactions in Multidisciplinary Design: A Game Theoretic Approach. AIAA J. 1997;35(8):1387–1392. doi:10.2514/2.248.

10. Takai S. A Game-Theoretic Model of Collaboration in Engineering Design. ASME J Mech Des. 2010;132(5):051005. doi:10.1115/1.4001205.

11. Heydari B, Dalili K. Emergence of Modularity in System of Systems: Complex Networks in Heterogeneous Environments. IEEE Syst J. 2013;9(1):223–231. doi:10.1109/JSYST.2013.2281694.

12. Bayrak AE, Sha Z. Integrating Sequence Learning and Game Theory to Predict Design Decisions Under Competition. ASME J Mech Des. 2021;143(5):051401. doi:10.1115/1.4048222.

13. Avşar AZ, Grogan PT. Effects of Differential Risk Attitudes in Collaborative Systems Design. Systems Eng. 2023;26(6):770–782. doi:10.1002/sys.21687.

14. Salado A. Interpreting Integrability. Systems Eng. 2023;(in press):1–10. doi:10.1002/sys.21709.

15. Sugden R. A Theory of Focal Points. Econ J. 1995; p. 533–550. doi:10.2307/2235016.

16. de Heus P, Hoogervorst N, Van Dijk E. Framing Prisoners and Chickens: Valence Effects in the Prisoner's Dilemma and the Chicken Game. J Exp Soc Psychol. 2010;46(5):736–742. doi:10.1016/j.jesp.2010.04.013.

17. Brams SJ. Theory of Moves. Am Sci. 1993;81(6):562–570.

18. Sen A. The Contributions of John Nash to Game Theory and Economics. Econ Political Wkly. 2015;p. 75–80.

19. Durlauf S, Blume L. Behavioural and Experimental Economics. London, United Kindom: Palgrave Macmillan London; 2010. doi:https://doi.org/10.1057/9780230280786.

20. Macy MW, Flache A. Learning Dynamics in Social Dilemmas. Proc Natl Acad Sci USA. 2002;99(suppl 3):7229–7236. doi:10.1073/pnas.092080099.

21. Rapoport A. Strategy and Conscience. New York, NY, United States: Harper and Row; 1964.

22. Oye KA. Explaining Cooperation under Anarchy: Hypotheses and Strategies. World Polit. 1985;38(1):1–24. doi:10.2307/2010349.

23. Bruns BR. Names for Games: Locating 2×2 Games. Games. 2015;6(4):495–520. doi:10.3390/g6040495.

24. Ahn TK, Ostrom E, Schmidt D, Shupp R, Walker J. Cooperation in PD Games: Fear, Greed, and History of Play. Public Choice. 2001;106(1-2):137–155. doi:10.1023/A:1005219123532.

25. Kitts JA, Leal DF, Felps W, Jones TM, Berman SL. Greed and Fear in Network Reciprocity: Implications for Cooperation among Organizations. PLOS ONE. 2016;11(2):e0147264. doi:10.1371/journal.pone.0147264.

26. Nowak MA, May RM. Evolutionary Games and Spatial Chaos. Nature. 1992;359(6398):826–829. doi:10.1038/359826a0.

27. Hauert C. Effects of Space in 2×2 Games. Int J Bifurcat Chaos. 2002;12(07):1531–1548. doi:10.1142/S0218127402005273.

28. Nowak MA, Sigmund K. Evolutionary Dynamics of Biological Games. Science. 2004;303(5659):793–799. doi:10.1126/science.1093411.

29. Hummert S, Bohl K, Basanta D, Deutsch A, Werner S, Theißen G, et al. Evolutionary Game Theory: Cells as Players. Mol Biosyst. 2014;10(12):3044–3065. doi:10.1039/C3MB70602H.

30. Traulsen A, Santos FC, Pacheco JM. Evolutionary Games in Self-Organizing Populations. In: Gross T, Sayama H, editors. Adaptive Networks: Theory, Models and Applications. Springer; 2009. p. 253–267. doi:10.1007/978-3-642-01284-6_12.

31. van Segbroeck S, Santos FC, Pacheco JM, Lenaerts T. Coevolution of Cooperation, Response to Adverse Social Ties and Network Structure. Games. 2010;1(3):317–337. doi:10.3390/g1030317.

32. Brown SP, Taddei F. The Durability of Public Goods Changes the Dynamics and Nature of Social Dilemmas. PLOS ONE. 2007;2(7):e593. doi:10.1371/journal.pone.0000593.

33. Werner B. Mathematical Models of Cell Population Dynamics. Ph.D. thesis. University of Lübeck. Lübeck, Germany; 2013.

34. Selten R. An Axiomatic Theory of a Risk Dominance Measure for Bipolar Games with Linear Incentives. Games Econ Behav. 1995;8(1):213–263. doi:10.1016/S0899-8256(05)80021-1.

35. Grogan PT, Ho K, Golkar A, de Weck OL. Multi-actor Value Modeling for Federated Systems. IEEE Syst J. 2018;12(2):1193–1202. doi:10.1109/JSYST.2016.2626981.

36. Young HP. The Evolution of Conventions. Econometrica. 1993; p. 57–84. doi:10.2307/2951778.

37. Grogan PT, Bayrak A. Operational and Strategical Decisions in Engineering Design Games. In: *Proceedings of the ASME 2018 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, Québec City, QC, Canada. American Society of Mechanical Engineers; 2018. p. V02BT03A024. doi:10.1115/DETC2018-85317.

38. Campos-Mercade P. The Volunteer's Dilemma Explains the Bystander Effect. J Econ Behav Organ. 2021;186:646–661. doi:10.1016/S0167-7187(99)00037-5.

39. Frahm G. Rational Choice and Strategic Conflict: The Subjectivistic Approach to Game and Decision Theory. Berlin, Germany: Walter de Gruyter GmbH & Co KG; 2019. doi:10.1515/9783110596106.

40. Perea A. Epistemic Game Theory: Reasoning and Choice. Cambridge, United Kindom: Cambridge University Press; 2012. doi:10.1017/CBO9780511844072.

41. Newton J. Agency Equilibrium. Games. 2019;10(1):14. doi:10.3390/g10010014.

42. Wiszniewska-Matyszkiel A. Belief-distorted Nash Equilibria: Introduction of a New Kind of Equilibrium in Dynamic Games with Distorted Information. Annals of Operations Research. 2016;243(1):147–177. doi:10.1080/15710882.2013.824482.

43. Güth W, Kalkofen B. Unique Solutions for Strategic Games: Equilibrium Selection Based on Resistance Avoidance. In: Beckman MJ, Krelle W, editors. Lecture Notes in Economics and Mathematical Systems. vol. 328. Berlin, Germany: Springer-Verlag; 1989. p. 1–200. doi:10.1007/978-3-642-48367-7.

44. Kalai E, Lehrer E. Subjective Games and Equilibria. Games Econ Behav. 1995;8(1):123–163. doi:10.1016/S0899-8256(05)80019-3.

45. Battigalli P, Guaitoli D. Conjectural Equilibria and Rationalizability in a Game with Incomplete Information. In: Battigalli P, Montesano A, Panunzi F, editors. Decisions, Games and Markets. Springer; 1997. p. 97–124. doi:10.1007/978-1-4615-6337-2_4.

46. Harsanyi JC. Games with Incomplete Information Played by "Bayesian" Players, I–III Part I. The Basic Model. Manag Sci. 1967;14(3):159–182. doi:10.1287/mnsc.14.3.159.

47. Güth W. On the Inconsistency of Equilibrium Refinement. Theory Decis. 2002;53(4):371–392. doi:10.1023/A:1024194725086.

48. Peleg B, Tijs S. The Consistency Principle for Games in Strategic Form. Int J Game Theory. 1996;25(1):13–34. doi:10.1007/BF01254381.

49. Norde H, Potters J, Reijnierse H, Vermeulen D. Equilibrium Selection and Consistency. Games Econ Behav. 1996;12(2):219–225. doi:10.1006/game.1996.0014.

50. Valencia-Romero A. Strategy Dynamics in Collective Systems Design. Ph.D. thesis. Stevens Institute of Technology. Hoboken, NJ, United States; 2021.

51. OEIS Foundation Inc. The On-Line Encyclopedia of Integer Sequences; 2024. Available from: https://oeis.org/.

52. La Haye R. Binary Relations on the Ppower Set of an n-Element Set. J Integer Seq. 2009;12(09.2.6).

53. Glance NS, Huberman BA. The Dynamics of Social Dilemmas. Sci Am. 1994;270(3):76–81.

54. Weil RL. The N-person Prisoner's Dilemma: Some Theory and a Computer-oriented Approach. Behav Sci. 1966;11(3):227–234. doi:10.1002/bs.3830110310.

55. Poundstone W. Prisoner's Dilemma. New York, NY, United States: Anchor Books; 1992.

56. Goehring DJ, Kahan JP. The Uniform n-Person Prisoner's Dilemma Game: Construction and Test of an Index of Cooperation. J Confl Resolut. 1976;20(1):111–128. doi:10.1177/002200277602000104.

57. Diekmann A. Volunteer's Dilemma. J Confl Resolut. 1985;29(4):605–610. doi:10.1177/0022002785029004003.