Check for
updates

# Emergent Collaboration in Social Purpose Games

**Robert P. Gilles[1] · Lina Mallozzi[2]** (ID) **· Roberta Messalli[3]**

## Abstract

We study a class of non-cooperative aggregative games—referred to as *social purpose games*—in which the payoffs depend separately on a player's own strategy (individual benefits) and on a function of the strategy profile which is common to all players (social benefits) weighted by an individual benefit parameter. This structure allows for an asymmetric assessment of a common social benefit across players. We show that these games have a weighted potential, and we investigate its properties. We investigate the payoff structure and the uniqueness of Nash equilibria and social optima. Furthermore, following the literature on partial cooperation, we investigate the leadership of a single coalition of cooperators, while the rest of players act as non-cooperative followers. In particular, we show that social purpose games admit the emergence of a stable coalition of cooperators for the subclass of *strict* social purpose games. As a particular application, we study a standard formulation of the *tragedy of the commons*. We show that there emerges a single stable coalition of cooperators that curbs the over-exploitation of the common resource.

**Keywords** Partial cooperation · Leadership equilibrium · Potential games · Aggregative games · Tragedy of the commons

✉ Lina Mallozzi
  mallozzi@unina.it

  Robert P. Gilles
  r.gilles@qub.ac.uk

  Roberta Messalli
  robertamessalli@gmail.com

[1] Management School, The Queen's University of Belfast, Riddel Hall, 185 Stranmillis Road, Belfast BT9 5EE, UK

[2] Department of Mathematics and Applications, University of Naples Federico II, Via Claudio 21, 80125 Naples, Italy

[3] Department of Economic and Statistical Science, University of Naples Federico II, Complesso Monte Sant'Angelo 21, 80125 Naples, Italy

**JEL Classification** C71 · D71

## 1 Introduction: Endogenous Partial Cooperation in Games

In this paper, following the literature of additively separable aggregative games [8, 13] and in line with the asymmetry considered in [29], we introduce a class of non-cooperative games—referred to as *Social Purpose Games*—for which the payoff of each player depends on his own strategy and on an aggregation of the complete strategy profile, which is the same for all players. A player's payoff is now an affine combination of these two parts, respectively, an individual cost function and a common social benefits component.

The purpose of this paper is to investigate the properties of and relationship between Nash equilibria and social optima in these social purpose games. Furthermore, we show that for a subclass of these social purpose games, there emerges a stable coalition of cooperators founded on the presence of a limited form of farsightedness in the corresponding definition of stability.

The main feature of social purpose games is that they capture the tension between social benefits and individual costs. As such this class of social purpose games includes implementations of the tragedy of the commons, pollution abatement games, and public good provision games. In most of these social purpose games, individual optimality leads players to under-utilise social benefits in favour of direct individualistic payoffs. We show that, indeed, this is a general feature of social purpose games, exemplified by the social suboptimality of the Nash equilibria. If we interpret the strategies in these social purpose games to represent a chosen level of "effort", the total Nash equilibrium generated level of effort is socially suboptimal. We argue that partial collaboration through the emergence of a stable coalition of cooperators, partially abates the suboptimality of the resulting effort levels. Therefore, this fundamental insight holds for a substantial class of aggregative games.

For our analysis, we introduce two subclasses of social purpose games. In a *regular* social purpose game, all payoff functions are continuous and (weakly) concave. Regularity of a social purpose game guarantees the existence of the main equilibrium concepts. A social purpose game is *strict* if the payoff structures are strictly convex, guaranteeing uniqueness of the main equilibrium concepts.

Jensen [19] already pointed out that classes of aggregative games can be characterised through certain forms of potential functions. We show here that social purpose games admit a weighted potential function. This is a substantially stronger property than Jensen showed for more general classes of aggregative games. Furthermore, the existence of a weighted potential leads to the question regarding the relationship between the maximisers of the potential function and the Nash equilibria of these games. We show that for a special class of social purpose games, the set of Nash equilibria actually coincides with the set of potential maximisers.

Concerning the Nash equilibria of social purpose games, we show a wide range of properties. Under standard continuity properties, any social purpose game admits at least one Nash equilibrium, including the class of regular social purpose games. For the class of *strict* social purpose games, the Nash equilibrium is unique due to the strict convexity of the corresponding payoff structures.

Similarly, under these conditions, social purpose games generally admit at least one social optimum, while for the subclass of strict social purpose games, the social optimum is unique. The fundamental tension between social and individual benefit leads to the natural conclusion

that, as stated above, for a large subclass of social purpose games, the Nash equilibrium levels of effort are socially suboptimal. The next goal of our paper is to investigate how partial cooperation can alleviate tension between socially optimal and equilibrium levels of effort.

*Partial cooperation in social purpose games* Social purpose games have a specific structure that allows the emergence of stable partial cooperation among players. Our investigation is rooted in the work on partial cooperation in a range of types of non-cooperative games by Chander and Tulkens [7], Mallozzi and Tijs [26–28] and Chakrabarti et al. [4, 5]. In this literature, one considers the formation of a single "coalition of cooperators" that collectively determines a joint strategy to maximise its collective payoffs—being the sum of the individual members' payoffs. This is akin to cartel formation in oligopolistic market games [10] and international treaty writing in the context of environmental abatement situations [12, 20, 31]. There are two natural behavioural hypotheses that can be considered in this context.

First, one may assume that the coalition of cooperators acts as a single player under standard best response rationality. The resulting stable strategy profiles are referred to as *partial cooperative equilibria*, which existence can be established under relatively mild conditions [4, 5]. This avenue is not investigated in our analysis, since in many applications, the coalition of cooperators would not have a competitive position in the corresponding decision-making processes.

Instead, in most natural applications, the coalition of cooperators is found to have a position of *leadership* in a decision-making hierarchy in the sense of von Stackelberg [38]. The hypothesis that the coalition of cooperators assumes a leadership position as a first mover in relation to the non-cooperating players in the game was seminally proposed by Chander and Tulkens [7]. This conception has been developed further as the notion of a *partial cooperative leadership equilibrium* (PCLE) in the cited literature on partial cooperation in general non-cooperative games. It is clear that the leadership position of the coalition of cooperators gives it an advantage in comparison with the standard partial cooperative equilibrium payoff. Nevertheless, this leadership position seems natural and has been observed in the context of social purpose games, in particular for public good provision games, pollution abatement games, and (oligopolistic) market games.

In our study of partial cooperation in social purpose games, we limit ourselves to the investigation of PCLE under the formation of a coalition of cooperators. In particular, we can show that under relatively mild conditions, there exists a unique PCLE in a social purpose game. This conception, therefore, allows us to investigate how a coalition of cooperators forms. In particular, we presented the emergence of a stable coalition of cooperators using the von Neumann–Morgenstern ("vNM") standards [37], which considers two partial stability properties: (1) *Internal stability*—every cooperator will obtain a lower payoff upon leaving the coalition of cooperators, and (2) *External stability*—every non-cooperator receives a lower payoff upon joining the existing coalition of cooperators.

In the class of strict social purpose games, we formulate vNM stability based on the decisions by individual marginal players to join or abandon a proposed coalition of cooperators. This could also be captured in terms of an accompanying abstract meta-game, which equilibrium states represent forms of farsightedness in decision-making. We refer here to the equilibrium concepts investigated by [23] and [22] that relate closely to the vNM stability notion investigated here for strict social purpose games.

We show for a subclass of the strict social purpose games there indeed emerge stable coalitions of cooperators that abate the social suboptimality of the corresponding equilibrium effort levels.

*Relationship to the literature* An aggregative game is founded on the hypothesis that each payoff function depends on the corresponding player's strategy as well as on some

aggregation of all selected strategies. Classical examples of aggregation are the unweighted sum and the mean. The concept of aggregative games goes back to [33], who considers as aggregation function the summation of all the players' strategies. Later, this concept has been studied in the case of other aggregation functions, and it has been generalised to the concept of quasi-aggregative games. For this we refer to [1, 8, 9, 11, 13, 19, 24, 36]. In this literature, there are many games that present an aggregative structure: among them, we mention Cournot and Bertrand games, patent races, models of contests of fighting and model with aggregate demand externalities.

In additively separable aggregative games, each payoff function is a sum of a function that depends on an aggregation of strategies and a function that depends on player's own strategy. The model of additively separable aggregative games appeared in the literature, among others, in the context of International Environmental Agreements (IEA), studying the formation of stable IEA in the case in which each country's choice variable is emission and then extending the results to the dual case, i.e. the case where the choice variable is abatement effort [12].

Also public good provision games are in the context of additively separable aggregative games [3] where each player consumes a certain amount of a private good and donates a certain other amount to the supply of the public good. Thus, the payoff function of each player turns to depend not only on the quantity of private good that he consumes, but also on all contributions to the public good made by all individuals. [29] investigate the impact of asymmetry in a voluntary public goods environment by proposing an improved design that explicitly isolates individual incentives, without assuming a dominant strategy.

*Structure of the paper* In Sect. 2 of this paper, we investigate thoroughly the quintessential social purpose game, namely a standard implementation of the tragedy of the commons [17]. We show that the effects of the over-exploitation of the commons can be mitigated by partial cooperation among the players and that such cooperation can be stable. Stable partial cooperation is shown to mitigate the over-exploitation of the commons.

In Sect. 3, we introduce the class of general social purpose games and identify some relevant sub-classes. We show that social purpose games in general admit a weighted potential and that for a sub-class of these games the potential maximisers coincide with the set of Nash equilibria. We discuss social optima and their properties, showing in particular that for the class of regular social purpose games, there is the expected relationship between Nash equilibrium and socially optimal strategies, referring to the over-exploitation of the commons.

Section 4 introduces the notion of partial cooperation in social purpose games. We define the notion of a partial cooperative leadership equilibrium and prove its existence for regular social purpose games. Subsequently, we introduce a stability concept in the formation of partial cooperation and show that it is natural to expect that strict social purpose games admit the emergence of stable partial cooperation. This is fully developed for an application to quadratic payoff functions.

## 2 A Motivating Case: The Tragedy of the Commons

The "tragedy of the commons" refers to the classical problem of the overuse and exploitation of a common resource through free-riding in a non-cooperative setting. Traditionally, the common resource referred to a common tract of land in a medieval village for the grazing of

cattle owned by the village's peasants. If peasants freely access and use the land, a situation of over-exploitation arises, resulting in the depletion of the commons for collective use [21].[1]

The tragedy of the commons was formulated in game-theoretic terms by Hardin [17] and has been considered a totemic reference in many contributions to the social and biological sciences [14]. Hardin [17] only considered the non-cooperative case of unlimited and free extraction from the commons. Ostrom [32] challenged Hardin's reductionism by pointing out that in many social situations the commons was and still is successfully governed through the application of institutional and behavioural solutions. Gilles et al. [15] introduce a mathematical model of an institutional solution on managing the commons from a public good perspective based on Ostrom's insights.

Here, we pursue a third perspective on the management of the commons by investigating the endogenous emergence of a coalition of cooperators that collective regulate their extraction from the commons, while non-members of this coalition of cooperators selfishly extract. We can show that the extraction from the commons under such *partial cooperation* significantly improves collective wealth generation and welfare. We illustrate the benefits from endogenous partial cooperation in the classical tragedy of the commons game by considering a simple example.

*A non-cooperative extraction game* Suppose that there is a finite commonly owned resource that has a total size of one (1). There are $n \in \mathbb{N}$ with $n \geqslant 3$ users of this resource, who are individually free to extract any benefit from the common resource with the understanding that any future benefit from the resource would be limited by the extent of today's usage. Hence, future use is based on the remainder of the common resource at the conclusion of today's collective extraction.

This results in a standard non-cooperative game $\langle N, (S_i)_{i \in N}, (\pi_i)_{i \in N} \rangle$ where $N = \{1, 2, \ldots, n\}$ is the set of users and $S_i = [0, 1]$ is the set of individual extraction levels. We apply a standard payoff function that considers benefits from current usage and the utility of future usage with equal weight for every $i \in N$. Hence,

$$\pi_i(x) = x_i \left( 1 - \sum_{j \in N} x_j \right) = x_i (1 - X_N) \quad \text{for every } x \in S = \prod_{j \in N} S_j = [0, 1]^n \subset \mathbb{R}^n.$$

We denote by $X_T = \sum_{i \in T} x_i$ the total extraction of a coalition $T \subseteq N$ with the convention that $X_\varnothing = 0$. Note that $(1 - X_N)$ therefore represents the total size of the commons left for future usage.[2]

*Nash equilibria and social optima* We summarise the resulting Nash equilibrium[3] and the unique social (Pareto) optimum[4] outcomes in the following table:

---

[1] The tragedy of the commons has significant appeal and application in our contemporary global economy, referring to contemporary issues such as the exploitation of natural resources—including fish stocks, fresh water sources, mining of ores, and oligopolistic commodity markets—as well as the global environmental conditions. The tragedy of the commons is known as one of the most fundamental examples of a *social dilemma*.

[2] We remark that $1 - X_N < 0$ refers to the destruction of the commons, resulting in negative payoffs for all users.

[3] A Nash equilibrium is a strategy profile $x^{\text{NE}} \in S$ in which all users do a best response to what other users extract from the commons, i.e. $\pi_i(x^{\text{NE}}) = \max_{x_i \in S_i} \pi_i(x_i, x_{-i}^{\text{NE}})$.

[4] A social optimum is a strategy profile $x^{\text{SO}} \in S$ that maximises the collective welfare, i.e. $\sum_{i \in N} \pi_i(x^{\text{SO}}) = \max_{x \in S} \sum_{i \in N} \pi_i(x)$.

| | $x_i$ | $X_N$ | $\pi_i$ | $\sum \pi_i$ |
|---|---|---|---|---|
| Nash equilibrium | $\frac{1}{n+1}$ | $\frac{n}{n+1}$ | $\frac{1}{(n+1)^2}$ | $\frac{n}{(n+1)^2}$ |
| Social optimum | $\frac{1}{2n}$ | $\frac{1}{2}$ | $\frac{1}{4n}$ | $\frac{1}{4}$ |

We observe here the well-known conclusion that there significant over-extraction in the Nash equilibrium, which reduces welfare uniformly for all users. Indeed, we note that $X_N^{NE} > X_N^{SO}$ as well as $\sum \pi_i^{NE} < \sum \pi_i^{SO}$ for all $n \geq 3$.

*Partial cooperation in the tragedy of the commons* Next, we consider a hybrid of best response rationality and socially optimal decision-making by allowing users to collaborate to extract in a collectively rational fashion. In particular, consider that a coalition of cooperators $C \subset N$ decides collectively about their extraction rate, while the non-collaborating users $j \in N \backslash C$ act according to best response rationality. Furthermore, we assume that the coalition of cooperators $C$ assumes a Stackelberg leadership position and acts as a first-mover. Hence, users in $C$ determine their coordinated extraction rates taking into account the best responses of all non-cooperators $j \in N \backslash C$. This is referred to as a *partial cooperative leadership equilibrium* (PCLE) in [4, 5].

Suppose that the coalition of cooperators has size $|C| = m < n$. In a PCLE, non-cooperators $j \in N \backslash C$ optimise their payoffs given $X_C$ as well as $(x_k)_{k \in N \backslash C, k \neq j}$. This results for the non-cooperator $j \in N \backslash C$ in solving

$$\max_{0 \leq x_j \leq 1} \pi_j \left( x_j, x^{NC}, x^C \right) = x_j (1 - X_C - X_{NC})$$

where $X_C = \sum_{i \in C} x_i$ and $X_{NC} = \sum_{h \in N \backslash C} x_h$. Solving this simultaneously for all $j \in N \backslash C$, given $X_C$, we conclude that the best response for $j \in N \backslash C$ is given by

$$x_j(X_C) = \frac{1 - X_C}{n - m + 1} \quad \text{and} \quad \sum_{j \in N \backslash C} x_j(X_C) = \frac{(n - m)(1 - X_C)}{n - m + 1} \tag{1}$$

Next, the coalition of cooperators $C$ collectively determine their collective extraction, given the optimal decisions of all non-cooperators, by solving

$$\max_{x^C \in [0,1]^C} \sum_{i \in C} \pi_i \left( x^C, \left( x_j(X_C) \right)_{j \in N \backslash C} \right) = X_C \left( 1 - X_C - \sum_{j \in N \backslash C} x_j(X_C) \right)$$

This results in the conclusion that the coalition of cooperators $C$ solves

$$\max_{x^c \in [0,1]^C} X_C \cdot \frac{1 - X_C}{n - m + 1} \tag{2}$$

leading to the conclusion that the optimal collective extraction is $X_C^L = \frac{1}{2}$. The resulting PCLE can be summarised as

$$x_i^L = x_C^L(m) = \frac{1}{2m} \qquad \text{for } i \in C \tag{3}$$

$$x_j^L = x_{NC}^L(m) = \frac{1}{2(n - m + 1)} \qquad \text{for } j \in N \backslash C \tag{4}$$

resulting in $X_C^L(m) = \frac{1}{2}$, $X_{NC}^L(m) = \frac{n-m}{2(n-m+1)}$ and $X_N^L(m) = 1 - \frac{1}{2(n-m+1)} = \frac{2(n-m)+1}{2(n-m+1)}$. Clearly, the resulting PCLE outcomes form a hybrid between Nash equilibrium and social

optimum extraction. The resulting PCLE payoffs reflect this as

$$\pi_i^L = \pi_C^L(m) = \frac{1}{4m(n-m+1)} \qquad \text{for } i \in C \qquad (5)$$

$$\pi_j^L = \pi_{NC}^L(m) = \frac{1}{4(n-m+1)^2} \qquad \text{for } j \in N \setminus C \qquad (6)$$

*Stable partial cooperation* The tragedy of the commons as formulated here allows for the consideration of stability in the process of partial cooperation under leadership. We call a coalition of cooperators $C$ *stable* if no non-cooperator $j \in N \setminus C$ would like to join the coalition $C$ and no member $i \in C$ would like to leave the coalition of cooperators.[5] Formally, we say that a coalition of cooperators $C \subseteq N$ is **stable** if

$$\pi_{NC}^L(m) \geqslant \pi_C^L(m+1) \qquad \text{as well as} \qquad \pi_C^L(m) \geqslant \pi_{NC}^L(m-1) \qquad (7)$$

It can be verified that there are many $(n, m)$ pairs that satisfy these inequalities. In particular, this holds for $n = 8$ and $m = 5$. We verify this for these particular sizes of the coalition of cooperators for population size $n = 8$. The next table summarises the main resulting PCLE strategies and extraction rates for all sizes $2 \leqslant m < n = 8$ of the coalition of cooperators.

| $m$ | $x_C^L$ | $x_{NC}^L$ | $X_N$ | $\pi_C^L$ | $\pi_{NC}^L$ |
|---|---|---|---|---|---|
| 2 | $\frac{1}{4}$ | $\frac{1}{14}$ | $\frac{13}{14}$ | $\frac{1}{56}$ | $\frac{1}{196}$ |
| 3 | $\frac{1}{6}$ | $\frac{1}{12}$ | $\frac{11}{12}$ | $\frac{1}{72}$ | $\frac{1}{144}$ |
| 4 | $\frac{1}{8}$ | $\frac{1}{10}$ | $\frac{9}{10}$ | $\frac{1}{80}$ | $\frac{1}{100}$ |
| 5* | $\frac{1}{10}$ | $\frac{1}{8}$ | $\frac{7}{8}$ | $\frac{1}{80}$ | $\frac{1}{64}$ |
| 6 | $\frac{1}{12}$ | $\frac{1}{6}$ | $\frac{5}{6}$ | $\frac{1}{72}$ | $\frac{1}{36}$ |
| 7 | $\frac{1}{14}$ | $\frac{1}{4}$ | $\frac{3}{4}$ | $\frac{1}{56}$ | $\frac{1}{16}$ |

From the table of the computed PCLE, we conclude that indeed $m^* = 5$ is a stable size of the coalition. Indeed, $\pi_C^L(5) = \frac{1}{80} > \pi_{NC}^L(4) = \frac{1}{100}$ as well as $\pi_{NC}^L(5) = \frac{1}{64} > \pi_C^L(6) = \frac{1}{72}$.

We can summarise the properties of stable partial cooperation in the tragedy of the commons as follows:

**Proposition 2.1** *A coalition of cooperators $C^*$ is stable in the tragedy of the commons if and only if its size $m^* = |C^*|$ satisfies*

$$\frac{m^*+1}{n-m^*+1} \geqslant \frac{n-m^*+1}{n-m^*} \qquad \text{as well as} \qquad \frac{n-m^*+2}{n-m^*+1} \geqslant \frac{m^*}{n-m^*+2}.$$

*Moreover, it holds that $X_N^{SO} < X_N^L(m^*) < X_N^{NE}$.*

**Proof** The first assertion can easily be derived through some simple calculus. Regarding the second assertion, we note that $\frac{m+1}{n-m+1} \geqslant \frac{n-m+1}{n-m}$ is equivalent to

$$\left[ n - \left( \tfrac{1}{2}m - \tfrac{1}{2} \right) \right]^2 \leqslant \tfrac{1}{4}m^2 - \tfrac{1}{2}m - \tfrac{3}{4}$$

Hence, a stable coalition of cooperators satisfies

$$\left[ n - \left( \tfrac{1}{2}m - \tfrac{1}{2} \right) \right]^2 < \tfrac{1}{4}m^2 - \tfrac{1}{2}m + \tfrac{1}{4} = \left( \tfrac{1}{2}m - \tfrac{1}{2} \right)^2, \ \text{or } n < 2m - 1.$$

---

[5] This introduces a form of vNM stability [37] in this simple player-based formation process of the coalition of cooperators.

This, in turn, implies that $X_N^L(m) = \frac{2(n-m)+1}{2(n-m+1)} < \frac{n}{n+1} = X_N^{NE}$, which shows the assertion. □

The analysis of the PCLE in the setting of the tragedy of the commons shows that there emerges self-regulation of the extraction from the commons through partial cooperation under a standard of rational behaviour considered by Kimya [23] through his notion of an "equilibrium coalitional behaviour". In this paper, we explore a more general class of non-cooperative games that exhibits this feature. This is developed over the following sections.

## 3 Social Purpose Games

A *normal form game* is a list $G = \langle N, (S_i)_{i=1}^n, (\pi_i)_{i=1}^n \rangle$ such that

- $N = \{1, \ldots, n\}$ is a given finite set of players, where $n \in \mathbb{N}$ is the number of players;
- for every $i \in N$, $S_i$ is a non-empty *strategy set* for player $i$, and;
- for every $i \in N$, $\pi_i \colon S \to \mathbb{R}$ is the *payoff function* of player $i$ with $S = \prod_{i \in N} S_i$ being the set of all *strategy tuples* in $G$.

A *Nash equilibrium* in the game $G = \langle N, (S_i)_{i=1}^n, (\pi_i)_{i=1}^n \rangle$ is a strategy tuple $x^* \in S$ such that for every $i \in N$: $\pi_i(x^*) \geqslant \pi_i\left(x_{-i}^*, x_i\right)$ for every alternative strategy $x_i \in S_i$, where the tuple of strategies of all players except $i$ is denoted by $x_{-i} = (x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n) \in \prod_{j \neq i} S_j$.

A *social optimum* in a game $G$ is a strategy profile $x^{SO} \in \arg\max_{x \in S} \sum_{i \in N} \pi_i(x)$, defined as a collective, utilitarian wealth maximiser.

We now introduce a social purpose game as a non-cooperative normal form game with a specific payoff structure.

**Definition 3.1** A **social purpose game** is a list $\Gamma = \langle N, \overline{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ that defines an associated non-cooperative game $G_\Gamma = \langle N, (S_i)_{i=1}^n, (\pi_i)_{i=1}^n \rangle$ such that

- $N = \{1, \ldots, n\}$ is a finite player set, where the number of players is $n \in \mathbb{N}$ such that $n \geqslant 2$;
- $\overline{Q} > 0$ is a positive number such that every $i \in N$: $S_i = [0, \overline{Q}]$, endowed with the standard Euclidean topology;
- $H \colon \mathbb{R} \to \mathbb{R}$ is a function; and
- For every $i \in N$: $\alpha_i > 0$ is a parametric weight and $h_i, g_i \colon [0, \overline{Q}] \to \mathbb{R}$ are functions such that for every $x = (x_1, \ldots, x_n) \in S = [0, \overline{Q}]^N$ player $i$'s payoffs are given by

$$\pi_i(x) = \alpha_i H \left( \sum_{j=1}^n h_j(x_j) \right) - g_i(x_i) \tag{8}$$

Throughout we use $\Gamma$ synonymously with $G_\Gamma$.

All payoffs in a social purpose game depend separately on player $i$'s own strategy, through the term $g_i(x_i)$ that represents an individual cost of player $i$ of executing strategy $x_i$, and on a common term $H\left(\sum_{i=1}^n h_i(x_i)\right)$ that represents a common social benefit for all players. Therefore, the function $H$ is naturally interpreted as a *social benefit function*, while for every player $i \in N$, the function $g_i$ represents an *individual cost function*.

The weight $\alpha_i > 0$ measures a *propensity to prefer the common benefit*. For ease of the following analysis, we assume that all players in $N$ are ranked according to their preference for the generated social benefit, i.e. such that $0 < \alpha_1 \leqslant \cdots \leqslant \alpha_n$.

The functions $(h_i)_{i \in N}$ can naturally be interpreted as assessments of individual contributions in the aggregated contribution, represented as $\sum_{i \in N} h_i$. This formulation allows for a wide range of different representations of aggregation through selections of the functions $H$ and $(h_i)_{i \in N}$. We refer to the functions $H$ and $(h_i, g_i)_{i \in N}$ as the *constituting functions* of the social purpose game $\Gamma$.

A social purpose game $\Gamma$ is clearly an aggregative game—making the insights from, e.g. [1, 19] applicable. The class of social purpose games includes the classes of various well-known aggregative games such as *pollution abatement games* and, more generally, *public good provision games*, which are both characterised by $\alpha_i = 1$ for all $i \in N$.

We note that the tragedy of the commons situation considered in Sect. 2 is indeed a social purpose game. This is made explicit by noting that the payoff function of the tragedy of the commons can be written in the required form through a monotonic transformation by taking the (natural) logarithm of her payoff function. In particular, for $i \in N$, $\log \pi_i(x) = \log \left( 1 - \sum_{i \in N} x_i \right) - (- \log x_i)$, corresponding to $\alpha_i = 1$, $h_i(x_i) = x_i$, $g_i(x_i) = - \log x_i$ and $H(x) = \log(1 - x)$.

*Classes of social purpose games*

Thus far we have not imposed any properties on the constituting functions $H$ and $(h_i, g_i)_{i \in N}$. It is natural to require these functions to satisfy certain properties. In particular, we consider the case in which each $h_i$ is the identity function. We combine this with convexity properties on the constituting functions. This results into a class of social purpose games that includes the tragedy of the commons and public good provision games.

**Definition 3.2** Let $\Gamma = \langle N, \overline{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ be a social purpose game, generating the normal form representation $G_\Gamma$.

(i) The social purpose game $\Gamma$ is called **regular** if

- all functions $h_i$, $i \in N$, are the identity functions $h_i(x_i) = x_i$ and all functions $H$ and $(g_i)_{i \in N}$ are continuously differentiable;
- the common benefit function $H$ is continuously differentiable, increasing and concave—implying that its derivative $H'$ is weakly decreasing on $X$, and;
- the individual cost function $g_i$ is continuously differentiable, increasing and convex in $x_i$ for every $i \in N$—implying that its derivative $g_i'$ is weakly increasing for every $i \in N$.

(ii) The social purpose game $\Gamma$ is called **strict** if $\Gamma$ is regular and, additionally, for every $i \in N$ the individual cost function $g_i$ is twice differentiable and strictly convex—implying that $g_i'' > 0$.

Regularity of social purpose games imposes the concavity of the payoff functions. This implies that regular social purpose games describe a situation in which players make a contribution to a common goal or resource and that these contributions are subject to weakly decreasing returns to scale, individually as well as socially.

Strict social purpose games additionally impose that the payoff structure is strictly concave. This implies that one can think of these games as extractive situations from a common resource that is subject to strictly decreasing marginal returns. One can also refer to these strict social purpose games as "extraction games" to emphasise the most common and obvious application of this class of games.

## 3.1 Weighted Potentials and Nash Equilibria

In the following proposition, we show that each social purpose game is a weighted potential game with weights $\alpha_i$ for all players $i \in N$ [30]. This result is stronger than the insight established in [19] for the larger class of aggregative games concerning the weaker notion of *best response potentials*.

**Proposition 3.3** *Every social purpose game* $\Gamma = \langle N, \overline{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ *is a weighted potential game for weight vector* $\alpha = (\alpha_1, \ldots, \alpha_n)$. *The corresponding* $\alpha$-*potential is given by*

$$P(x) = H\left(\sum_{i=1}^{n} h_i(x_i)\right) - \sum_{i=1}^{n} \frac{1}{\alpha_i} g_i(x_i), \tag{9}$$

*satisfying that* $\pi_i(y_i, x_{-i}) - \pi_i(z_i, x_{-i}) = \alpha_i \left(P(y_i, x_{-i}) - P(z_i, x_{-i})\right)$ *for all* $i \in N$.

The proof of the proposition is a straightforward computation to check the asserted property.

We say that social purpose game $\Gamma$ *admits* Nash equilibria if the associated normal form game $G_\Gamma$ has Nash equilibria. Using established insights from potential game theory, we can prove quite straightforwardly the existence of Nash equilibrium for the class of social purpose games.

**Proposition 3.4** *Consider a social purpose game* $\Gamma$.

(a) *If all constituting functions* $H$ *and* $(h_i, g_i)_{i \in N}$ *are continuous, then* $\Gamma$ *admits at least one Nash equilibrium.*

(b) *If the social purpose game* $\Gamma$ *is strict, then* $\Gamma$ *admits a unique Nash equilibrium which is identical to the unique potential maximiser.*

(c) *If all constituting functions* $H$ *and* $(h_i, g_i)_{i \in N}$ *are continuously differentiable, then every interior Nash equilibrium* $x^{\text{NE}} = (x_1^{\text{NE}}, \ldots, x_n^{\text{NE}}) \in \left(0, \overline{Q}\right)^N$ *is a solution to the set of equations given by*

$$H'\left(\sum_{j=1}^{n} h_j(x_j)\right) h_i'(x_i) = \frac{g_i'(x_i)}{\alpha_i} \quad \text{for every } i \in N. \tag{10}$$

(d) *If the common benefit function* $H$ *is linear, i.e.* $H(x) = Ax + B$ *for some* $A, B \in \mathbb{R}$, *then the set of Nash equilibria coincides with the set of potential maximisers, i.e.* $P^{\max} = \text{NE}(\Gamma)$.

**Proof** The hypotheses that $H$ is continuous and $h_i$ and $g_i$ are continuous for every $i \in N$, implying that the potential function $H$ is continuous on $[0, \overline{Q}]^N$. By compactness of $[0, \overline{Q}]^N$, there exists a maximum of $P$ and, thus, since for a weighted potential game $\arg\max P \subseteq \text{NE}_\Gamma$, where $\text{NE}_\Gamma$ is the set of Nash equilibria of $G_\Gamma$ [30], there exists at least one Nash equilibrium in $G_\Gamma$.

To show assertion (b), for any $s = \sum_{i=1}^{n} x_i$ let $x_i(s) \geqslant 0$ be the solution of $\alpha_i H'(s) = g_i'(x_i(s))$. Since $g_i'$ is strictly monotone, we find $x_i(s) = (g_i')^{-1}(\alpha_i H'(s))$. Since $\Gamma$ is a strict social purpose game and $(g_i')^{-1}$ is increasing, the map $s \to x_i(s)$ is continuous and decreasing on $[0, n\overline{Q}]$.

We define the map $F : s \in [0, n\overline{Q}] \mapsto \sum_{i=1}^{n} x_i(s) \in [0, n\overline{Q}]$, which is continuous and decreasing on $[0, n\overline{Q}]$, since $\sum_{i=1}^{n} x_i(0) > 0$. Furthermore, under our assumptions

$H'(0) > 0$, implying that $F$ has a unique fixed point $s^* \in [0, n\overline{Q}]$, i.e. $s^* = \sum_{i=1}^{n} x_i(s^*)$, corresponding to the unique Nash equilibrium.

To show assertion (c), consider an interior Nash equilibrium $x^{\text{NE}} = (x_1^{\text{NE}}, \ldots, x_n^{\text{NE}}) \in (0, \overline{Q})^N$. Then clearly, $x^{\text{NE}}$ satisfies the first-order conditions of the Nash best response optimisation problem given by

$$x_i^{\text{NE}} \in \arg\max_{0 < x_i < \overline{Q}} \pi_i \left( x_i, x_{-i}^{\text{NE}} \right).$$

Noting that

$$\frac{\partial \pi_i}{\partial x_i} \pi_i(x) = \alpha_i H' \left( \sum_{j=1}^{n} h_j(x_j) \right) h_i'(x_i) - g_i'(x_i),$$

we arrive immediately at the conclusion that assertion (c) is valid.

To show assertion (d), we note that if $H$ is linear, the corresponding social purpose game $\Gamma$ is a separable game [2]. Hence, using Mallozzi (2013, Proposition 6), $\Gamma$ admits an exact potential given by $Q(x) = A \sum_{i \in N} \alpha_i h_i(x_i) - \sum_{i \in N} g_i(x_i)$. Now, arg max $Q$ = arg max $P$ and, thus, the assertion follows immediately from Proposition 8 of [25]. $\qquad \square$

**Remark 3.5** In the case in which the constituting functions are continuously differentiable, any interior Nash equilibrium is a solution of the system of first-order conditions given by $\nabla P(x_1, \ldots, x_n) = 0$. This implies that every interior Nash equilibrium is a stationary point of the potential $P$. This allows the relatively easy computation of the Nash equilibria of a social purpose game.

Note here that given equations in Proposition 3.4(c) lead to the conclusion that for all $i, j \in N$:

$$\frac{g_i'(x_i^{\text{NE}})}{\alpha_i h_i'(x_i^{\text{NE}})} = \frac{g_j'(x_j^{\text{NE}})}{\alpha_j h_j'(x_j^{\text{NE}})}$$

Hence, each of these fractions represents a *shadow price* of the individual marginal benefit in terms of the common marginal benefit.

If $h_i$ is the identity function for all $i \in N$, this shadow price is simplified to $\rho_i = \frac{g_i'(x_i^{\text{NE}})}{\alpha_i}$. These shadow prices can be used to interpret the tensions between individual costs and benefits in social purpose game situations.

## 3.2 Social Optima

Recall that a strategy profile $x^{\text{SO}} \in [0, \overline{Q}]^N$ is a social optimum of the social purpose game $\Gamma$ if it maximises the collective payoff $\sum_{i \in N} \pi_i(x)$. The next proposition addresses the existence of social optima in social purpose games. For the following analysis, it is useful to introduce the *social welfare function* $W: [0, \overline{Q}]^N \to \mathbb{R}$ where

$$W(x) = \frac{1}{A} \sum_{i=1}^{n} \pi_i(x) = H \left( \sum_{i=1}^{n} h_i(x_i) \right) - \frac{\sum_{i=1}^{n} g_i(x_i)}{A} \tag{11}$$

with $A = \sum_{i \in N} \alpha_i$. Clearly, social optima maximise the function $W$ over $[0, \overline{Q}]^N$. This is used in the proof of the next proposition.

**Proposition 3.6** *Consider a social purpose game $\Gamma$.*

(a) *If the common benefit function H and all individual functions $(h_i, g_i)_{i \in N}$ are continuous, then there exists a social optimum $x^{SO} = (x_1^{SO}, \ldots, x_n^{SO}) \in [0, \overline{Q}]^N$.*

(b) *If $\Gamma$ is strict, then there exists a unique social optimum $x^{SO} \in [0, \overline{Q}]^N$ for $\Gamma$.*

**Proof** Under the stated assumptions, the social welfare function $W$ is continuous on $[0, \overline{Q}]^N$. By the Weierstrass theorem, $W$ has a maximum on $[0, \overline{Q}]^N$, which corresponds to a social optimum.

To show assertion (b), note that the constructed social welfare function $W$ is strictly concave on $[0, \overline{Q}]^N$. Hence, $W$ has a unique maximiser, corresponding to the unique social optimum for $\Gamma$.                                                                                    □

From Proposition 3.6(a), it follows immediately that every regular social purpose game admits a Nash equilibrium as well as a social optimum.

**Remark 3.7** We note that if all constituting functions are continuously differentiable, the social optimum solves $\nabla W(x_1, \ldots, x_n) = 0$ implying that for every $i \in N$:

$$H'\left(\sum_{j=1}^{n} h_j(x_j)\right) h_i'(x_i) = \frac{g_i'(x_i)}{A}$$

This compares to the shadow prices derived in Remark 3.5, indicating the well-established underinvestment in the common benefit in social purpose games.

Therefore, the social optimum $x_i^{SO}$ does not depend individually on $\alpha_i$ for every $i \in N$.

*Comparing Nash equilibria and social optima in social purpose games*

For the class of social purpose games, the comparison of the Nash equilibria and social optima is usually very informative. The next example considers a very simple payoff structure in which the difference between these two concepts is maximal on the strategy set $[0, \overline{Q}]$.

**Example 3.8** Consider a regular social purpose game $\Gamma = \langle N, \overline{Q}, H, (\alpha_i, h_i, g_i)_{i \in N} \rangle$ with the following payoff structure:

- $H\left(\sum_{i \in N} h_i(x_i)\right) = \sum_{i \in N} x_i$ is a standard utilitarian common benefit function;
- $g_i(x_i) = x_i$ for every $i \in N$, and
- $0 < \alpha_i < 1$ for every $i \in N$ such that $A = \sum_{i \in N} \alpha_i > 1$.

For any $i \in N$: $\pi_i(x) = \alpha_i \sum_{j \neq i} x_j + (\alpha_i - 1)x_i$ and, thus, the unique Nash equilibrium is given by $x_i^{NE} = 0$ for all $i \in N$.

On the other hand, the social optimum is identified by maximising the social welfare function

$$\sum_{i \in N} \pi_i(x) = A \sum_{i \in N} x_i - \sum_{i \in N} x_i = (A - 1) \sum_{i \in N} x_i$$

which has a unique maximum identified as being given by $x_i^{SO} = \overline{Q}$ for every $i \in N$.

In regular social purpose games, the collective contributions in Nash equilibrium are always lower than the collective contributions required for a social optimum. This refers to the commonly accepted property that is exhibited in public good provision situations.

**Proposition 3.9** *Suppose that $\Gamma$ is a regular social purpose game. Then for every Nash equilibrium $x^{\mathrm{NE}}$ and every social optimum $x^{\mathrm{SO}}$ of $G_\Gamma$, it holds that*

$$\sum_{i=1}^{n} x_i^{\mathrm{NE}} \leqslant \sum_{i=1}^{n} x_i^{\mathrm{SO}}. \tag{12}$$

**Proof** Let us suppose by contradiction that $\sum_{i=1}^{n} x_i^{\mathrm{NE}} > \sum_{i=1}^{n} x_i^{\mathrm{SO}}$. Since $H'$ is decreasing in $\sum_{i=1}^{n} x_i$, then $H'\left(\sum_{i=1}^{n} x_i^{\mathrm{NE}}\right) \leqslant H'\left(\sum_{i=1}^{n} x_i^{\mathrm{SO}}\right)$. From Remark 3.5, for any $i \in N$: $\frac{g_i'(x_i^{\mathrm{NE}})}{\alpha_i} \leqslant \frac{g_i'(x_i^{\mathrm{SO}})}{A}$, where $A = \sum_{j=1}^{n} \alpha_j > 0$. Since, for any $i \in N$, $0 < \alpha_i < A$ and $g_i'$ is positive, it follows that

$$\frac{g_i'(x_i^{\mathrm{NE}})}{A} < \frac{g_i'(x_i^{\mathrm{NE}})}{\alpha_i} \leqslant \frac{g_i'(x_i^{\mathrm{SO}})}{A}. \tag{13}$$

Since $g_i'$ is increasing in $x_i$, we obtain that $x_i^{\mathrm{NE}} \leqslant x_i^{\mathrm{SO}}$. Hence, we have shown that $x_i^{\mathrm{NE}} \leqslant x_i^{\mathrm{SO}}$ for any $i \in N$. This, in turn, implies that $\sum_{i=1}^{n} x_i^{\mathrm{NE}} \leqslant \sum_{i=1}^{n} x_i^{\mathrm{SO}}$, contradicting the hypothesis that $\sum_{i=1}^{n} x_i^{\mathrm{NE}} > \sum_{i=1}^{n} x_i^{\mathrm{SO}}$. This contradiction proves the assertion. $\square$

**Remark 3.10** We can easily generalise the previous result to the case in which functions $h_i$ for all $i \in N$ are not the identity functions, instead assuming that $\frac{g_i'(\cdot)}{h_i'(\cdot)}$ are increasing functions in $x_i$ for all players $i \in N$.

This assumption simply means that an increase in player $i$'s strategy has a greater impact on the individual cost $g_i(x)$ than on the common benefit $H\left(\sum_i h_i(x_i)\right)$ through which player $i$'s contribution is assessed.

We also note that if we assume that the common benefit function $H$ is convex, we lose the property stated in Proposition 3.9.

**Example 3.11** We consider the non-regular social purpose game $\Gamma$ defined by $n = 2$, $\overline{Q} = 1$ and payoff functions $\pi_1$ and $\pi_2$ given by $\alpha_1 = \alpha_2 = \frac{1}{2}$, $H(t) = 2t^2 - 4t$ and $g_1 = g_2 = 0$; then,

$$\pi_1(x_1, x_2) = \pi_2(x_1, x_2) = (x_1 + x_2)^2 - 2(x_1 + x_2). \tag{14}$$

We identify two Nash equilibria $\{(0, 0), (1, 1)\}$ that are also social optima. So, the statement in Proposition 3.9 is no longer true.

We next compare each component of a Nash equilibrium with the corresponding component of a social optimum. First, we consider an example of a strict social purpose game in which there are players $i \in N$ with $x_i^{\mathrm{NE}} < x_i^{\mathrm{SO}}$ as well as players $j \in N$ with $x_j^{\mathrm{NE}} > x_j^{\mathrm{SO}}$.

**Example 3.12** Let $\delta > 0$. Consider a strict social purpose game $\Gamma_\delta$ with $N = \{1, 2\}$ and $\overline{Q} = 1$ with

$$\pi_1(x_1, x_2) = \alpha_1 \log(x_1 + x_2 + \delta) - x_1^{1+\delta} \tag{15}$$

$$\pi_2(x_1, x_2) = \alpha_2 \log(x_1 + x_2 + \delta) - x_2^{1+\delta} \tag{16}$$

where $0 < \alpha_1 < 1 < \alpha_2$ with $A = \alpha_1 + \alpha_2 < 2$ and assuming that $\delta > 0$ is a sufficiently small parameter, implying that the game $\Gamma_\delta$ is indeed a strict social purpose game.

We investigate Nash equilibria as well as social optima of this class of strict social purpose games $\Gamma_\delta$ for $\delta > 0$ that are sufficiently small.

Nash equilibria: The game $\Gamma_\delta$ has a unique Nash equilibrium that can be determined by investigating the first-order conditions for an interior solution. We derive that for $i = 1, 2$:

$$\frac{\partial \pi_i}{\partial x_i} = 0 \quad \text{implies} \quad \frac{\alpha_i}{x_1 + x_2 + \delta} = (1 + \delta)x_i^\delta$$

For the case that $\delta$ is sufficiently small, $\alpha_1 < 1$ and $\alpha_2 > 1$, it follows that there is no interior Nash equilibrium and that the only Nash equilibrium is the corner equilibrium given by $x_1^{\text{NE}} = 0$ and $x_2^{\text{NE}} = 1$.

Social optimum: The social optimum can be determined by optimising the standard social welfare function given by

$$\pi_1(x_1, x_2) + \pi_2(x_1, x_2) = A \log(x_1 + x_2 + \delta) - x_1^{1+\delta} - x_2^{1+\delta}$$

The first-order conditions for this optimum can be derived and lead to the conclusion that

$$0 < x_1^{\text{SO}} = x_2^{\text{SO}} < \left( \frac{A}{2(1 + \delta)} \right)^{\frac{1}{1+\delta}} < 1$$

due to the property that $A < 2$ and that $\delta > 0$ is sufficiently small.

Hence, in this game, we have derived that $x_1^{\text{NE}} = 0 < x_1^{\text{SO}}$ as well as $x_2^{\text{NE}} = 1 > x_2^{\text{SO}}$, while $x_1^{\text{NE}} + x_2^{\text{NE}} = 1 < x_1^{\text{SO}} + x_2^{\text{SO}}$.

It may also happen that $x_i^{\text{NE}} < x_i^{\text{SO}}$ for any $i \in N$, as is shown in the application developed in Sect. 4.2.

## 4 Endogenous Emergence of Collaboration

Following the literature on partial cooperation in non-cooperative games developed in [4–6, 26–28], we consider the emergence of stable partial cooperative in social purpose games. We explore particularly the notion of a partial cooperative leadership equilibrium (PCLE) for this class of games prior to developing a stability notion and showing that there indeed might emerge stable partnerships in social purpose games under the PCLE concept. Through the PCLE notion and the definition of stable coalition formation, this incorporates a reduced form of vNM-farsightedness in the formation of a stable coalition of cooperators Kimya [23].

### 4.1 Partial Cooperative Leadership Equilibrium

We first consider two main notions of partial cooperative equilibrium, following the theory developed in [4, 27] as well as some of the insights obtained in [5]. The next hypothesis limits the scope of the coalitions of cooperators that are assumed to form, even though more generalised formulations are relatively easy to explore.

**Axiom 4.1** *Within the context of a social purpose game $\Gamma$, we have assumed throughout that players are ranked by their propensity to prefer the common benefit, i.e. $0 < \alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_n$.*

*Furthermore, we assume that only coalitions of players with the highest propensity to prefer the common benefits will cooperate. Hence, for a **level of cooperation** $k \in \{2, \ldots, n\}$,*

*we denote by* $C_k = \{n - k + 1, \ldots, n\}$ *the corresponding coalition of cooperators and by* $N_k = N \setminus C_k = \{1, \ldots, n - k\}$ *the complement of* $C_k$, *being the corresponding set of non-cooperators.*[6]

Axiom 4.1 implies that the higher ranked players with a higher propensity to prefer the common benefit component of their payoffs are assumed to identify the benefits for cooperation.

For level of cooperation $k \in \{2, \ldots, n\}$, we now denote by $x^{C_k} = (x_{n-k+1}, \ldots, x_n) \in \prod_{i=n-k+1}^{n} S_i = [0, \overline{Q}]^{C_k}$ and $x^{N_k} = (x_1, \ldots, x_{n-k}) \in \prod_{j=1}^{n-k} S_j = [0, \overline{Q}]^{N_k}$. In particular, $x^{C_k}$ is collectively selected by the coalition $C_k$, while $x^{N_k}$ is competitively selected by individual players $j \in N_k$.

*Equilibrium concepts under partial cooperation*

In this setting, there naturally result two types of equilibrium concepts to be considered for any given level of cooperation $k \in \{2, \ldots, n\}$. The first one is the *partial cooperative equilibrium* concept—based on the same logic as temporal "best response" rationality on which the standard Nash equilibrium concept is founded—in which every non-cooperator $i \in N_k = \{1, \ldots, n - k\}$ acts competitively or individually, while all cooperating players $j \in C_k = \{n - k + 1, \ldots, n\}$ act cooperatively as a single decision maker. Hence, in this partial cooperative equilibrium concept all decision makers in the set $\{1, \ldots, n - k, C_k\}$ act as Nash optimisers. We refer to [4, 5] for a full analysis of partial cooperative equilibria in general non-cooperative games. Here, we limit our discussion to the second notion of equilibrium under partial cooperation.

Alternatively, for any given level of cooperation $k \in \{2, \ldots, n\}$, we might consider the notion of *partial cooperative leadership equilibrium* (PCLE). In such an equilibrium, it is assumed that the coalition of cooperators $C_k$ acts as a single Stackelberg leader [38] and all non-cooperators $j \in N_k$ act as Stackelberg followers. In order to explicitly introduce this leader–follower equilibrium concept, for any $x^{C_k} \in [0, \overline{Q}]^{C_k}$, we denote by $\Gamma_k(x^{C_k}) = \langle N_k, [0, \overline{Q}]^{N_k}, \omega^{x^{C_k}} \rangle$ the normal form game—called the *conditional partial cooperative game* for $x^{C_k}$—given by player set $N_k = N \setminus C_k$ of non-cooperators whose strategy set is $S_j = [0, \overline{Q}]$ and who have a conditional payoff function $\omega_j^{x^{C_k}} : [0, \overline{Q}]^{N_k} \to \mathbb{R}$ defined by

$$\omega_j^{x^{C_k}} \left( x^{N_k} \right) = \pi_j \left( x^{N_k}, x^{C_k} \right) \qquad j \in N_k = N \setminus C_k. \tag{17}$$

We denote by $\mathrm{NE}(x^{C_k}) \subset [0, \overline{Q}]^{N_k}$ the set of Nash equilibria that emerge in this conditional non-cooperative game $\Gamma_k(x^{C_k})$.

Note that under the assumptions of Proposition 3.4, $\mathrm{NE}(x^{C_k}) \neq \varnothing$ but we cannot guarantee the uniqueness of the Nash equilibrium in $\Gamma_k(x^{C_k})$. Thus, in line with [4, 27], in order to select one Nash equilibrium among followers, we assume that cooperators are pessimistic in the sense that the coalition of the cooperators, the leader, supposes that the followers' (non-cooperators) choice is the worst for herself and select a maximin strategy. As a consequence, we introduce

$$\tilde{\pi} \left( x^{C_k} \right) = \min_{x^{N_k} \in \mathrm{NE}(x^{C_k})} \sum_{i \in C_k} \pi_i \left( x^{N_k}, x^{C_k} \right) \tag{18}$$

---

[6] We remark here that $k = 0$ or $k = 1$ are meaningless in the context of partial cooperation. Furthermore, the level of cooperation $k = n$ refers to a case of social optimality, since all players in the game act as cooperators. The latter is actually a potential valid level of cooperation.

and

$$\widetilde{X}_k = \left\{ \tilde{x}^{C_k} \in [0, \overline{Q}]^{C_k} \,\middle|\, \tilde{\pi}(\tilde{x}^{C_k}) = \max_{x^C \in [0,\overline{Q}]^C} \tilde{\pi}(x^C) \right\} \tag{19}$$

An action tuple $(x_*^{N_k}, x_*^{C_k}) \in [0, \overline{Q}]^{N_k} \times [0, \overline{Q}]^{C_k}$ is a *partial cooperative leadership equilibrium* for the game $\Gamma$ if $x_*^{C_k} \in \widetilde{X}_k$ and

$$x_*^{N_k} \in \arg\min_{x^{N_k} \in \mathrm{NE}(x_*^{C_k})} \sum_{i \in C_k} \pi_i \left( x^{N_k}, x_*^{C_k} \right).$$

The following result states the conditions under which a PCLE exists in a social purpose game.

**Proposition 4.2** *For every regular social purpose game $\Gamma$ and every level of cooperation $k \in \{2, \ldots, n\}$, there exists at least one partial cooperative leadership equilibrium.*

**Proof** Assume that $\Gamma$ is a regular social purpose game, i.e. $H$ is concave and increasing, $g_i$ are convex and continuous for all $i \in N$ and $h_i(x_i) = x_i$ for all $i \in N$. Take any level of cooperation $k \in \{2, \ldots, n\}$ with the corresponding coalition of cooperators $C_k = \{n - k + 1, \ldots, n\}$. We first show the following claim: □

**Claim A** *The correspondence $\mathbf{E} \colon [0, \overline{Q}]^{C_k} \to 2^{[0,\overline{Q}]^{N_k}}$ defined by*

$$x^{C_k} \in [0, \overline{Q}]^{C_k} \mapsto \mathbf{E}(x^{C_k}) = \mathrm{NE}(x^{C_k}) \subset [0, \overline{Q}]^{N_k} \tag{20}$$

*is non-empty, compact valued and upper hemi-continuous.*

**Proof of Claim A:** Since Proposition 3.4 holds, for any $x^{C_k} \in [0, \overline{Q}]^{C_k}$, $\mathrm{NE}(x^{C_k}) \neq \varnothing$, implying that $\mathbf{E}$ is non-empty valued.

Let $\left( x_p^{C_k} \right)_{p \in \mathbb{N}}$ a sequence such that $x_p^{C_k} \to x^{C_k} \in [0, \overline{Q}]^{C_k}$. Let us take $x_p^{N \backslash C} \in \mathrm{NE}(x_p^{C_k})$ such that $x_p^{N_k} \to x^{N_k}$ as $p \to \infty$. Since $x_p^{N_k} \in \mathrm{NE}(x_p^{C_k})$, we have

$$\alpha_j H \left( \sum_{j=1}^{n-k} x_{j,p}^{N_k} + \sum_{i=n-k+1}^{n} x_{i,p}^{C_k} \right) - g_j \left( x_{j,p}^{N_k} \right) \geqslant$$
$$\alpha_j H \left( x_j^{'N_k} + \sum_{l=1,l\neq j}^{n-k} x_{l,p}^{N_k} + \sum_{i=n-k+1}^{n} x_{i,p}^{C_k} \right) - g_j \left( x_j^{'N_k} \right)$$

for every non-cooperator $j \in N_k$ and for any $x_j^{'N_k} \in [0, \overline{Q}]$.

If $p \to \infty$ by continuity of all functions, we obtain

$$\alpha_j H \left( \sum_{j=1}^{n-k} x_j^{N_k} + \sum_{i=n-k+1}^{n} x_i^{C_k} \right) - g_j \left( x_j^{N_k} \right) \geqslant$$
$$\alpha_j H \left( x_j^{'N_k} + \sum_{l=1,l\neq j}^{n-k} x_l^{N_k} + \sum_{i=n-k+1}^{n} x_i^{C_k} \right) - g_j \left( x_j^{'N_k} \right)$$

for every non-cooperator $j \in N_k$ and for any $x_j^{'N_k} \in [0, \overline{Q}]$. Thus, $x^{N_k} \in \mathrm{NE}(x^{C_k})$ leading to the conclusion that $\mathbf{E}$ is closed valued.

Finally, the set $\text{NE}(x^{C_k})$ is closed and compact for any $x^{C_k} \in [0, \overline{Q}]^{C_k}$, since $\text{NE}(x^{C_k}) \subset [0, \overline{Q}]^{N_k}$ is closed as shown above and $[0, \overline{Q}]^{N_k}$ is compact. Finally, since every closed correspondence with compact co-domain is upper hemi-continuous, this implies the assertion of the claim.

From Claim A and the fact that each payoff function $\pi_i$ for $i \in C_k$ and $\pi_j$ for $j \in N_k$ is continuous and quasi-concave, with reference to the proof of [4, Theorem 2.6], there exists at least one PCLE in $\Gamma$. This shows the assertion of Proposition 4.2.

We remark that the regularity condition for the existence of PCLE in social purpose games as stated in Proposition 4.2 is not necessary as a simple example makes clear.

**Example 4.3** Consider a social purpose game with three players $n = 3$, $\overline{Q} = 1$, a convex common benefit function given by $H(t) = t^2$, and for every $i \in \{1, 2, 3\}$ a payoff structure determined by $\alpha_i = i$, $h_i(x) = x$, $g_i(x) = 0$. Hence, for every $i \in \{1, 2, 3\}$: $\pi_i(x_1, x_2, x_3) = i(x_1 + x_2 + x_3)^2$.

The game is not regular. However, it is easy to see that $(1, 1, 1)$ is a PCLE with players 2 and 3 forming the coalition of cooperators, assuming a leadership position.

The next example shows explicitly that the mapping $\mathbf{E}$ as constructed in the proof of Proposition 4.2 does not have to be single-valued.

**Example 4.4** Consider the regular social purpose game $\Gamma$ with $n = 2$, $\overline{Q} = 1$ and payoff functions $\pi_1$ and $\pi_2$ given by choosing $\alpha_1 = \alpha_2 = 1$, $H(t) = t$, $g_1 = x_1$, $g_2 = 0$. Then, $\pi_1(x_1, x_2) = x_2$ and $\pi_2(x_1, x_2) = x_1 + x_2$.

By considering the player 2 as the only cooperator, for any given $x_2$, the other player maximised $\pi_1(x_1, x_2) = x_2$ with respect to $x_1$. The multi-valued map $\mathbf{E}$ of Claim A assigns to every $x_2$ the whole strategy set $[0, 1]$. The non-cooperative player 1 determines her optimal strategy $\bar{x}_2 = 1$ by solving the problem

$$\max_{x_2} \left[ \min_{x_1 \in [0,1]} x_1 + x_2 \right]$$

Hence, the map $\mathbf{E}$ is indeed multi-valued.

**Remark 4.5** Denoting by $x_*^L = \left( x_*^{N_k}, x_*^{C_k} \right)$ a partial cooperative leadership equilibrium at level of cooperation $k \in \{2, \ldots, n\}$ and given $C_k \subset N$ as the corresponding coalition of cooperators, under the assumptions of Proposition 3.9, we obtain that $\sum_{i \in C_k} x_i^{\text{NE}} \leqslant \sum_{i \in C_k} x_{*,i}^L$.

## 4.2 Formation of Stable Coalitions of Cooperators

Given the existence of a partial cooperative leadership equilibrium, the next objective is to address the following research question: is there a way to endogenously determine the number of cooperators in a partial cooperative framework?

For this, we introduce additional hypotheses to guarantee that a standard vNM stability conception [37] can be implemented.[7] In particular, it is required that in the game under consideration, there is a unique PCLE for every level of cooperation. We already established in Proposition 4.2 that every regular social purpose game admits at least one PCLE for every level of cooperation. Furthermore, for every strict social purpose game the constructed

---

[7] We note that in the literature, the vNM-stability concept is at the foundation of formulations of farsightedness in decision-making. We refer to [10, 16, 18, 34, 35] and [23] for further elaborations.

mapping **E** in the proof of Proposition 4.2 is a single-valued function. Therefore, we limit ourselves to this class of social purpose games throughout the following discussion.

**Axiom 4.6** *Throughout the following discussion we only consider strict social purpose games* $\Gamma$ *that admit a unique Nash equilibrium* $x^{\mathrm{NE}}$ *as well as a unique partial cooperative leadership equilibrium (PCLE) for every level of cooperation* $k \in \{2, \ldots, n\}$.

*Recalling Axiom 4.1, for every level of cooperation* $k \in \{2, \ldots, n\}$, *representing the case that* $C_k = \{n - k + 1, \ldots, n\}$ *forms as the coalition of cooperators, we denote by* $x_i^L(k) \in [0, \overline{Q}]$ *the unique PCLE strategy for every* $i \in N$ *and by* $\pi_i^L(k) = \pi_i \left( x^L(k) \right)$ *the resulting PCLE payoff of player* $i \in N$.

Under Axiom 4.6, we use the well-known vNM-stability notion. Originally, this notion of coalitional stability simply means that no player inside the coalition has an incentive to abandon her membership of the coalition of cooperators ("internal" stability) and no player outside the coalition of cooperators has an incentive to join the coalition of cooperators ("external" stability). In the applied formulation below for a level of cooperation $k$, this is only expressed for the two marginal players $n - k \in N_k$ and $n - k + 1 \in C_k$, whereas before $C_k = \{n - k + 1, \ldots, n\}$ is the corresponding coalition of cooperators under cooperation level $k$. This is founded on the specific role of these marginal players in context of Axioms 4.1. This definition has already been applied in the motivating case of the tragedy of the commons as set out in Sect. 2. The formalisation of this notion of stability is as follows.

**Definition 4.7** Consider a strict social purpose game $\Gamma$ that satisfies Axioms 4.1 and 4.6. We refer to the level of cooperation $k \in \{2, \ldots, n\}$ as

(i) **Internally stable** in $\Gamma$ if $\pi_{n-k+1}^L(k) \geqslant \pi_{n-k+1}^L(k - 1)$, and
(ii) **Externally stable** in $\Gamma$ if $k = n$ or $\pi_{n-k}^L(k) \geqslant \pi_{n-k}^L(k + 1)$

where for ease of notation we denote by $x^L(1) = x^{\mathrm{NE}}$.

The level of cooperation $k^* \in \{2, \ldots, n\}$ is **stable** if $k^*$ is an internally as well as externally stable level of cooperation in $\Gamma$.

The next result states the necessary and sufficient conditions for stability in the formation of a stable coalition of cooperators in an arbitrary *strict* social purpose game satisfying Axiom 4.6.

**Proposition 4.8** *Let* $\Gamma$ *be a strict social purpose game satisfying Axiom 4.6 and consider the level of cooperation* $k \in \{2, \ldots, n\}$. *Denote by* $q = n - k + 1 \in C_k$ *the marginal cooperator and* $r = n - k \in N_k$ *the marginal non-cooperator.[8] Then the following properties apply:*

*(i) The level of cooperation* $k = 2$ *is stable if and only if*

$$\frac{1}{\alpha_{n-1}} \left( g_{n-1}\left( x_{n-1}^L(2) \right) - g_{n-1}\left( x_{n-1}^{\mathrm{NE}} \right) \right) \leqslant H\left( \sum_{i \in N} x_i^L(2) \right) - H\left( \sum_{i \in N} x_i^{\mathrm{NE}} \right)$$

*as well as*

$$\frac{1}{\alpha_{n-2}} \left( g_{n-2}\left( x_{n-2}^L(2) \right) - g_{n-2}\left( x_{n-2}^L(3) \right) \right) \geqslant H\left( \sum_{i \in N} x_i^L(2) \right) - H\left( \sum_{i \in N} x_i^L(3) \right).$$

---

[8] We remark that as a consequence $C_{k-1} = C_k \setminus \{q\}$ and $C_{k+1} = C_k \cup \{r\}$.

(ii) *For $k \in \{3, \ldots, n-1\}$, it holds that the level of cooperation $k$ is stable if and only if*

$$\frac{1}{\alpha_q} \left( g_q \left( x_q^L(k) \right) - g_q \left( x_q^L(k-1) \right) \right) \leqslant H \left( \sum_{i \in N} x_i^L(k) \right) - H \left( \sum_{i \in N} x_i^L(k-1) \right)$$

*as well as*

$$\frac{1}{\alpha_r} \left( g_r \left( x_r^L(k) \right) - g_r \left( x_r^L(k+1) \right) \right) \geqslant H \left( \sum_{i \in N} x_i^L(k) \right) - H \left( \sum_{i \in N} x_i^L(k+1) \right).$$

(iii) *The level of cooperation $k = n$ is stable if and only if*

$$\frac{1}{\alpha_1} \left( g_1 \left( x_1^{SO} \right) - g_1 \left( x_1^L(n-1) \right) \right) \leqslant H \left( \sum_{i \in N} x_i^{SO} \right) - H \left( \sum_{i \in N} x_i^L(n-1) \right).$$

**Proof** We only state the proof for $k \in \{3, \ldots, n-1\}$. The proofs for the two extreme cases $k \in \{2, n\}$ are omitted, but follow immediately from the reasoning developed below.

We compare the payoffs of player $q = n - k + 1 \in C_k$ in the stability conditions if she cooperates with $C_k$ or acts as a non-cooperator $q \in N_{k-1}$. Indeed, if she cooperates, she receives payoff

$$u_q(C) = \pi_q^L(k) = \alpha_q H \left( \sum_{i \in N} x_i^L(k) \right) - g_q \left( x_q^L(k) \right)$$

$$= \alpha_q H \left( \sum_{i \in C_{k-1}} x_i^L(k) + x_q^L(k) + \sum_{j \in N_k} x_j^L(k) \right) - g_q \left( x_q^L(k) \right)$$

and, if she does not cooperate with $C_k$, she would receive

$$u_q(NC) = \pi_i^L(k-1)$$

$$= \alpha_q H \left( \sum_{i \in N} x_i^L(k-1) \right) - g_q \left( x_q^L(k-1) \right)$$

$$= \alpha_q H \left( \sum_{i \in C_{k-1}} x_i^L(k-1) + x_q^L(k-1) + \sum_{j \in N_k} x_j^L(k-1) \right) - g_q \left( x_q^L(k-1) \right).$$

Internal stability requires now that $u_q(C) \geqslant u_q(NC)$. This is equivalent to the first condition in the assertion.

For the external stability condition of the level of cooperation $k \in \{3, \ldots, n-1\}$, we compare the payoffs of the marginal non-cooperator $r = n - k \in N_k$ if she cooperates with $C_k$ to form $C_{k+1} = C_k \cup \{r\}$ or acts as a non-cooperator $r \in N_k$. Indeed, if she cooperates with $C_k$ she receives payoff

$$u_r(C) = \pi_r^L(k+1)$$

$$= \alpha_r H \left( \sum_{i \in N} x_i^L(k+1) \right) - g_r \left( x_r^L(k+1) \right)$$

$$= \alpha_r H \left( \sum_{i \in C_k} x_i^L(k+1) + x_r^L(k+1) + \sum_{j \in N_{k+1}} x_j^L(k+1) \right) - g_r \left( x_r^L(k+1) \right)$$

and, if she does not cooperate with $C_k$ and remains in $N_k$, she receives

$$u_r(NC) = \pi_r^L(k) = \alpha_r H\left(\sum_{i \in N} x_i^L(k)\right) - g_r\left(x_r^L(k)\right)$$

$$= \alpha_r H\left(\sum_{i \in C_k} x_i^L(k) + x_r^L(k) + \sum_{j \in N_{k+1}} x_j^L(k)\right) - g_r\left(x_r^L(k)\right)$$

External stability now requires that $u_r(NC) \geqslant u_r(C)$, which is equivalent to the second condition of the assertion.                                                                        □

## Application: Stability in a Quadratic Social Purpose Game

Consider the simple strict social purpose game $\Gamma^s$ with players $N = \{1, \ldots, n\}$ and $\overline{Q} > 0$ sufficiently large.[9] For $x = (x_1, \ldots, x_n) \in \mathbb{R}_+^n$, the payoff for player $i \in N$ is given by

$$\pi_i(x) = \alpha_i \sum_{i \in N} x_i - x_i^2 \quad \text{with } \alpha_i > 0 \quad i \in N. \tag{21}$$

Clearly, $\Gamma^s$ is a strict social purpose game, which unique Nash equilibrium is given by

$$x_i^{\text{NE}} = \tfrac{1}{2}\alpha_i \quad \text{resulting in } X^{\text{NE}} = \sum_{i \in N} x_i^{\text{NE}} = \tfrac{A}{2} \text{ with } A = \sum_{i \in N} \alpha_i. \tag{22}$$

The social optimum is unique as well and determined as

$$x_i^{\text{SO}} = \frac{1}{2}A \quad \text{resulting in } X^{\text{SO}} = \sum_{i \in N} x_i^{\text{SO}} = \tfrac{n}{2}A. \tag{23}$$

Clearly, $\sum_{i=1}^n x_i^{\text{NE}} \leqslant \sum_{i=1}^n x_i^{\text{SO}}$ and, moreover, $x_i^{\text{NE}} < x_i^{\text{SO}}$ for every $i \in N$.

Now, for every $i \in N$, the Nash equilibrium payoffs are determined as $\pi_i(x^{\text{NE}}) = \frac{\alpha_i}{4}(2A - \alpha_i) > 0$ and the socially optimal payoffs are given by $\pi_i(x^{\text{SO}}) = \frac{A}{4}(2\alpha_i n - A)$. Note that $\pi_i(x^{\text{SO}}) > 0$ if and only if $\alpha_i > \frac{A}{2n}$. Finally, $\pi_i(x^{\text{NE}}) < \pi_i(x^{\text{SO}})$ if and only if $\frac{A}{\alpha_i} - \frac{\alpha_i}{A} < 2(n-1)$.

Next, we determine the partial cooperative leadership equilibrium (PCLE) for any level of cooperation of the simple social purpose game $\Gamma^s$. Denote for any non-empty coalition $\varnothing \neq M \subseteq N$ by $A_M = \sum_{h \in M} \alpha_h > 0$.

Recalling that $0 < \alpha_1 \leqslant \alpha_2 \leqslant \cdots \leqslant \alpha_n$, select any level of cooperation $k \in \{2, \ldots, n\}$ and let $C_k = \{n-k+1, \ldots, n\} \subset N$ be the corresponding coalition of $k$ cooperators with the highest propensity for the generated common benefits. The non-cooperators are now the players in $N_k = N \setminus C_k = \{1, \ldots, n-k\}$ with the lowest preference for the collectively generated benefit.

To compute the resulting unique PCLE for the level of cooperation $k$, note that every non-cooperator selects a best response to all other players' strategies. Hence, $x_j^L(k) = \frac{1}{2}\alpha_j$ for every non-cooperator $j \in N_k$. Note that this best response is independent on the chosen strategy $x_{C_k}$ of the coalition of cooperators $C_k$.

---

[9] Hence, the strategic variable $x_i$ can be assumed as if taken from $\mathbb{R}_+$ for every $i \in N$. Given the chosen formulation of the payoff structure, it is sufficient to select $\overline{Q} = A = \sum_{i \in N} \alpha_i > 0$.

The cooperators in $C_k$ determine their strategies collectively to maximise their collective payoff $\pi_{C_k} = \sum_{i \in C_k} \pi_i$, given $(x_j^L)_{j \in N_k}$. This is again independent of the chosen strategies of all players $j \in N_k$ and results into

$$x_i^L(k) = \frac{1}{2} A_{C_k} = \frac{1}{2} \sum_{i=n-k+1}^{n} \alpha_i \qquad \text{for every cooperator } i \in C_k.$$

Hence, in the partial cooperative leadership equilibrium, we have that

$$X^L(k) = \sum_{h \in N} x_h^L(k) = \frac{1}{2} A_{N_k} + \frac{k}{2} A_{C_k} \tag{24}$$

$$\pi_j^L(k) = \alpha_j \left( \frac{1}{2} A_{N_k} + \frac{k}{2} A_{C_k} \right) - \frac{1}{4} \alpha_j^2 \qquad \text{for every } j \in N_k \tag{25}$$

$$\pi_i^L(k) = \alpha_i \left( \frac{1}{2} A_{N_k} + \frac{k}{2} A_{C_k} \right) - \frac{1}{4} A_{C_k}^2 \qquad \text{for every } i \in C_k \tag{26}$$

From these payoff levels, we deduce that for every non-cooperator $j \in N_k$: $\pi_j^{\text{NE}} < \pi_j^L(k)$ and that for every cooperator $i \in C_k$ it holds that $\pi_i^{\text{NE}} < \pi_i^L(k)$ if and only if $\frac{A_{C_k}}{\alpha_i} - \frac{\alpha_i}{A_{C_k}} < 2(k-1)$.

We next turn to the question whether partial cooperation in this simple social purpose game $\Gamma^s$ results in a stable coalition of cooperators.

**Proposition 4.9** *Under the partial cooperative leadership equilibrium concept, a level of cooperation $k \in \{2, \ldots, n-1\}$ in the game $\Gamma^s$ is stable if and only if*

$$\alpha_{n-k+1} \geqslant \frac{A_{C_k}}{1 + \sqrt{2(k-1)}} \quad \text{as well as} \quad \alpha_{n-k} \leqslant \frac{A_{C_k}}{\sqrt{2k}}. \tag{27}$$

*Furthermore, $k = n$ is a stable level of cooperation in $\Gamma^s$ to form $C_n = N$ to implement the social optimum if and only if $\alpha_1 \geqslant \frac{A_N}{1 + \sqrt{2(n-1)}}$.*

**Proof** Here, and in the following, for simplicity of notation in formulas, we let $q = n - k + 1$ and $r = n - k$ denote the marginal cooperator $q \in C_k$ and the marginal non-cooperator $r \in N_k$, respectively.

If player $q$ cooperates with $C_k$, she receives a payoff of

$$u_q(C) = \pi_q^L(k) = \alpha_q \left( \frac{1}{2} A_{N_k} + \frac{k}{2} A_{C_k} \right) - \frac{1}{4} A_{C_k}^2$$

and if she does not cooperate with $C_k$ she receives

$$u_q(NC) = \alpha_q \left( \frac{1}{2} A_{N_k} + \frac{1}{2} \alpha_q + \frac{k-1}{2} A_{C_k-q} \right) - \frac{1}{4} \alpha_q^2.$$

Internal stability requires now that $u_q(C) \geqslant u_q(NC)$. This is equivalent to

$$\frac{\alpha_q}{2} k A_{C_k} - \frac{1}{4} A_{C_k}^2 \geqslant \alpha_q \left( \frac{k-1}{2} (A_{C_k} - \alpha_q) + \frac{\alpha_q}{2} \right) - \frac{1}{4} \alpha_q^2$$

or $A_{C_k}^2 - \alpha_q^2 \leqslant 2\alpha_q A_{C_k} + 2k\alpha_q^2 - 4\alpha_q^2$, leading to $A_{C_k}^2 - 2\alpha_q A_{C_k} + \alpha_q^2 = (A_{C_k} - \alpha_q)^2 \leqslant 2(k-1)\alpha_q^2$. This is equivalent to

$$A_{C_k} - \alpha_q \leqslant \sqrt{2(k-1)} \, \alpha_q \tag{28}$$

which is equivalent to the first part of (27).

If the marginal non-cooperator $r = n - k \in N_k$ cooperates with $C_k$ to form $C_{k+1}$, she receives payoff

$$u_r(C) = \alpha_r \left( \frac{1}{2}(A_{N_k} - \alpha_r) + \frac{k+1}{2}(A_{C_k} + \alpha_r) \right) - \frac{1}{4}\left(A_{C_k} + \alpha_r\right)^2$$

and, if she does not cooperate and remains member of $N_k$, she receives

$$u_r(NC) = \pi_r^L(k) = \alpha_r \left( \frac{1}{2}A_{N_k} + \frac{k}{2}A_{C_k} \right) - \frac{1}{4}\alpha_r^2.$$

External stability now requires that $u_r(NC) \geqslant u_r(C)$, which is equivalent to

$$\frac{k}{2}\alpha_r A_{C_k} - \frac{1}{4}\alpha_r^2 \geqslant \alpha_r \left( \frac{k}{2}\alpha_r + \frac{k+1}{2}A_{C_k} \right) - \frac{1}{4}\left(A_{C_k} + \alpha_r\right)^2$$

or

$$A_{C_k}^2 + 2\alpha_r A_{C_k} \geqslant \alpha_r \left[ 2(k+1)A_{C_k} + 2k\alpha_r - 2k A_{C_k} \right] = \alpha_r \left( 2A_{C_k} + 2k\alpha_r \right) \qquad (29)$$

which is equivalent to $A_{C_k}^2 \geqslant 2k\alpha_r^2$ and, hence, the second part of (27).

Finally, $k = n$ is a stable level of cooperation if and only if $u_1(C) \geqslant u_1(NC)$. This is equivalent to $A_N - \alpha_1 \leqslant \sqrt{2(n-1)}\alpha_1$ as derived above. □

# References

1. Acemoglu D, Jensen MK (2013) Aggregative comparative statics. Games Econ Behav 81:27–49
2. Balder E.(1997) Remarks on Nash equilibria for games with additively coupled payoffs. Econ Theor 9:161–167
3. Bergstrom TC, Blume L, Varian HA (1986) On the private provison of public goods. J Public Econ 29:25–49
4. Chakrabarti S, Gilles RP, Lazarova EA (2011) Strategic behavior under partial cooperation. Theor Decis 71:175–193
5. Chakrabarti S, Gilles RP, Lazarova EA (2018) Partial cooperation in strategic decision situations. Theor Decis 85:455–478
6. Chakrabarti S, Gilles RP, Lazarova EA (2021) Stability of cartels in multi-market cournot oligopolies. Manch Sch 89:70–85
7. Chander P, Tulkens H (1997) The core of an economy with multilateral environmental externalities. Internat J Game Theory 26:379–401
8. Corchón LC (1994) Comparative statics for aggregative games the strong concavity case. Math Soc Sci 28:151–165
9. Cornes R, Hartley R (2005) Asymmetric contests with general technologies. Econ Theor 26:923–946
10. D'Aspremont C, Jacquemin A, Gabszewicz JJ, Weymark JA (1983) On the stability of collusive price leadership. Can J Econ 16:17–25

11. De Miguel V, Xu H (2009) A Stochastic multiple-leader Stackelberg model: analysis, computation, and application. Oper Res 57:1220–1235
12. Diamantoudi E, Sartzetakis ES (2006) Stable IEA: an analytical approach. J Public Econ Theory 8:247–263
13. Dubey P, Haimanko O, Zapechelnyuk A (2006) Strategic complements and substitutes, and potential games. Games Econom Behav 54:77–94
14. Frischmann BM, Marciano A, Ramello GB (2019) Tradegy of the commons after 50 years. J Econ Perspect 33:211–228
15. Gilles RP, Pesce M, Diamantaras D (2020) The provision of collective goods through a social division of labour. J Econ Behav Organ 178:287–312
16. Greenberg J (1990) The theory of social situations: an alternative game-theoretic approach. Cambridge University Press, Cambridge
17. Hardin G (1968) The tragedy of the commons. Science 162:1243–1248
18. Harsanyi J (1974) An equilibrium-point interpretation of stable sets and a proposed alternative defnition. Manage Sci 20:1472–1495
19. Jensen MK (2010) Aggregative games and best-reply potentials. Econ Theor 43:45–66
20. Kwon OS (2006) Partial international coordination in the great fish war. Environ Resour Econ 33:463–483
21. Lloyd WF (1833) Two lectures on the checks to population: delivered before the University of Oxford, in Michaelmas Term 1832. J.H. Parker Publishers, Oxford
22. Karos OS, Robles L (2021) Full farsighted rationality. Games Econom Behav 130:409–424
23. Kimya M (2020) Equilibrium coalitional behavior. Theor Econ 15:669–714
24. Mallozzi L, Messalli R (2017) Multi-leader multi-follower model with aggregative uncertainty. Games 8:25
25. Mallozzi, L. (2013): "An application of optimization theory to the study of equilibria for games: a survey. CEJOR Cent. Eur J Oper Res 21:523–539
26. Mallozzi L, Tijs S (2008) Conflict and cooperation in symmetric potential games. Int Game Theory Rev 10:245–256
27. Mallozzi L, Tijs S (2009) Coordinating choice in partial cooperative equilibrium. Econ Bull 29:1–6
28. Mallozzi L, Tijs S (2012) Stackelberg vs Nash assumption in partially cooperative games. Czech Econ Rev 6:5–13
29. McGuinty M, Milam G (2013) Public good provisions by asymmetric agents: experimental evidence. Soc Choice Welfare 40:1159–1170
30. Monderer D, Shapley LS (1996) Potential games. Games Econ Behav 14:124–143
31. Olmstead SM (2014) Climate change adaptation and water resource management: a review of the literature. Energy Econ 46:500–509
32. Ostrom E (1990) Governing the commons: the evolution of institutions for collective action. Cambridge University Press, Cambridge
33. Selten R (1970) Aggregations theorie des linear-quadratischen modells. In: Selten R (ed) Preispolitik der Mehrproduktenunternehmung in der statischen Theorie. Springer, pp 113–125
34. Ray, D. and R. Vohra (2015) The farsighted stable set. Econometrica 83:977–1011
35. Ray D, Vohra R (2019) Maximality in the farsighted stable set. Econometrica 87:1763–1779
36. Vives X (1990) Nash equilibrium with strategic complementaries. J Math Econ 19:305–321
37. von Neumann J, Morgenstern O (1947) Theory of games and economic behavior, 3rd edn. Princeton University Press, Princeton
38. von Stackelberg H (1934) Marktform und Gleichgewicht, Habilitation, University of Vienna, translated as Market Structure and Equilibrium, Damien Bazin, Lynn Urch, and Rowland Hill (translators), 2011 ed. Springer