games

MDPI

Article

The Two-Stage Game Approach to Coalition Formation: Where We Stand and Ways to Go

Achim Hagen [1], Pierre von Mouche [2] and Hans-Peter Weikard [2,*]

[1] Resource Economics Group, Humboldt Universiät zu Berlin, 10117 Berlin, Germany; achim.hagen@hu-berlin.de
[2] Section Economics, Wageningen University, 6708 PB Wageningen, The Netherlands; pierre.vanmouche@wur.nl
* Correspondence: hans-peter.weikard@wur.nl

Received: 19 July 2019; Accepted: 16 December 2019; Published: 1 January 2020

Abstract: Coalition formation is often analysed in an almost non-cooperative way, as a two-stage game that consists of a first stage comprising membership actions and a second stage with physical actions, such as the provision of a public good. We formalised this widely used approach for the case where actions are simultaneous in each stage. Herein, we give special attention to the case of a symmetric physical game. Various theoretical results, in particular, for cartel games, are provided. As they are crucial, recent results on the uniqueness of coalitional equilibria of Cournot-like physical games are reconsidered. Various concrete examples are included. Finally, we discuss research strategies to obtain results about equilibrium coalition structures with abstract physical games in terms of qualitative properties of their primitives.

Keywords: binary action game; cartel game; Cournot-like game; coalition formation; equilibrium coalition structure; potential game; symmetric game; two-stage game

1. Introduction

Non-cooperative game theory plays an important role in the modern theory of coalition formation.[1] Modelling coalition formation as a two-stage game under almost non-cooperative conditions is a very promising approach, albeit theoretically challenged. The roots of this approach can be found in articles such as [3,4] in the context of industrial organisation. However, its further development took place especially in the context of environmental economics, starting with articles such as [5,6]; see [7–9] for overviews.

In the present article, we provide a formalisation of a specific variant of the two-stage game approach in the case of complete information, transferable payoffs, and in each stage, independent simultaneous actions. We simply refer to such a game as a "two-stage coalition formation game." Formally, it is defined by providing a player set N, a membership rule R, and a game in strategic form Γ with N as the player set, called "physical game"; we denote it by $(N; R; \Gamma)$. In the first stage players choose a membership action. This leads via the membership rule to a coalition structure $\mathcal{C}$ and a corresponding coalitional game $\Gamma_{\mathcal{C}}$ which is played in the second stage. In this stage the members of each coalition in $\mathcal{C}$ cooperate and each coalition behaves like a single player, called "meta player." After each meta player has chosen a physical action, each player obtains his payoff belonging to the

[1] See [1] for a fundamental discussion of this approach. Also see [2] for an overview of coalition formation for economic models.

played physical action profile. Definition 2 provides a formal definition of $(N; R; \Gamma)$. If we speak below about a two-stage coalition formation game, we always mean such a game.[2]

The topics of the present article are the equilibrium coalition structures for a two-stage coalition formation game $(N; R; \Gamma)$. In the literature, up to now, the notion of equilibrium coalition structure presupposes that the two-stage coalition formation game is regular, meaning that each possible coalitional game $\Gamma_{\mathcal{C}}$ has a unique Nash equilibrium. Assuming regularity, the procedure to solve the game then is as follows: the unique Nash equilibrium corresponds to a physical action for each individual player, which in turn corresponds to a payoff for each individual player. In this way the two-stage game $(N; R; \Gamma)$ leads to a game in strategic form G, referred to as an "effective game." Finally, G has to be solved. Usually this is done by determining the Nash equilibria set of G.[3] Finally, the membership rule R provides for each Nash equilibrium $\mathbf{m}$ of G, an equilibrium coalition structure $R(\mathbf{m})$.

An important special case of a two-stage coalition formation game is the cartel game, where players in the first stage decide whether or not to cooperate. This leads to a coalition structure where there is a (possibly empty) coalition of cooperators and various singleton coalitions. Initially put forward to analyse the incentives of firms to join an industrial cartel ([3]), applications of cartel games have spread to other areas. In particular, the game has been used to examine incentives of countries to join international environmental agreements.[4]

A two-stage coalition formation game $(N; R; \Gamma)$ contains the very basic structure shared by various coalition formation games that have additional features like sharing and support ([14,15]).[5] We will disregard such extensions for three main reasons: (i) The 2-stage structure of the game may be lost;[6] (ii) There are many possibilities for such extensions; and (iii) most importantly, because our "simple" coalition formation game $(N; R; \Gamma)$ is, in our opinion, not sufficiently understood. For example, although there are various articles on two-stage coalition formation games, there is, to the best of our knowledge, no theoretical result about equilibrium coalition structures that holds for an abstract class of physical games in terms of qualitative properties (such as convexity, monotonicity and symmetry) of the primitives of the physical game; only results for concrete (mostly symmetric) physical games (mostly with linear or quadratic conditional payoff functions) are available (see, for example, [4,16–20]). This casts serious doubt on the robustness of these results. For the further development of the theory of basic two-stage coalition formation games, various mathematical problems have to be addressed. Therefore, the first thing to do is to make the mathematical structure of such a game more transparent. Highlighting the formal structure is one aim of the present article. As [21] also did for the cartel game, transparency is obtained, among other things, by associating a game in strategic form (i.e., the effective game G) with a two-stage coalition formation game. In this way all results for games in strategic form apply via its effective game G to two-stage coalition formation games. We discuss this further in Section 11.

Our article is self-contained. It is organised as follows. Section 2 formulates the game rules of a two-stage coalition formation game $(N; R; \Gamma)$, and Section 4 provides its formalisation together with an explanation how the game is solved. Section 2 provides also a first example. The formal definition of a two-stage coalition formation game in Section 4 uses various notions which are defined in Section 3. Section 5 introduces a class of physical games which are quite popular in the literature

---

2 A variant of this game that has also obtained a lot of attention concerns the situation where, in the second stage, first some specific coalitions takes action, and they are followed by the other coalitions. We do not consider this type of coalition formation game, as such a game is, in fact, a three-stage game.

3 For cartel games (see below), one may also look for so-called "semi-strict Nash equilibria."

4 Recent applications address, e.g., agriculture [10], biodiversity conservation [11], and vaccination [12]. For further ideas for possible extensions of the model, see [13].

5 Loosely speaking, "support" concerns the possibility for cartel games that players who do not choose for cooperation support the cooperators.

6 For example, sharing leads to a 3-stage game and support even to a 4-stage game.

about two-stage coalition formation games: Cournot-like games. Sufficient conditions are provided for the regularity of a two-stage coalition formation game with a Cournot-like game as a physical game. Section 7 deals with two-stage coalition formation games with a symmetric membership rule and a symmetric physical game; preparations for the results therein take place in Section 6. Section 9 further analyses cartel games. As the effective game of a cartel game is a binary action game, Section 8 first presents various useful results about binary action games, particularly concerning the existence of potentials. Section 10 presents some additional examples. In Section 11, we conclude with a discussion of important next steps in order to obtain more comprehensive results.

2. The Rules of the Game

To allow a clear analysis of the two-stage coalition formation game, we describe the rules of the game, and in the next section, provide the ingredients for dealing with these games.

The intended game is a two-stage game with complete information and transferable payoffs, where the players choose independently and simultaneously a (pure) action in each stage. Further rules of the game are as follows.

- The player set is $N = \{1, 2, \dots, n\}$ with $n \geq 2$. Each player i is characterised by a "membership" action set M_i, a "physical" action set X_i and with $\mathbf{X} := X_1 \times \cdots \times X_n$, by a payoff function $f_i : \mathbf{X} \to \mathbb{R}$. Let $\mathbf{M} := M_1 \times \cdots \times M_n$.
- At the first stage, the players choose, simultaneously and independently, a membership action from their membership action sets. If each player i chooses m_i, then this gives a membership action profile $\mathbf{m} \in \mathbf{M}$. This action profile leads via a given membership rule R to the coalition structure $R(\mathbf{m})$ (which is a partition of N).
- In the second stage, the players of each coalition in $R(\mathbf{m})$ coordinate their choices of a physical action like a single player, called a "meta player." The meta players simultaneously and independently choose a physical action by choosing a physical action for each of their members. The in this way a defined action profile for the meta players corresponds to a physical action profile $\mathbf{x} \in \mathbf{X}$. Each meta player $C \in R(\mathbf{m})$ obtains a payoff $\hat{P}_C(\mathbf{x}) = \sum_{l \in C} f_l(\mathbf{x})$ and each player i obtains a payoff $f_i(\mathbf{x})$.

As time is not explicitly modelled, such a game has an almost static real-world structure. Even for this, seemingly simple, real-word structure of the game it is not trivial how to solve the game. Usually, as mentioned in the introduction (also see Subsection 4.3), regularity is assumed; i.e., that there is a unique Nash equilibrium for every membership action profile $\mathbf{m}$ of the game played in the second stage, with player set $R(\mathbf{m})$, referred to as "coalitional game." Denoting with $\hat{\mathbf{x}}(\mathbf{m})$ the with this equilibrium corresponding action profile, this leads to a payoff $g_i(\mathbf{m}) = f_i(\hat{\mathbf{x}}(\mathbf{m}))$ for each player i and in this way to a game in strategic form G, referred to as "effective game": this game has player set N, for player i action set M_i and payoff function g_i. We refer to g_i as the "effective payoff function" of player i. Finally, the Nash equilibria of G are determined; if $\mathbf{m}$ is such an equilibrium, then $R(\mathbf{m})$ is an equilibrium coalition structure. (Formal definitions will be given later.)

As the remainder of the article is quite abstract and general, it may be helpful to consider a simple clarifying example of a cartel game. In a cartel game, each player has $M = \{0, 1\}$ as membership action set. The membership rule, here denoted by R_c, assigns to each membership action profile $\mathbf{m} \in M^n$ a coalition structure $R_c(\mathbf{m})$ as follows: all players i with $m_i = 1$ jointly form a coalition and each player i with $m_i = 0$ forms a singleton coalition. We refer to such a coalition structure as a "cartel coalition structure." As with a physical game, we take a concrete, symmetric, Cournot-like public good game.

Example 1. *Consider the cartel game with a physical game, the Cournot-like public good game, with common action set $X = \mathbb{R}_+$ and with payoff functions*

$$f_i(\mathbf{x}) = -\frac{1}{2} a x_i^2 + b \sum_{l=1}^{n} x_l$$

where $a, b > 0$.

The following notations for $\mathbf{m} \in M^n$ are useful:

$$\hat{S}(\mathbf{m}) = \{i \in N \mid m_i = 1\} \text{ and } \hat{S}_\star(\mathbf{m}) = \{i \in N \mid m_i = 0\}.$$

Consider the game in the second stage for a given action profile $\mathbf{m}$ chosen in the first stage. In this game the payoff function $\hat{P}_C$ of meta player $C \in R_c(\mathbf{m})$ in terms of the action profile $\mathbf{x} \in \mathbf{X} = X^n$ equals

$$\hat{P}_C(\mathbf{x}) = \begin{cases} -\frac{1}{2}ax_i^2 + bx_i + b\sum_{l \neq i} x_l \text{ if } C = \{i\} \text{ with } i \in \hat{S}_\star(\mathbf{m}), \\ \sum_{i \in C}(-\frac{a}{2}x_i^2 + \#\hat{S}(\mathbf{m})\, bx_i) + \#\hat{S}(\mathbf{m})\, b \sum_{l \notin C} x_l \text{ if } C = \hat{S}(\mathbf{m}). \end{cases}$$

It is clear that in this coalitional game, each meta player has a strictly dominant action: meta player $C = \{i\}$ has b/a as a strictly dominant action and meta player $C = \hat{S}(\mathbf{m})$ has a strictly dominant action wherein each of its members plays $\#\hat{S}(\mathbf{m})b/a$. Thus, this game has a unique Nash equilibrium, the two-stage coalition formation game is regular and for the with this equilibrium corresponding action profile $\hat{\mathbf{x}}(\mathbf{m})$ we have

$$\hat{x}_i(\mathbf{m}) = \begin{cases} \frac{b}{a} \text{ if } m_i = 0, \\ \#\hat{S}(\mathbf{m})\frac{b}{a} \text{ if } m_i = 1. \end{cases}$$

The effective payoff function of player i is the function $g_i : M^n \to \mathbb{R}$ given by $g_i(\mathbf{m}) = f_i(\hat{\mathbf{x}}(\mathbf{m}))$. Thus

$$g_i(\mathbf{m}) = \begin{cases} -\frac{a}{2}(\frac{b}{a})^2 + b\Big(\#\hat{S}(\mathbf{m})\#\hat{S}(\mathbf{m})\frac{b}{a} + (n - \#\hat{S}(\mathbf{m}))\frac{b}{a}\Big) \text{ if } m_i = 0, \\ -\frac{a}{2}(\#\hat{S}(\mathbf{m})\frac{b}{a})^2 + b\Big(\#\hat{S}(\mathbf{m})\#\hat{S}(\mathbf{m})\frac{b}{a} + (n - \#\hat{S}(\mathbf{m}))\frac{b}{a}\Big) \text{ if } m_i = 1. \end{cases}$$

The next step is to consider the Nash equilibria of the effective game G with player set N, common action set M^n, and payoff functions g_i. The game G usually is solved by determining its Nash equilibria. As $\#M = 2$, G is a binary action game. The action profile $\mathbf{m}$ being a Nash equilibrium of G comes down to $g_i'(\mathbf{m}) \leq 0$ $(i \in N)$; here $g_i'(\mathbf{m})$ is the marginal payoff of i at $\mathbf{m}$; i.e., the payoff change at $\mathbf{m}$ when player i changes his action from m_i to $1 - m_i$. As, for $\mathbf{m}$ with $\#\hat{S}(\mathbf{m}) = 1$, it holds that $R_c(\mathbf{m}) = R_c(\mathbf{0}) = \{\{l\} \mid l \in N\}$, $g_i'(\mathbf{0}) = 0$ $(i \in N)$ follows, and thus $\mathbf{0}$ is a Nash equilibrium. However, as we will see below, there are other (more interesting) Nash equilibria.

We have, writing $s = \#\hat{S}(\mathbf{m})$ and letting $w_0(s) = (s^2 - s + n - \frac{1}{2})\frac{b^2}{a}$ $(0 \leq s \leq n-1)$ and $w_1(s) = (\frac{1}{2}s^2 - s + n)\frac{b^2}{a}$ $(1 \leq s \leq n)$,

$$g_i(\mathbf{m}) = \begin{cases} w_0(s) \text{ if } m_i = 0, \\ w_1(s) \text{ if } m_i = 1. \end{cases}$$

This implies $g_i'(\mathbf{m}) = w_1(s+1) - w_0(s)$ if $m_i = 0$ and $g_i'(\mathbf{m}) = w_0(s-1) - w_1(s)$ if $m_i = 1$. With

$$t(s) = w_1(s) - w_0(s-1) \text{ if } 1 \leq s \leq n,$$

we have derived the formula

$$g_i'(\mathbf{m}) = \begin{cases} t(s+1) \text{ if } m_i = 0, \\ -t(s) \text{ if } m_i = 1. \end{cases}$$

Letting $t(0) = +\infty$ and $t(n+1) = -\infty$, this formula implies

$$\mathbf{m} \text{ is a Nash equilibrium } \Leftrightarrow t(s+1) \leq 0 \leq t(s).$$

With the above explicit expression for $g_i(\mathbf{m})$ we find for $1 \leq s \leq n$

$$t(s) = \Big((\frac{1}{2}s^2 - s + n) - ((s-1)^2 - (s-1) + n - \frac{1}{2})\Big)\frac{b^2}{a} = (-\frac{1}{2}s^2 + 2s - \frac{3}{2})\frac{b^2}{a}.$$

So $t(0) = +\infty,\ t(1) = 0,\ t(2) = 1/2,\ t(3) = 0,\ t(4) = -3/2,\ t(5) = -14, \ldots, t(n+1) = -\infty$.

This implies that, for $n \geq 3$, $\mathbf{m} \in M^n$ *is a Nash equilibrium if and only if* $\hat{S}(\mathbf{m}) \in \{0, 2, 3\}$ *and the equilibrium coalition structures are exactly the cartel coalition structures with* 0, 2 *or* 3 *cooperators. For* $n = 2$ *the equilibrium coalition structures are exactly the cartel coalition structures with* 0 *or* 2 *cooperators.*

Observe that in this example G is a symmetric aggregative game,[7] *there exists an effective equilibrium cartel coalition structure (i.e., an equilibrium cartel coalition structure not equal to* $\{\{1\}, \{2\}, \ldots, \{n\}\}$*), that the sign of* $t(s)$ *does not depend on a and b and that the functions* w_0 *and* w_1 *are increasing.* $\diamond$

3. Fundamental Objects

In this section we provide the fundamental formalisations for dealing with two-stage coalition formation games.

3.1. Games in Strategic Form

A **game in strategic form** Γ is an ordered 3-tuple

$$\Gamma = (I; (X_i)_{i \in I}; (f_i)_{i \in I}),$$

where I is a non-empty finite set, every X_i is a non-empty set and every f_i is a function

$$f_i : \mathbf{X}_I \to \mathbb{R},$$

where

$$\mathbf{X}_I := \prod_{i \in I} X_i;$$

i.e., the product of the family $(X_i)_{i \in I}$. The set I is called **player set** and its elements **players**. The set X_i is called the **action set** of player i and its elements **actions** of player i.[8] The function f_i is called the **payoff function** of player i and the elements of $\mathbf{X}_I$, being by I indexed families $(x_i)_{i \in I}$ with $x_i \in X_i$, are called **action profiles**. For $i \in I$, we write

$$\hat{\imath} := I \setminus \{i\} \text{ and } \mathbf{X}_{\hat{\imath}} := \prod_{j \in \hat{\imath}} X_j.$$

For $i \in I$ and $\mathbf{z} = (z_j)_{j \in \hat{\imath}} \in \mathbf{X}_{\hat{\imath}}$, define the **conditional payoff function** $f_i^{(\mathbf{z})} : X_i \to \mathbb{R}$ by

$$f_i^{(\mathbf{z})}(x_i) \ := \ f_i(x_i; \mathbf{z});$$

here, $(x_i; \mathbf{z})$ is a by I indexed family with x_i for the element with index i and z_j for the element with index $j \neq i$. Also define for $i \in I$ the **best-reply correspondence** $R_i : \mathbf{X}_{\hat{\imath}} \multimap X_i$ by

$$R_i(\mathbf{z}) := \operatorname{argmax} f_i^{(\mathbf{z})}.$$

Thus, the best-reply correspondence for a player assigns to each given strategy profile of his opponents, the set of actions that maximise his payoff. An action profile $\mathbf{x} = (x_i)_{i \in I} \in \mathbf{X}_I$ is a **(Nash) equilibrium** of Γ if, for all $i \in I$, writing (given i) again $\mathbf{x} = (x_i; \mathbf{z})$, x_i is a maximiser of the conditional payoff function $f_i^{(\mathbf{z})}$, i.e., $x_i \in R_i(\mathbf{z})$. We denote by

$$E(\Gamma)$$

---

[7] If wished, see Subsection 3.1 for these notions.

[8] The actions here are intended to be pure (and not mixed) actions.

the set of Nash equilibria of Γ.

Often, when dealing with games in strategic form, one takes $N = \{1, 2, \ldots, n\}$ for the player set, and then, for the set of action profiles, instead of the product $\mathbf{X}_N = \prod_{i \in N} X_i$, the Cartesian product $\mathbf{X} = \times_{i=1}^n X_i = X_1 \times \cdots \times X_n$. However, when dealing below with coalitional games where the players are subsets (so-called "meta players"), it is more natural not to number these subsets with $1, 2, \ldots$, but just to identify them with the subsets themselves; this then leads to a game with action profiles in a family product (instead of a Cartesian product). Below, in Definition 2, when defining our intended two-stage coalition formation game $(N; R; \Gamma)$, the game in strategic form Γ will have as player set $N = \{1, 2, \ldots, n\}$ and the action profiles of the coalitional games will be elements of a family product.

A game in strategic form $\Gamma = (N; (X_i)_{i \in N}; (f_i)_{i \in N})$ where $N = \{1, 2, \ldots, n\}$ and $X_1 = \cdots = X_n =: X$ is **symmetric** if for each permutation π of N, every $i \in N$ and every $\mathbf{x} \in \mathbf{X} = X^n$

$$f_{\pi(i)} = f_i \circ T_{\pi^{<-1>}}; \tag{1}$$

i.e., $f_{\pi(i)}(x_1, \ldots, x_n) = f_i(x_{\pi(1)}, \ldots, x_{\pi(n)})$ for every $(x_1, \ldots, x_n) \in X^n$; here, the mapping $T_\pi : X^n \to X^n$ is defined by

$$T_\pi(x_1, \ldots, x_n) = (x_{\pi^{<-1>}(1)}, \ldots, x_{\pi^{<-1>}(n)}).$$

Many economic games, like Cournot-like games (see Section 5) are aggregative games. Aggregative games admit special, powerful techniques (see for instance [22–24]). Various definitions exist for the notion of an aggregative game. Here, we present a general one for a game in strategic form with player set N and actions sets X_i, subsets of $\mathbb{R}_+$. In order to do so, let $Y := \sum_{l \in N} X_l$ and $T_i := \sum_{j \neq i} X_j$.[9] Also, for $i \in N$, let $\Delta_i := \{(x_i, y) \in X_i \times Y \mid y - x_i \in T_i\}$. Next note that the following two properties are equivalent:

1. For every $i \in N$ and $z \in T_i$ there exists a function $\tilde{f}_i^{(z)} : X_i \to \mathbb{R}$ such that

$$f_i(\mathbf{x}) = \tilde{f}_i^{(\sum_{j \neq i} x_l)}(x_i) \ (\mathbf{x} \in \mathbf{X}). \tag{2}$$

2. For every $i \in N$ there exists a function $\pi_i : \Delta_i \to \mathbb{R}$ such that $f_i(\mathbf{x}) = \pi_i(x_i, \sum_l x_l)$ $(\mathbf{x} \in \mathbf{X})$.

Having said this, the game is **aggregative** if it satisfies one, thereby satisfying both properties. Note that if $\pi_1 = \cdots = \pi_n$, then ($X_1 = \cdots = X_n$ and) the game is symmetric.

3.2. Coalition Structures

In this subsection n is a positive integer and $N := \{1, 2, \ldots, n\}$. A **partition of the set** N is a set with as elements non-empty disjoint subsets of N whose union is N.[10] A related notion is a **partition of the positive integer** n; this is a finite sequence $(k_1, \ldots, k_s)$ of positive integers such that $k_1 \geq k_2 \geq \cdots \geq k_s$ and $k_1 + \cdots + k_s = n$.[11] We denote the set of partitions of N by $\mathscr{C}_N$ and refer to a partition of N also as **coalition structure**. This last notion is used in order to define now below the notion of congruent "coalition structures."[12]

Given a coalition structure $\mathcal{C}$, we denote for $i \in N$ by $\mathcal{C}_{[i]}$ the unique element of $\mathcal{C}$ with

$$i \in \mathcal{C}_{[i]}.$$

---

9 The sums here are Minkowski sums.

10 For example, there are 5 partitions of $\{1, 2, 3\}$: $\{\{1\}, \{2\}, \{3\}\}$, $\{\{1\}, \{2, 3\}\}$, $\{\{2\}, \{1, 3\}\}$, $\{\{3\}, \{1, 2\}\}$, and $\{\{1, 2, 3\}\}$.

11 For example, there are 11 partitions of 6: $(6), (5,1), (4,2), (4,1,1), (3,3), (3,2,1), (3,1,1,1), (2,2,2), (2,2,1,1), (2,1,1,1,1)$, and $(1,1,1,1,1,1)$. The number p_n of partitions of n is a rapidly increasing function of n. For example, $p_1 = 1$, $p_4 = 5$, $p_9 = 30$, $p_{50} = 204226$, $p_{200} = 3972999029388$. The number of coalition structures of N is given by the Bell-number B_n. One has $B_0 = 1, B_1 = 1, B_2 = 2, B_3 = 5, \ldots$. The following formula holds: $B_{n+1} = \sum_{k=0}^n \binom{n}{k} B_k$.

12 Congruent coalition structures play an important role, as we shall see, in the case of symmetric physical games.

Given a coalition structure $\mathcal{C}$, the sizes (i.e., number of elements) of the subsets in $\mathcal{C}$ induce in a natural way a partition $[\mathcal{C}]$ of n. On $\mathscr{C}_N$ the relation $\sim$ defined by

$$\mathcal{C} \sim \mathcal{C}' \text{ means } [\mathcal{C}] = [\mathcal{C}'], \tag{3}$$

which is an equivalence relation.[13] We refer to $\mathcal{C} \sim \mathcal{C}'$ as $\mathcal{C}$ and $\mathcal{C}'$ are **congruent**. It is clear that coalition structures $\mathcal{C}, \mathcal{C}'$ of N are congruent if and only if there exists a permutation π of N such that $\mathcal{C}' = \{\pi(C) \mid C \in \mathcal{C}\}$.

Finally, the following additional terminology may be useful: we call the coalition structure $\{\{1\}, \ldots, \{n\}\}$ the **ineffective** coalition structure and refer to every other coalition structure as **effective** coalition structure.

3.3. Coalitional Equilibria

In this subsection n is again a positive integer and $N = \{1, 2, \ldots, n\}$. Consider a game in strategic form Γ with player set N:

$$\Gamma = (N; (X_i)_{i \in N}; (f_i)_{i \in N}).$$

For every partition $\mathcal{C}$ of N we are going to define a game in strategic form $\Gamma_{\mathcal{C}}$ with a player set the elements of $\mathcal{C}$, being non-empty subsets of N. Therefore, it is useful to refer to a non-empty subset of N as **coalition** and to a partition of N as a **coalition structure** (of N).

Given the game in strategic form Γ, we introduce the following notations. First, for a coalition C

$$K_C := \prod_{l \in C} X_l.$$

So an element k_C of K_C is a by C indexed family $(k_{C;l})_{l \in C}$ with $k_{C;l} \in X_l$ $(l \in C)$:

$$k_C = (k_{C;l})_{l \in C}.$$

Second, for a coalition structure $\mathcal{C}$, we define the mapping $J_{\mathcal{C}} : \prod_{C \in \mathcal{C}} K_C \to \mathbf{X}$ by

$$J_{\mathcal{C}}((k_C)_{C \in \mathcal{C}}) := (k_{\mathcal{C}_{[i]};i})_{i \in N}. \tag{4}$$

We refer to it as the **canonical mapping**. We call $J_{\mathcal{C}}(\mathbf{k})$ the with **k associated action profile** in Γ. Note that $J_{\mathcal{C}}$ is a bijection.

Having these notations, we are ready to formalise the intended notion of coalitional equilibrium (with physical game Γ) as already outlined in section 1.

Definition 1. *Given a game in strategic form $\Gamma = (N; (X_i)_{i \in N}; (f_i)_{i \in N})$ and a coalition structure $\mathcal{C}$ of N, the (**with $\mathcal{C}$ associated**) game in strategic form $\Gamma_{\mathcal{C}}$ is defined as the game in strategic form*

$$\Gamma_{\mathcal{C}} := (\mathcal{C}; (K_C)_{C \in \mathcal{C}}; (P_C)_{C \in \mathcal{C}}) \text{ where } P_C := \sum_{i \in C} f_i \circ J_{\mathcal{C}}. \ \diamond$$

So $P_C : \mathbf{K}_{\mathcal{C}} \to \mathbb{R}$ and $P_C(\mathbf{k}) = \sum_{i \in C} f_i(J_{\mathcal{C}}(\mathbf{k}))$.

It may be appropriate to define for meta player C his **payoff function in terms of physical action profiles** by

$$\hat{P}_C(\mathbf{x}) = P_C(J_C^{<-1>}(\mathbf{x})).$$

---

[13] For example, if $N = \{1,2,3,4,5,6,7\}$, then for the coalition structure $\mathcal{C} = \{\{3,4,5\}, \{1,2\}, \{6\}, \{7\}\}$, we have $[\mathcal{C}] = (3,2,1,1)$ and $\mathcal{C}_{[2]} = \{1,2\}$. And for $\mathcal{C}' = \{\{1,4,5\}, \{2,3\}, \{6\}, \{7\}\}$, we have $[\mathcal{C}'] = (3,2,1,1)$, $\mathcal{C}'_{[2]} = \{2,3\}$ and $\mathcal{C} \sim \mathcal{C}'$.

So $\hat{P}_C(\mathbf{x}) = \sum_{i \in C} f_i(\mathbf{x})$.

Clearly the payoff function P_C of the coalitional game $\Gamma_{\mathcal{C}}$ is completely determined by $\mathcal{C}, C$ and the $f_i \ (i \in C)$.[14] The intended interpretation is that in $\Gamma_{\mathcal{C}}$ the players inside each coalition coordinate their actions. We also will refer to the elements of $\mathcal{C}$ as **meta players**. And we refer to $\Gamma_{\mathcal{C}}$ also as a **coalitional game**. A Nash equilibrium of $\Gamma_{\mathcal{C}}$ is also called a **coalitional Nash equilibrium** of Γ; more precisely, we speak of a $\mathcal{C}$-equilibrium of Γ. As far as we know, coalitional games first were considered in [26]. For more on coalitional equilibria (and related notions), see [27].

The action sets $\mathbf{X}_C$ of $\Gamma_{\mathcal{C}}$ are typically more dimensional. Note that if

$$\mathcal{C} = \{\{1\}, \{2\}, \dots \{n\}\},$$

then $\Gamma_{\mathcal{C}} = \Gamma$ and a $\mathcal{C}$-equilibrium of Γ is nothing else than a Nash equilibrium of Γ. And if

$$\mathcal{C} = \{N\},$$

then a $\mathcal{C}$-equilibrium is nothing else than a maximiser of the total payoff function $\sum_{i \in N} f_i$.

3.4. Membership Rules

In this subsection n again is a positive integer and $N = \{1, 2, \dots, n\}$. A **membership rule** (for N) is a mapping

$$R : \mathbf{M} \to \mathscr{C}_N,$$

where the M_l non-empty sets and $\mathbf{M} = \times_{i=1}^{n} M_i$, with the following property: the ineffective coalition structure $\{\{1\}, \dots, \{n\}\}$ is in the image of R.[15] A classification of membership rules can be found in [28].

Given a membership rule $R : \mathbf{M} \to \mathscr{C}_N$, a coalition structure is said to be **possible** if it belongs to $R(\mathbf{M})$. Thus, the ineffective coalition structure is possible.

We call a membership rule **symmetric** if the membership action sets are identical and for each permutation π of N and $\mathbf{m} \in \mathbf{M}$

$$R(T_\pi(\mathbf{m})) = \{\pi(C) \mid C \in R(\mathbf{m})\}. \tag{5}$$

The following result should be clear:

Proposition 1. *For a symmetric membership rule R, a permutation π of N and action profiles $\mathbf{m}, \mathbf{m}' \in \mathbf{M}$ it holds that $\mathbf{m}' = T_\pi(\mathbf{m}) \ \Rightarrow R(\mathbf{m}) \sim R(\mathbf{m}')$.*[16] $\diamond$

4. Two-Stage Coalition Formation Games

4.1. Notion

After having presented the rules of the game in Section 2, we now focus on the formal definition of the two-stage coalition formation game.

---

14 This construction may not always be realistic. For example, in the case where the physical game Γ is a Cournot oligopoly, one may imagine that the resulting cost function for a meta-player C in $\Gamma_{\mathcal{C}}$ is not obtained as a sum $\sum_{i \in C} c_i(k_{C;i})$ of the individual cost functions (see, for example, [25]).

15 Note that we do not assume that R is injective. For example R_c in Definition 9 below is not injective.

16 We note that for a symmetric membership rule R, the implication $R(\mathbf{m}) \sim R(\mathbf{m}') \Rightarrow \mathbf{m}' = T_\pi(\mathbf{m})$ for some permutation π of N does not necessarily hold. For example it does not hold for the cartel membership rule R_c in Definition 9.

Definition 2. *A* ***two-stage coalition formation game*** *is a* 3*-tuple*

$$(N; R; \Gamma)$$

where $\Gamma = (N; (X_i)_{i \in N}; (f_i)_{i \in N})$ *is a game in strategic form with player set* $N = \{1, 2, \ldots, n\}$ *with* $n \geq 2$*, called the* ***physical game*** *and* $R : \mathbf{M} \to \mathscr{C}_N$ *is a membership rule for* N*.* $\diamond$

We call M_i the **membership action set** of player i and X_i his **physical action set**.

An additional possibility for the above game is to allow for payoff transfers between the players. This occurs, for example, if players who join a coalition redistribute their payoffs according to some sharing rule; for example, equal sharing, where the total payoff is divided equally among the coalition members. Although sharing is an important issue, we will not deal with it, since a general formalisation is complicated (for example, the additional rules may refer explicitly to the chosen membership actions and physical actions). In addition, sharing would destroy the two-stage structure of the game.[17]

4.2. Effective Game

As a two-stage coalition formation game is a two-stage game with, in both stages, simultaneous and independent actions, it is not so clear how to "solve" such a game ([31]); an extra complication here is that the player set of the second stage is, in general, not equal to that of the first stage.

The existing literature handles the solving issue by "looking for subgame perfect equilibria": first one solves the second stage by determining for each possible coalitional game[18] its Nash equilibria, and then the first one. In doing so it is assumed that each possible coalitional game has a unique Nash equilibrium. Below we shall make this precise.

Definition 3. *A two-stage coalition formation game* $(N; R; \Gamma)$ *is* ***regular*** *if for each possible coalition structure the coalitional game* $\Gamma_{\mathcal{C}}$ *has a unique Nash equilibrium. In this case this Nash equilibrium is denoted by*

$$\mathbf{e}^{(\mathcal{C})}. \ \diamond$$

Definition 4. *Let* $(N; R; \Gamma)$ *be a regular two-stage coalition formation game. Its* ***effective game*** *is the game in strategic form* $G = (N; (M_i)_{i \in N}; (g_i)_{i \in N})$ *defined by*

$$g_i(\mathbf{m}) := f_i(\hat{\mathbf{x}}(\mathbf{m})). \tag{6}$$

with $\hat{\mathbf{x}} : \mathbf{M} \to \mathbf{X}$ *given by*[19]

$$\hat{\mathbf{x}}(\mathbf{m}) = (\hat{x}_1(\mathbf{m}), \ldots, \hat{x}_n(\mathbf{m})) := J_{R(\mathbf{m})}(\mathbf{e}^{(R(\mathbf{m}))}).$$

We refer to the function g_i *as* ***effective payoff*** *of player i and to* $\hat{x}_i(\mathbf{m})$ *as* ***effective physical action*** *(of player i associated with* $\mathbf{m}$*).*[20] $\diamond$

---

[17] In the literature (for example, in [19,29,30]) various sharing rules are used which are incompatible with the game rules in Section 2 for the two-stage game. The problem is that these rules refer to effective payoffs (see Subsection 4.2) which only are known after the game has been solved. In particular, this applies to so-called "optimal sharing" that refers to the payoff of a player who cooperates in the case he would not have cooperated. However, as we shall see, in Subsection 9.2, the ideas related to optimal sharing will "survive" for situations where the effective game has the so-called deviation property D_1.

[18] I.e., coalitional game $\Gamma_{\mathcal{C}}$ where $\mathcal{C}$ is a possible coalition structure.

[19] Using the notation (4).

[20] The formal object of "effective payoff" has a close relation to what is called "valuation" in the theory of partition function games. However, their precise mathematical structure is different. Using effective payoffs and effective physical actions, the strategic form structure becomes much more visible.

So we have the formula

$$g_i(\mathbf{m}) = f_i(J_{R(\mathbf{m})}(\mathbf{e}^{(R(\mathbf{m}))}). \tag{7}$$

Note that the two-stage coalition formation game is almost completely non-cooperative; the only place where cooperative aspects enter is in the coordination of the choices by the coalition members.

4.3. Solving the Two-Stage Game

Consider a regular two-stage coalition formation game $(N; R; \Gamma)$ with effective game G. The game G is a game in strategic form with $\mathbf{M}$ as set of action profiles. In the literature the two-stage game is solved by determining the Nash equilibria set

$$E(G)$$

of G. For every $\mathbf{m} \in E(G)$ we refer to $R(\mathbf{m})$ as a **equilibrium coalition structure**.

In the context of a cartel game (with the cartel membership rule R_c), one also is interested in semi-strict Nash equilibria (see Section 9). Denoting the set of semi-strict Nash equilibria by

$$E_{ss}(G),$$

we refer for every $\mathbf{m} \in E_{ss}(G)$ to $R_c(\mathbf{m})$ as a **semi-strict equilibrium coalition structure**.

5. Cournot-Like Games

5.1. Notion

When dealing with two-stage coalition formation games, Cournot games and public good games are popular physical games. As such games have a common structure and admit a unified analysis, in [32] the following class of games was introduced:

Definition 5. *A **Cournot-like game** is a game in strategic form*

$$(N; (X_i)_{i \in N}; (f_i)_{i \in N})$$

where every X_i is a subset of[21] *$\mathbb{R}_+$ with $0 \in X_i$ and*

$$f_i(\mathbf{x}) = p_i(x_i) - x_i^{\beta_i} q_i(\sum_{l \in N} \gamma_l x_l)$$

where, with (the Minkowski sum) $Y := \sum_{l \in N} \gamma_l X_l$,

- *$p_i : X_i \to \mathbb{R}$ and $q_i : Y \to \mathbb{R}$;*
- *$\beta_i \in \{0, 1\}$ and $\gamma_l > 0$.* $\diamond$

The abstract class of Cournot-like games contains various heterogeneous Cournot oligopoly games: take every $\beta_i = 1$. It contains all[22] homogeneous Cournot oligopoly games: take, in addition, all q_i equal and each $\gamma_l = 1$. It also contains various public good games: take every $\beta_i = 0$.

5.2. Uniqueness of Coalitional Equilbria

Let us consider the regularity issue for two-stage coalition formation games with a Cournot-like game as physical game. This issue mainly concerns a Nash equilibrium semi-uniqueness

[21] Mostly X_i even is a proper real interval.
[22] Disregarding cases with finite action sets.

problem[23] as conditions for existence do not seem to be problematic in the relevant literature;[24] in particular equilibrium existence, results, à la Nikaido-Isoda, are useful (see [33]). Already in the case of one-dimensional action sets, for example in the classical Cournot oligopoly, the equilibrium semi-uniqueness problem poses more serious problems than the existence problem (see, for example, [34,35]). For coalitional games this problem is even more complicated, as in such such games action sets may be higher dimensional.

In [36], the above problem was approached by developing an equilibrium semi-uniqueness result for games in strategic form with higher dimensional action sets. The next theorem follows from the results in [36]. This theorem deals with Cournot-like games as physical games.

Theorem 1.

1. *Consider a homogeneous Cournot oligopoly game, with $f_i(\mathbf{x}) = a_i(x_i) - x_i b(\sum_{l \in N} x_l)$, with compact action sets, differentiable strictly concave a_i and with differentiable increasing convex b. Then for every coalition structure $\mathcal{C}$ the game has a unique $\mathcal{C}$-equilibrium.*
2. *Consider a public good game with $f_i(\mathbf{x}) = a_i(x_i) - b_i(\sum_{l \in N} x_l)$, with compact action sets, differentiable strictly concave a_i and with differentiable increasing convex b_i. Then for every coalition structure $\mathcal{C}$ the game has a unique $\mathcal{C}$-equilibrium.* ⋄

Proof. By Corollary 3 in [36]. □

Corollary 1. *A two-stage coalition formation game $(N; R; \Gamma)$ with as physical game Γ a Cournot-like game as in Theorem 1 is regular.* ⋄

Example 2. *A simple example of a Cournot-like game that does not have the property that for every coalition structure $\mathcal{C}$ there is a unique $\mathcal{C}$-equilibrium, is the following: each player has action set $\mathbb{R}_+$ and the payoff functions are*

$$f_i(\mathbf{x}) = \Big(\sum_{l \in N} x_l\Big)^{1/2} - x_i.$$

In order to see this, consider the coalitional game for the coalition structure $\{N\}$. This game has one player: the meta-player N. The payoff function of this game is the function $\mathbf{k} \mapsto n\left(\sum_{l \in N} k_l\right)^{1/2} - \sum_{l \in N} k_l$. The set of Nash equilibria of this game consists of the actions $\mathbf{k}$ with $\sum_{l=1}^{n} k_l = n^2/4$. Thus, there are infinitely many $\{N\}$-equilibria. ⋄

6. Coalitional Equilibria of Symmetric Games

The symmetry notion (1) for a game in strategic form Γ presupposes that each player has the same action set. This implies for a coalitional game $\Gamma_{\mathcal{C}}$ of Γ that "symmetric" may not be well-defined, even if Γ is symmetric, since the coalition structure may comprise meta players of unequal size. Proposition 2 below shows that for a symmetric Γ the Nash equilibria of the coalitional games of Γ nevertheless have some symmetry properties.

In the rest of this section we consider a game in strategic form $\Gamma = (N; (X_i)_{i \in N}; (f_i)_{i \in N})$.

Given a coalition structure $\mathcal{C}$ and a permutation π of N, let $\mathcal{C}'$ be the coalition structure $\mathcal{C}' = \{\pi^{<-1>}(C) \mid C \in \mathcal{C}\}$ and $\mathcal{U}_\pi : \prod_{C \in \mathcal{C}} K_C \to \prod_{C' \in \mathcal{C}'} K_{C'}$ be the mapping defined by

$$\mathcal{U}_\pi := J_{\mathcal{C}'}^{<-1>} \circ T_{\pi^{<-1>}} \circ J_{\mathcal{C}}.$$

[23] Equilibrium uniqueness comes down to equilibrium existence and to equilibrium semi-uniqueness, i.e., that there exists at most one equilibrium.

[24] Of course, for concrete games where one can show by straightforward calculation that there is at most one Nash equilibrium the problem is not serious.

$\mathcal{U}_\pi$ is, being a composition of bijections, a bijection. The reason to introduce $\mathcal{U}_\pi$ is that with it we have the result in Lemma 1(2). Denoting payoff functions of $\Gamma_{\mathcal{C}}$ with P_C and those of $\Gamma_{\mathcal{C}'}$ with $P'_{C'}$, we have for this situation the following lemma.

Lemma 1. *Suppose* Γ *is symmetric.*

1. *For every* $C \in \mathcal{C}$ *the identity* $P_C = P'_{\pi^{<-1>}(C)} \circ \mathcal{U}_\pi$ *holds.*
2. $\mathcal{U}_\pi(E(\Gamma_{\mathcal{C}})) = E(\Gamma_{\mathcal{C}'})$. $\diamond$

Proof. 1. With (1) we obtain $P_C = \sum_{i \in C} f_i \circ J_{\mathcal{C}} = \sum_{i \in \pi^{<-1>}(C)} f_{\pi(i)} \circ J_{\mathcal{C}} = \sum_{i \in \pi^{<-1>}(C)} f_i \circ T_{\pi^{<-1>}} \circ J_{\mathcal{C}} = \sum_{i \in \pi^{<-1>}(C)} f_i \circ J_{\mathcal{C}'} \circ \mathcal{U}_\pi = P'_{\pi^{<-1>}(C)} \circ \mathcal{U}_\pi$.

2. First we prove that $\mathcal{U}_\pi(E(\Gamma_{\mathcal{C}})) \subseteq E(\Gamma_{\mathcal{C}'})$. So suppose $\mathbf{k} \in E(\Gamma_{\mathcal{C}})$. Let $\mathbf{k}' = \mathcal{U}_\pi(\mathbf{k})$. We are going to prove that $\mathbf{k}' \in E(\Gamma_{\mathcal{C}'})$. In order to do so, we fix in $\Gamma_{\mathcal{C}'}$, a meta player D' and an action profile $\mathbf{k}''$ with $k'_{C'} = k''_{C'}$ ($C' \in \mathcal{C}'$ with $C' \neq D'$) of this game and show that

$$P'_{D'}(\mathbf{k}'') \leq P'_{D'}(\mathbf{k}').$$

Well, let $A \in \mathcal{C}$ be such that $D' = \pi^{<-1>}(A)$ and let $\mathbf{d} = \mathcal{U}_\pi^{<-1>}(\mathbf{k}'')$; we have

$$J_{\mathcal{C}}(\mathbf{d}) = (T_\pi \circ T_{\pi^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{d}) = (T_\pi \circ J_{\mathcal{C}'})(\mathbf{k}'').$$

We first prove

$$d_C = k_C \ (C \in \mathcal{C} \text{ with } C \neq A).$$

In order to do so, we fix $C \in \mathcal{C}$ with $C \neq A$ and $l \in C$.
As $\pi^{<-1>}(C) \neq D'$, we have $k'_{\pi^{<-1>}(C);\pi^{<-1>}(l)} = k''_{\pi^{<-1>}(C);\pi^{<-1>}(l)}$. With this

$$\begin{aligned} d_{C;l} &= (J_{\mathcal{C}}(\mathbf{d}))_l = ((T_\pi \circ J_{\mathcal{C}'})(\mathbf{k}''))_l = (J_{\mathcal{C}'}(\mathbf{k}''))_{\pi^{<-1>}(l)} = k''_{\pi^{<-1>}(C);\pi^{<-1>}(l)} \\ &= k'_{\pi^{<-1>}(C);\pi^{<-1>}(l)} = (J_{\mathcal{C}'}(\mathbf{k}'))_{\pi^{<-1>}(l)} \\ &= ((T_{\pi^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{k}))_{\pi^{<-1>}(l)} = (J_{\mathcal{C}}(\mathbf{k}))_l = k_{C;l}. \end{aligned}$$

As $\mathbf{k} \in E(\Gamma_{\mathcal{C}})$, we have $P_A(\mathbf{d}) \leq P_A(\mathbf{k})$. With this, as Γ is symmetric, we obtain with part 1, as desired,

$$P'_{D'}(\mathbf{k}'') = P'_{\pi^{<-1>(A)}}(\mathcal{U}_\pi(\mathbf{d})) = P_A(\mathbf{d}) \leq P_A(\mathbf{k}) = P'_{\pi^{<-1>(A)}}(\mathcal{U}_\pi(\mathbf{k})) = P'_{D'}(\mathbf{k}').$$

In the same way as above we can show that $\mathcal{U}_\pi^{<-1>}(E(\Gamma_{\mathcal{C}'})) \subseteq E(\Gamma_{\mathcal{C}})$. This implies $E(\Gamma_{\mathcal{C}'}) = (\mathcal{U}_\pi \circ \mathcal{U}_\pi^{<-1>})(E(\Gamma_{\mathcal{C}'})) \subseteq \mathcal{U}_\pi(E(\Gamma_{\mathcal{C}}))$. $\square$

Proposition 2. *Suppose* Γ *is symmetric. Fix a coalition structure* $\mathcal{C}$*. Suppose the game* $\Gamma_{\mathcal{C}}$ *has a unique Nash equilibrium* $\mathbf{k} = (k_C)_{C \in \mathcal{C}}$.

1. *For every* $C \in \mathcal{C}$*, the action* k_C *of meta player* C *is constant; i.e.,* $k_{C;i} = k_{C;j}$ $(i, j \in C)$.
2. *For every* $C, C' \in \mathcal{C}$ *with* $\#C = \#C'$*, it holds that* k_C *and* $k_{C'}$ *are the same constant.*
3. *For every* $C, C' \in \mathcal{C}$ *with* $\#C = \#C'$*, it holds that* $f_i(J_{\mathcal{C}}(\mathbf{k})) = f_j(J_{\mathcal{C}}(\mathbf{k}))$ $(i \in C, j \in C')$.

Further let $\mathcal{C}'$ *be a with* $\mathcal{C}$ *congruent coalition structure.*

4. $\Gamma_{\mathcal{C}'}$ *has a unique Nash equilibrium* $\mathbf{k}'$.
5. *For every* $C \in \mathcal{C}$ *and* $C' \in \mathcal{C}'$ *with* $\#C = \#C'$*, the actions* k_C *and* $k'_{C'}$ *are the same constant.*[25]

[25] By parts 1 and 4, these are indeed constant.

6. *For every $C \in \mathcal{C}$ and $C' \in \mathcal{C}'$ with $\#C = \#C'$, it holds that $f_i(J_{\mathcal{C}}(\mathbf{k})) = f_j(J_{\mathcal{C}'}(\mathbf{k}'))$ $(i \in C,\ j \in C')$.* $\diamond$

Proof. 1. By contradiction, suppose $C \in \mathcal{C}$ is such that $\mathbf{k}_C$ is not constant. Fix $l, m \in C$ such that $k_{C;l} \neq k_{C;m}$. Let π be a permutation of N with $\pi(B) = B$ $(B \in \mathcal{C})$ such that $\pi(l) = m$. Now $\mathcal{C}'$ in Lemma 1 equals $\mathcal{C}$. Let $\mathbf{k}' = \mathcal{U}_\pi(\mathbf{k})$. As $k'_{C;l} = (J_{\mathcal{C}}(\mathbf{k}'))_l = J_{\mathcal{C}}(\mathcal{U}_\pi(\mathbf{k})) = ((T_{\pi^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{k}))_l = (J_{\mathcal{C}}(\mathbf{k}))_{\pi(l)} = k_{C,\pi(l)} = k_{C,m} \neq k_{C,l}$, we have $\mathbf{k}' \neq \mathbf{k}$. By Lemma 1(2), also $\mathbf{k}' \in E(\Gamma_{\mathcal{C}})$, a contradiction with $\#E(\Gamma_{\mathcal{C}}) = 1$.

2. Fix a permutation π of N such that $\pi(C') = C$. Then $C' = \pi^{<-1>}(C)$. Fix $j \in C'$. By part 1, the proof is complete if we show that $k'_{C';j} = k_{C;\pi(j)}$. Well, $k'_{C';j} = (J_{\mathcal{C}'}(\mathbf{k}'))_j = ((J_{\mathcal{C}'} \circ \mathcal{U}_\pi)(\mathbf{k}))_j = ((T_{\pi^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{k}))_j = (J_{\mathcal{C}}(\mathbf{k}))_{\pi(j)} = k_{C;\pi(j)}$.

3. Let π be a permutation of N such that $\pi(i) = j$, $\pi(C) = C'$, $\pi(C') = C$ and $\pi = \mathrm{Id}$ on $N \setminus (C \cup C')$. With $\mathcal{C}'$ as in Lemma 1, again $\mathcal{C} = \mathcal{C}'$. Lemma 1(2) implies $\mathcal{U}_\pi(\mathbf{k}) = \mathbf{k}$ and therefore $(T_{\pi^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{k}) = J_{\mathcal{C}}(\mathbf{k})$. With this, we obtain, $f_j(J_{\mathcal{C}}(\mathbf{k})) = f_{\pi(i)}(J_{\mathcal{C}}(\mathbf{k})) = f_i((T_{\pi^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{k})) = f_i(J_{\mathcal{C}}(\mathbf{k}))$.

4. Fix a permutation π of N such that $\mathcal{C}' = \{\pi^{<-1>}(A) \mid A \in \mathcal{C}\}$. By Lemma 1(2), $\mathcal{U}_\pi(\mathbf{k})$ is the unique Nash equilibrium of $\Gamma_{\mathcal{C}'}$.

5. As $\mathcal{C} \sim \mathcal{C}'$ and $\#C = \#C'$, there exists a permutation σ of N with $\mathcal{C}' = \{\sigma^{<-1>}(A) \mid A \in \mathcal{C}\}$ and $C' = \sigma^{-1}(C)$. Fix such an σ. By the proof of part 4, $\mathbf{k}' = \mathcal{U}_\sigma(\mathbf{k})$. Take an arbitrary $j \in C'$; note that $\sigma 9j) \in C$. We obtain, as desired $k'_{C';j} = (J_{\mathcal{C}'}(\mathbf{k}'))_j = ((J_{\mathcal{C}'} \circ \mathcal{U}_\sigma)(\mathbf{k}))_j = ((T_{\sigma^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{k}))_j = (J_{\mathcal{C}}(\mathbf{k}))_{\sigma(j)} = k_{C;\sigma(j)}$.

6. Choose σ as in the proof of part 5 such that $\sigma(j) = i$. Now $f_i(J_{\mathcal{C}}(\mathbf{k})) = f_{\sigma(j)}(J_{\mathcal{C}}(\mathbf{k})) = f_j((T_{\sigma^{<-1>}} \circ J_{\mathcal{C}})(\mathbf{k})) = f_j(J_{\mathcal{C}'} \circ \mathcal{U}_\sigma)(\mathbf{k})) = f_j(J_{\mathcal{C}'}(\mathbf{k}'))$. $\square$

From part 2 of this proposition we see that the actions of the individual players in a unique coalitional equilibrium of a symmetric game only depend on the size of the coalition and not on their composition.

7. Case of a Symmetric Physical Game

Theorem 2. *Consider a regular two-stage coalition formation game $(N; R; \Gamma)$ with a symmetric physical game Γ and a symmetric membership rule R. Let G be the effective game.*

1. *G is symmetric.*

Further, for all $\mathbf{m}, \mathbf{m}' \in M^n$ and $i, j \in N$, writing $\mathcal{C} = R(\mathbf{m})$ and $\mathcal{C}' = R(\mathbf{m}')$, if $\mathcal{C}$ and $\mathcal{C}'$ are congruent, then

2. $\#\mathcal{C}_{[i]} = \#\mathcal{C}'_{[j]} \ \Rightarrow\ [\hat{x}_i(\mathbf{m}) = \hat{x}_j(\mathbf{m}') \ \wedge\ g_i(\mathbf{m}) = g_j(\mathbf{m}')]$. $\diamond$

Proof. 1. In order to prove that G is symmetric, we fix a permutation π of N, $j \in N$ and $\mathbf{m} \in \mathbf{M}$. By (1), we have to prove that $g_{\pi(j)}(\mathbf{m}) = g_j(T_{\pi^{<-1>}}(\mathbf{m})))$. Writing $\mathcal{C} = R(\mathbf{m})$ and $\mathcal{C}' = R(T_{\pi^{<-1>}}(\mathbf{m}))$, Proposition 1 guarantees that $\mathcal{C}$ and $\mathcal{C}'$ are congruent; note that, by (5), $\mathcal{C}' = \{\pi^{<-1>}(C) \mid C \in \mathcal{C}\}$. By (7) we have to prove that $f_{\pi(j)}(J_{\mathcal{C}}(\mathbf{e}^{(\mathcal{C})})) = f_j(J_{\mathcal{C}'}(\mathbf{e}^{(\mathcal{C}')}))$. So, with $\mathbf{k} = \mathbf{e}^{(\mathcal{C})}$, $\mathbf{k}' = \mathbf{e}^{(\mathcal{C}')}$ and $i = \pi(j)$, we have to prove that $f_i(J_{\mathcal{C}}(\mathbf{k})) = f_j(J_{\mathcal{C}'}(\mathbf{k}'))$. Fix $C \in \mathcal{C}$ with $i \in \mathcal{C}$. Now $j = \pi^{<-1>}(i) \in \pi^{<-1>}(C) \in \mathcal{C}'$. Applying Proposition 2(6) completes the proof.

2. Suppose $\#\mathcal{C}_{[i]} = \#\mathcal{C}'_{[j]}$. Write $\mathbf{k} = \mathbf{e}^{(R(\mathbf{m}))}$ and $\mathbf{k}' = \mathbf{e}^{(R(\mathbf{m}'))}$. Then $\hat{\mathbf{x}}(\mathbf{m}) = J_{\mathcal{C}}(\mathbf{k})$ and $\hat{\mathbf{x}}(\mathbf{m}') = J_{\mathcal{C}'}(\mathbf{k}')$. Therefore $\hat{x}_i(\mathbf{m}) = k_{\mathcal{C}_{[i]};i}$ and $\hat{x}_j(\mathbf{m}') = k'_{\mathcal{C}'_{[j]};j}$. Proposition 2(5) guarantees $\hat{x}_i(\mathbf{m}) = \hat{x}_j(\mathbf{m}')$.

Further, we have $g_i(\mathbf{m}) = f_i(J_{\mathcal{C}}(\mathbf{k}))$ and $g_i(\mathbf{m}') = f_i(J_{\mathcal{C}'}(\mathbf{k}'))$. As $\#\mathcal{C}_{[i]} = \#\mathcal{C}'_{[j]}$, Proposition 2(6) guarantees $g_i(\mathbf{m}) = g_j(\mathbf{m}')$. $\square$

Some readers may find Theorem 2 intuitively clear. Yes it is. However, the literature uses reasoning based on this theorem, but does not provide a proof. Concerning this, it is good to note that there is almost no literature about general properties of symmetric games in strategic form and systematic studies are lacking; an exception is the recent [37]. Most text books on game theory even do not give a formal definition of a symmetric game. Additionally, there are various misunderstandings about the definition of such a game.[26]

8. Binary Action Games

In the next section we deal with a special type of two-stage coalition formation game: the cartel game. As the effective game of a cartel game is a binary action game, we first consider below, binary action games. The content of this section is based on [19,38]. For various results also, a proof will be given.

8.1. Internal and External Stability

In the rest of this section we consider, if not stated otherwise, a **binary action game**; i.e., a game in strategic form with player set $N = \{1, 2, \ldots, n\}$ where each player has $M = \{0, 1\}$ as action set. We denote the game by G and the payoff function of player i by g_i, so $g_i : M^n \to \mathbb{R}$.

Again as in Example 1, for $\mathbf{m} \in M^n$, let

$$\hat{S}(\mathbf{m}) = \{i \in N \mid m_i = 1\}, \ \ \hat{S}_\star(\mathbf{m}) = \{i \in N \mid m_i = 0\}.$$

We refer to the players in $\hat{S}(\mathbf{m})$ as **cooperators** and to the players in $\hat{S}_\star(\mathbf{m})$ as **non-cooperators**.

For $i \in N$ define $T_i : M^n \to M^n$ as follows: $T_i(\mathbf{m})$ is the action profile obtained from $\mathbf{m} = (m_1, \ldots, m_n)$ when player i changes his action (i.e., replaces m_i by $1 - m_i$). Let $g'_i(\mathbf{m})$ be the marginal payoff of i at $\mathbf{m}$; i.e.,

$$g'_i(\mathbf{m}) = g_i(T_i(\mathbf{m})) - g_i(\mathbf{m}).$$

Note that $g'_i(T_i(\mathbf{m})) = -g'_i(\mathbf{m})$ and that $\mathbf{m}$ is a Nash equilibrium if and only if $g'_i(\mathbf{m}) \leq 0$ for all $i \in N$.

An interesting notion for binary action games is that of "semi-strict Nash equilibrium":

Definition 6. *An action profile* $\mathbf{m}$ *is*

1. ***Weakly internally stable*** *if* $g'_i(\mathbf{m}) \leq 0$ *for each cooperator* i.
2. ***Strictly externally stable***, $g'_i(\mathbf{m}) < 0$ *for each non-cooperator* i.
3. *A* ***semi-strict*** *Nash equilibrium if* $\mathbf{m}$ *is weakly internally stable and strictly externally stable.* ⋄

These notions of internal and external stability were essentially introduced in [3] and differ entirely from those in other types of games.

We denote the set of semi-strict Nash equilibria of G by

$$E_{ss}(G)$$

and the set of weakly internally stable action profiles by

$$\mathcal{I}(\hat{S}).$$

Note that $\mathcal{I}(\hat{S}) \neq \emptyset$ as $\mathbf{0} \in \mathcal{I}(\hat{S})$. Let

$$U_{\hat{S}}(\mathbf{m}) := \{i \in \hat{S}_\star(\mathbf{m}) \mid g'_i(\mathbf{m}) \geq 0\}.$$

Thus, $U_{\hat{S}}(\mathbf{m})$ is the set of players that violate the strict inequality for strict external stability; $\mathbf{m}$ is strictly externally stable if and only if $U_{\hat{S}}(\mathbf{m}) = \emptyset$.

---

[26] For example, the stone-paper-scissors (bi-matrix) game is not symmetric.

The next result is at the base for various results about binary action games.

Proposition 3. *Consider an action profile* $\mathbf{m}$.

1. $\mathbf{m} \in \mathcal{I}(\hat{S}) \Rightarrow U_{\hat{S}}(T_i(\mathbf{m})) \neq \emptyset \; (i \in \hat{S}(\mathbf{m}))$.
2. $U_{\hat{S}}(\mathbf{m}) = \emptyset \Rightarrow T_i(\mathbf{m}) \notin \mathcal{I}(\hat{S}) \; (i \in \hat{S}_\star(\mathbf{m}))$.
3. *If* $\mathbf{m}$ *is a semi-strict Nash equilibrium, then for every* $i \in N$ *it holds that* $T_i(\mathbf{m})$ *is not a semi-strict Nash equilibrium.* ⋄

Proof. See Proposition 3 in [38] (see also Lemma 1 in [19]). □

Proposition 3(1) states that if an action profile $\mathbf{m}$ is weakly internally stable, then for every cooperator i the action profile $T_i(\mathbf{m})$ is strictly externally unstable. Its part 2 indicates that if an action profile $\mathbf{m}$ is strictly externally stable, then for every non-cooperator i the action profile $T_i(\mathbf{m})$ is weakly internally unstable. Part 3 shows a typical property of semi-strict Nash equilibria: if $\mathbf{m}$ is such an equilibrium, then if one player changes his action, the resulting action profile is no longer a semi-strict Nash equilibrium.

8.2. Deviation Property D_1

In [38] three so-called deviation properties (i.e., D_1, D_2 and D_3) for binary action games are defined. In the present article we only consider the following one:

Definition 7. *A binary action game has the* ***deviation property*** D_1 *if* $g'_j(\mathbf{m}) \leq 0 \Leftrightarrow g'_i(\mathbf{m}) \leq 0$ *for each action profile* $\mathbf{m}$ *and cooperators* i, j. ⋄

Terminology: a finite sequence $\mathbf{m}^{(0)}, \mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(k)}$ of action profiles is an **enlargement sequence** if $\hat{S}(\mathbf{m}^{(0)}) \subset \hat{S}(\mathbf{m}^{(1)}) \subset \cdots \subset \hat{S}(\mathbf{m}^{(k)})$. An enlargement sequence $\mathbf{m}^{(0)}, \mathbf{m}^{(1)}, \cdots,$ $\mathbf{m}^{(k)}$ is called **elementary** if $\#\hat{S}(\mathbf{m}^{(l+1)}) = \#\hat{S}(\mathbf{m}^{(l)}) + 1$ for all l.

Theorem 3. *Consider a binary action game with the deviation property* D_1.

1. *Let* $\mathbf{m} \in M^n$. *Then* $i \in U_{\hat{S}}(\mathbf{m}) \Rightarrow T_i(\mathbf{m}) \in \mathcal{I}(\hat{S})$.
2. *For each weakly internally stable action profile* $\mathbf{m}^{(0)}$ *there exists an elementary enlargement sequence* $\mathbf{m}^{(0)}, \mathbf{m}^{(1)}, \ldots, \mathbf{m}^{(k)}$ *of weakly internally stable action profiles where* $\mathbf{m}^{(k)}$ *is a semi-strict Nash equilibrium.*
3. *The game has a semi-strict Nash equilibrium.* ⋄

Proof. See Proposition 5 and Theorem 1 in [38]. □

Theorem 3(2) implies that each binary action game with the deviation property D_1 has a semi-strict Nash equilibrium: indeed, take $\mathbf{m}^{(0)} = \mathbf{0}$.

Definition 8. *A binary action game:*

1. *Is **super-additive** if for each action profile* $\mathbf{m}$ *and* $i \in \hat{S}_\star(\mathbf{m})$

$$\sum_{l \in \hat{S}(T_i(\mathbf{m}))} g_l(T_i(\mathbf{m})) \geq \sum_{l \in \hat{S}(\mathbf{m})} g_l(\mathbf{m}) + g_i(\mathbf{m});$$

2. *Has **weak negative spillovers** if for each action profile* $\mathbf{m}$ *and* $i, k \in \hat{S}_\star(\mathbf{m})$ *with* $i \neq k$

$$g_i(T_k(\mathbf{m})) \leq g_i(\mathbf{m}). \ \diamond$$

Super-additivity thus means that: given an action profile, the payoff of a meta player consisting of the cooperators together with a non-cooperator is at least the sum of the payoffs of the cooperators and this non-cooperator. Weak negative spillovers, in contrast, indicate that the payoff of a non-cooperator decreases or stays the same if another non-cooperator becomes a cooperator.

Proposition 4. *Suppose G has the deviation property* D_1*, is super-additive and has weak negative spillovers.*

1. *For all* $\mathbf{m} \in \mathcal{I}(\hat{S})$ *and* $j \in S_\star(\mathbf{m})$*, it holds that* $T_j(\mathbf{m}) \in \mathcal{I}(\hat{S})$*.*
2. *Each action profile is weakly internally stable.*
3. $E_{ss}(G) = \{(1, 1, \ldots, 1)\}$*.* $\diamond$

Proof. 1. Let $\mathbf{p} = T_j(\mathbf{m})$. Note that $\mathbf{m} = T_j(\mathbf{p})$ and $\hat{S}(\mathbf{p}) = \hat{S}(\mathbf{m}) \cup \{j\}$. We obtain

$$\sum_{l \in \hat{S}(\mathbf{p})} g_l(T_l(\mathbf{p})) = g_j(\mathbf{m}) + \sum_{l \in \hat{S}(\mathbf{m})} g_l(T_j T_l(\mathbf{m}))$$

$$\leq g_j(\mathbf{m}) + \sum_{l \in \hat{S}(\mathbf{m})} g_l(T_l(\mathbf{m})) \leq g_j(\mathbf{m}) + \sum_{l \in \hat{S}(\mathbf{m})} g_l(\mathbf{m})$$

$$\leq \sum_{l \in \hat{S}(T_j(\mathbf{m}))} g_l(T_j(\mathbf{m})) = \sum_{l \in \hat{S}(\mathbf{p})} g_l(\mathbf{p}).$$

Here the first inequality holds by weak negative spillovers (noting that $j, l \in \hat{S}_\star(T_l(\mathbf{m}))$ and $l \neq j$), the second by $\mathbf{m} \in \mathcal{I}(\hat{S})$, and the third by super-additivity. Thus $\sum_{l \in \hat{S}(\mathbf{p})} g'_l(\mathbf{p}) \leq 0$. The deviation property D_1 implies $\mathbf{p} \in \mathcal{I}(\hat{S})$.

2. By part 1, as $\mathbf{0} \in \mathcal{I}(\hat{S})$.

3. As $\mathbf{1} = (1, \ldots, 1)$ is strictly externally stable, part 2 implies that $\mathbf{1} \in E_{ss}(G)$. Next, by contradiction we prove $E_{ss}(G) \subseteq \{\mathbf{1}\}$. So suppose $\mathbf{m} \in E_{ss}(G)$ with $\mathbf{m} \neq \mathbf{1}$. As $\mathbf{m} \in E_{ss}(G)$, we have $U_S(\mathbf{m}) = \emptyset$. As $\mathbf{m} \neq \mathbf{1}$, we can fix $i \in \hat{S}_\star(\mathbf{m})$. By Proposition 3(2), $T_i(\mathbf{m}) \notin \mathcal{I}(\hat{S})$, a contradiction with part 2. □

8.3. Symmetric Binary Action Games

For a symmetric game in strategic form with a unique Nash equilibrium, this equilibrium may not be strict as the following bi-matrix game shows:

$$\begin{pmatrix} 1;1 & 1;0 & 0;1 \\ 0;1 & 0;0 & 1;-1 \\ 1;0 & -1;1 & 0;0 \end{pmatrix}.$$

But, as part 2 of the following proposition shows, it is strict if the game is a binary action game.

Proposition 5. *Consider a symmetric binary action game G.*

1. *G has the deviation property D_1 (and thus, a semi-strict Nash equilibrium).*
2. *If G has a unique Nash equilibrium, then this equilibrium is strict.* $\diamond$

Proof. See Theorem 5 and Corollary 1 in [38]. $\square$

Lemma 2. *Consider a symmetric binary action game; let $S = \hat{S}$ or $S = \hat{S}_\star$. Then for all action profiles $\mathbf{m}, \mathbf{m}'$ and players i, j with $\#S(\mathbf{m}) = \#S(\mathbf{m}')$, $i \in S(\mathbf{m})$ and $j \in S(\mathbf{m}')$, it holds that $g_i(\mathbf{m}) = g_j(\mathbf{m}')$ and $g_i(T_i(\mathbf{m})) = g_j(T_j(\mathbf{m}'))$.* $\diamond$

Proof. By Lemma 2 in [38]. $\square$

For a symmetric binary action game, define[27]

$$\begin{aligned} w_0(s) &= g_1(0, \mathbf{1}_s, \mathbf{0}) \ (0 \le s \le n-1), \qquad (8) \\ w_1(s) &= g_1(\mathbf{1}_s, \mathbf{0}) \ (1 \le s \le n), \qquad (9) \end{aligned}$$

and the function $t : \{0, 1, \ldots, n+1\} \to \mathbb{R} \cup \{-\infty, +\infty\}$ by

$$t(s) := \begin{cases} +\infty \text{ if } s = 0, \\ g_1(\mathbf{1}_s, \mathbf{0}) - g_1(0, \mathbf{1}_{s-1}, \mathbf{0}) = w_1(s) - w_0(s-1) \text{ if } 1 \le s \le n, \\ -\infty \text{ if } s = n+1. \end{cases} \qquad (10)$$

We refer to t as the **stability function**.

Theorem 4. *Suppose G is symmetric. Let $\mathbf{m} \in M^n$; write $s = \#\hat{S}(\mathbf{m})$.*

1. *$\mathbf{m}$ is a Nash equilibrium if and only if, $t(s+1) \le 0 \le t(s)$.*
2. *$\mathbf{m}$ is a semi-strict Nash equilibrium if and only if, $t(s+1) < 0 \le t(s)$.* $\diamond$

Proof. By Lemma 2, writing $s = \#\hat{S}(\mathbf{m})$, for $i \in \hat{S}(\mathbf{m})$ that $g_i(\mathbf{m}) = g_1(\mathbf{1}_s, \mathbf{0})$, $g_i(T_i(\mathbf{m})) = g_1(0, \mathbf{1}_{s-1}, \mathbf{0})$, and therefore, $g'_i(\mathbf{m}) = g_1(0, \mathbf{1}_{s-1}, \mathbf{0}) - g_1(\mathbf{1}_s, \mathbf{0}) = g'_1(\mathbf{1}_s, \mathbf{0})$. This in turn implies for $i \in \hat{S}_\star(\mathbf{m})$ that $g'_i(\mathbf{m}) = -g'_i(T_i(\mathbf{m})) = -g_1(0, \mathbf{1}_s, 0) + g_1(\mathbf{1}_{s+1}, \mathbf{0}) = -g'_1(\mathbf{1}_{s+1}, \mathbf{0})$. Having seen this, observe for all $\mathbf{m} \in M^n$ and $i \in N$

$$g'_i(\mathbf{m}) = \begin{cases} -t(\#\hat{S}(\mathbf{m})) \text{ if } i \in \hat{S}(\mathbf{m}), \\ t(\#\hat{S}(\mathbf{m}) + 1) \text{ if } i \in \hat{S}_\star(\mathbf{m}). \end{cases} \qquad (11)$$

2. We use (11). Definition 6(a) together with $t(0) = \infty$ to imply that $\mathbf{m}$ with $\hat{S}(\mathbf{m}) \ge 1$ is weakly internally stable if and only if $t(s) \ge 0$. Definition 6(b) together with $t(n+1) = -\infty$ implies that $\mathbf{m}$ is strictly externally stable if and only if, $t(s+1) < 0$. Thus, the desired results follows.

1. In the same way as 2. $\square$

Proposition 6. *If G is symmetric, then G is an aggregative game.* $\diamond$

Proof. Note that, for every $i \in N$, the set $T_i = \sum_{l \ne i} M_l$ equals $\{0, \ldots, n-1\}$. For $z \in T_i$, define $\tilde{g}_i^{(z)} : M_i \to \mathbb{R}$ by $\tilde{g}_i^{(z)}(m_i) := g_i(m_i, \mathbf{1}_z, 0, \ldots, 0)$. Lemma 2 (with $j = i$) implies, as desired by (2), that $g_i(\mathbf{m}) = g_i(m_i; \mathbf{m}_{\hat{\imath}}) = g_i(m_i; \mathbf{1}_{\sum_{l \ne i} m_l}, 0, \ldots, 0) = \tilde{g}_i^{(\sum_{l \ne i} m_l)}(m_i)$. $\square$

[27] $\mathbf{1}_s = (1, 1, \ldots, 1)$ with s times a 1.

8.4. Potentials

The concept of potential game originated in [39]. As far as we know they still did not find their way into the two-stage game approach to coalition formation. Roughly said, a game is said to be a potential game if the incentive of all players to change their strategy can be expressed using a single global function called the potential function. Nowadays, there exist various types of potential games. Potential games have interesting properties: a maximiser of the potential function is a pure Nash equilibrium, and a potential provides insight into (finite-time) convergence of an iterated game towards a Nash equilibrium and is helpful for performing comparative statics in case the game also has an aggregative structure ([23]). For an extensive treatment of potential games, we refer the reader to [40].

Theorem 5. *A binary action game with the deviation property D_1 is a generalised, ordinal potential game. If in addition for each player his best-reply correspondence is single-valued, then the game is an ordinal potential game.* $\diamond$

Proof. See Theorem 3 in [38]. $\square$

By Proposition 5 and Theorem 3, each symmetric binary action game is a generalised ordinal potential game. But we can do better, as the next theorem shows.

Theorem 6. *Consider a symmetric binary action game G.*

1. *G is an exact potential game. Even, the function $P : M^n \to \mathbb{R}$ defined by $P(\mathbf{m}) := \sum_{l=1}^{\#\hat{S}(\mathbf{m})} t(l)$ is an exact potential.*
2. *G is a congestion game.* $\diamond$

Proof. 1. By Theorems 4 and 5 in [38].
2. By part 1, as each finite exact potential game is a congestion game ([39]). $\square$

9. Cartel Games

9.1. Notion

An important example of a membership rule is the cartel membership rule, which is defined as follows.

Definition 9. *The **cartel membership rule** is the membership rule $R_c : \mathbf{M} \to \mathscr{C}_N$ defined by $M_i = M = \{0, 1\}$ $(i \in N)$ and*[28]

$$R_c(\mathbf{m}) := \begin{cases} \{\hat{S}(\mathbf{m})\} \cup \cup_{i \in \hat{S}_\star(\mathbf{m})} \{\{i\}\} \text{ if } \mathbf{m} \neq \mathbf{0}, \\ \{\{1\}, \ldots, \{n\}\} \text{ if } \mathbf{m} = \mathbf{0}. \end{cases}$$

Here $\hat{S}, \hat{S}_\star : \mathbf{M} \to \mathbb{R}$ are defined by

$$\hat{S}(\mathbf{m}) := \{i \in N \mid m_i = 1\} \text{ and } \hat{S}_\star(\mathbf{m}) := \{i \in N \mid m_i = 0\}. \; \diamond$$

We refer to the coalition structures in $R_c(\mathbf{M})$ as **cartel coalition structures**.

Note that

$$\#\hat{S}(\mathbf{m}) \leq 1 \;\Rightarrow\; R_c(\mathbf{m}) = \{\{1\}, \ldots, \{n\}\} \tag{12}$$

and that $R_c : \mathbf{M} \setminus \{\mathbf{0}\} \to \mathscr{C}_N$ is injective.

---

[28] Note that we distinguish between two cases as $\{\emptyset\} \cup \{\{1\}, \ldots, \{n\}\} \neq \{\{1\}, \ldots, \{n\}\}$.

Definition 10. *A **cartel game** is a two-stage coalition formation game* $(N; R; \Gamma)$ *where the membership rule* R *is the cartel membership rule* R_c. $\diamond$

If $\#\hat{S}(\mathbf{m}) = \#\hat{S}(\mathbf{m}')$, then there exists a permutation π of N with $\mathbf{m}' = T_\pi(\mathbf{m})$. Therefore, Proposition 1 implies

$$\#\hat{S}(\mathbf{m}) = \#\hat{S}(\mathbf{m}') \;\Rightarrow\; R_c(\mathbf{m}) \sim R_c(\mathbf{m}').$$

The cartel membership rule can be generalised as follows by the following membership rule $R_{cc} : \mathbf{M} \to \mathscr{C}_N$ defined by $M_i = \{0, 1, \ldots, p\}$ $(i \in N)$ and

$$R_{cc}(\mathbf{m}) := \cup_{k=1,\ldots,p \text{ with } \hat{S}_k(\mathbf{m}) \neq \emptyset} \{\hat{S}_k(\mathbf{m})\} \cup \cup_{i \in \hat{S}_0(\mathbf{m})} \{\{i\}\}$$

where $p \geq 1$ and $\hat{S}_k(\mathbf{m}) = \{i \in N \mid m_i = k\}$.[29]

Proposition 7. *The membership rule* R_{cc} *is symmetric.* $\diamond$

Proof. Let π be a permutation of N and $\mathbf{m} \in \mathbf{M}$.

First note that $\hat{S}_k(T_\pi(\mathbf{m})) = \pi^{<-1>}(\hat{S}_k(\mathbf{m}))$. Indeed: $i \in \hat{S}_k(T_\pi(\mathbf{m})) \Leftrightarrow k = (T_\pi(\mathbf{m}))_i = m_{\pi(i)} \Leftrightarrow \pi(i) \in \hat{S}_k(\mathbf{m}) \Leftrightarrow k \in \pi^{<-1>}(\hat{S}_k(\mathbf{m}))$. This implies, as desired, $R_{cc}(T_\pi(\mathbf{m})) = \{\hat{S}_1(T_\pi(\mathbf{m})), \ldots, \hat{S}_s(T_\pi(\mathbf{m}))\} \cup \cup_{i \in \hat{S}_0(T_\pi(\mathbf{m}))} \{\{i\}\} = \{\pi^{<-1>}(\hat{S}_1(\mathbf{m})), \ldots, \pi^{<-1>}(\hat{S}_s(\mathbf{m}))\} \cup \cup_{i \in \pi^{<-1>}(\hat{S}_0(\mathbf{m}))} \{\{i\}\} = \{\pi^{<-1>}(C) \mid C \in R_{cc}(\mathbf{m})\}$. $\square$

9.2. Equilibrium Coalition Structures

In this subsection we consider a regular cartel formation game $(N; R_c; \Gamma)$. Again, G denotes its effective game.

As all membership action sets M_i equal $M := \{0, 1\}$, G is a binary action game. A useful observation is that G has, by (6) and (12), the following property:

$$g_i(\mathbf{m}) = g_i(\mathbf{0}) \text{ for all } i \in N \text{ and } \mathbf{m} \in M^n \text{ with } \#\hat{S}(\mathbf{m}) = 1. \tag{13}$$

(13) implies that for each player i the best reply to the strategy profile $\mathbf{0}$ of its opponents equals $\{0, 1\}$. Thus, no best-reply correspondence is single-valued.

With the stability function t as given by (10), (13) implies for a symmetric G

$$t(1) = 0. \tag{14}$$

For the effective game G the literature not only considers its Nash equilibria set $E(G)$ but also its semi-strict Nash equilibrium set

$$E_{ss}(G).$$

As $E_{ss}(G) \subseteq E(G)$, we then have for every $\mathbf{m} \in E_{ss}(G)$ a semi-strict equilibrium coalition structure $R_c(\mathbf{m})$. Property (13) implies

$$\#\hat{S}(\mathbf{m}) \leq 1 \;\Rightarrow\; \mathbf{m} \in \mathcal{I}(\hat{S}), \;\; \mathbf{0} \in E(G), \;\; \mathbf{0} \notin E_{ss}(G). \tag{15}$$

As $R_c(\mathbf{0}) = \{\{1\}, \ldots, \{n\}\}$, we obtain:

---

29 So choosing an action not equal to 0 means that a player is willing to cooperate. Concerning real world interpretations: each of the actions $1, \ldots, p$ may represent different "circumstances" for cooperation.

Proposition 8. *The ineffective coalition structure* $\{\{1\},\ldots,\{n\}\}$ *is for each cartel game an equilibrium coalition structure.* $\diamond$

However, (15) does not exclude that the ineffective coalition structure is a semi-strict equilibrium cartel coalition structure: indeed, if $\mathbf{m}$ is a semi-strict Nash equilibrium with $\#\hat{S}(\mathbf{m}) = 1$, then the ineffective coalition structure is a semi-strict equilibrium cartel coalition structure.

Example 3. *Consider again, Example 1. There we found that for* $n \geq 3$ *the equilibrium coalition structures are exactly the cartel coalition structures with zero, two, or three cooperators and that for* $n = 2$ *the equilibrium coalition structures are exactly the cartel coalition structures with zero or two cooperators. With Theorem 6(2), we see that the semi-strict equilibrium coalition structures are for* $n \geq 3$ *exactly the cartel coalition structures with* 3 *cooperators and for* $n = 2$ *exactly the cartel coalition structures with* 2 *cooperators.* $\diamond$

Proposition 9. *If* G *is a super-additive cartel game and has the deviation property* D_1*, then each action profile* $\mathbf{p}$ *with* $\#\hat{S}(\mathbf{p}) = 2$ *is weakly internally stable.* $\diamond$

Proof. We may suppose that $\hat{S}(\mathbf{p}) = \{1,2\}$. So $\hat{S}_\star(\mathbf{p}) = N \setminus \{1,2\}$. We have to prove that $g_1(\mathbf{p}) \geq g_1(T_1(\mathbf{p})) \wedge g_2(\mathbf{p}) \geq g_2(T_2(\mathbf{p}))$. As the game is a cartel game, this becomes $g_1(\mathbf{p}) \geq g_1(\mathbf{0}) \wedge g_2(\mathbf{p}) \geq g_2(\mathbf{0})$. As the game has the deviation property D_1, we have $g_1(\mathbf{p}) \geq g_1(T_1(\mathbf{p})) \Leftrightarrow g_2(\mathbf{p}) \geq g_2(T_2(\mathbf{p}))$. Thus, the proof is complete if we can show that

$$g_1(\mathbf{p}) + g_2(\mathbf{p}) \geq g_1(\mathbf{0}) + g_2(\mathbf{0}).$$

Applying Definition 8 with $\mathbf{m} = (1,0,\ldots,0)$ and $i = 2$ (and noting that $T_i(\mathbf{m}) = \mathbf{p}$) gives the result. □

Theorem 7. *Consider a regular cartel game* $(N; R_c; \Gamma)$ *with a symmetric physical game* Γ*. Let* G *be its effective game.*

1. *G is a symmetric game, an aggregative game, an exact potential game, and a congestion game and has the deviation property D_1. The function $P(\mathbf{m}) = \sum_{l=1}^{\#\hat{S}(\mathbf{m})} t(l)$ is an exact potential.*
2. *G has a semi-strict Nash equilibrium, and therefore, a semi-strict equilibrium cartel coalition structure.*
3. *If G has a unique Nash equilibrium, then $\mathbf{1}$ is this equilibrium, this equilibrium is strict, $\{N\}$ is a semi-strict equilibrium cartel coalition structure and there is no other equilibrium cartel coalition structure.*
4. *If there exists a membership action profile $\mathbf{m}$ with two cooperators such that $g_i(\mathbf{m}) \geq g_i(\mathbf{0})$ for some cooperator i, then there exists an effective semi-strict equilibrium cartel coalition structure.* $\diamond$

Proof. 1. Noting that by Proposition 7 the membership rule R_c is symmetric, Theorem 2(1) applies and guarantees that G is symmetric. Next, apply Theorem 6, Proposition 5(1) and Proposition 6.

2. As, by part 1, G is a symmetric binary action game, it has by Proposition 5(1) a semi-strict Nash equilibrium.

3. Let $\mathbf{m}$ be this equilibrium. By Proposition 5(2), $\mathbf{m}$ is strict. As G is symmetric, $\mathbf{m}$ is symmetric. As, by (15), $\mathbf{0}$ is not a semi-strict Nash equilibrium, $\mathbf{m} = \mathbf{1}$ follows. Thus $R_c(\mathbf{m}) = \{N\}$ is a semi-strict equilibrium cartel coalition structure. Of course, there is no other equilibrium cartel coalition structure.

4. As $\#\hat{S}(\mathbf{m}) = 2$, we have $\#\hat{S}(T_i(\mathbf{m})) = 1$, and therefore, by (13), $g_i(\mathbf{m}) \geq g_i(\mathbf{0}) = g_i(T_i(\mathbf{m}))$. So $g_i(\mathbf{m}) \geq g_i(T_i(\mathbf{m}))$. As by part 1 the game has the deviation property D_1 it follows that $g_k(\mathbf{m}) \geq g_k(T_k(\mathbf{m}))$ for all $k \in \hat{S}(\mathbf{m})$. Thus $\mathbf{m}$ is weakly internally stable. Theorem 3(2) implies that there exists a semi-strict Nash equilibrium $\mathbf{a}$ with $\#\hat{S}(\mathbf{a}) \geq \#\hat{S}(\mathbf{m}) = 2$. Thus, $R_c(\mathbf{a})$ is an effective semi-strict equilibrium cartel coalition structure. □

The nice thing in Theorem 7(1,2,3) is that its results almost solely depend on the qualitative properties of the primitives of the physical game, namely being symmetric. We write "almost" as there is the additional assumption that the cartel game is regular.

It is well-known that each maximiser of a finite exact potential game is a Nash equilibrium. Let us now, for Example 1 identify the Nash equilibria obtained by maximising the potential P in Theorem 7(5):

Example 4. *Consider again, Example 1. There we obtain for the exact potential P in Theorem 7(1), writing $s = \#\hat{S}(\mathbf{m})$,*

$$P(\mathbf{m}) = \sum_{l=1}^{s} t(l) = \sum_{l=1}^{s}(-\frac{1}{2}l^2 + 2l - \frac{3}{2})\frac{b^2}{a} = (-\frac{1}{6}s^3 + \frac{3}{4}s^2 - \frac{7}{12}s)\frac{b^2}{a}.$$

The maximisers of P are the strategy profiles with 2 or 3 cooperators (i.e., $s = 2$ or $s = 3$). $\diamond$

Proposition 10. *Consider a regular cartel game $(N; R_c; \Gamma)$. Suppose that the effective game G has deviation property D_1.*

1. *G is a generalised ordinal potential game.*
2. *G has a semi-strict Nash equilibrium, and therefore, a semi-strict equilibrium cartel coalition structure.*
3. *If there exists a membership action profile $\mathbf{m}$ with $\#\hat{S}(\mathbf{m}) \geq 2$ such that $g_i(\mathbf{m}) \geq g_i(\mathbf{0})$ for some player $i \in \hat{S}(\mathbf{m})$, then there exists an effective semi-strict equilibrium coalition structure.* $\diamond$

Proof. 1. By Theorem 3 in [38].
2. By Theorem 3(2).
3. Exactly the same (but omitting "by part 1" there) as Theorem 7(4). $\square$

10. Further Examples

To illustrate and compLemmant our general results from the previous section, we now determine the (semi-strict) equilibrium cartel coalition structures for some concrete cartel formation games with a Cournot-like game as physical game.

Example 5. *Consider the cartel game with a physical game the Cournot-like public good game with common action set $X = \mathbb{R}_+$ and with payoff functions*

$$f_i(\mathbf{x}) = -c_i(x_i) + b_i \sum_{l=1}^{n} x_l$$

where c_i is continuously differentiable, strictly convex, and strictly increasing with $\lim_{x_i \to \infty} c_i'(x_i) = \infty$, $b_i > 0$ and $b_i \geq c_i'(0)$. With this example we provide a more general variant of Example 1: it will relax the assumption of a quadratic (cost) function c_i.

Note that $c_i : \mathbb{R}_+ \to [c_i(0), \infty[$, $c_i' : \mathbb{R}_+ \to [c_i'(0), \infty[$, and $c_i'^{<-1>} : [c_i'(0), \infty[\to \mathbb{R}_+$. For $A \subseteq N$ we denote $b(A) := \sum_{l \in A} b_l$.

The payoff function $\hat{P}_C$ of meta player $C \in R_c(\mathbf{m})$ in terms of the action profile $\mathbf{x} \in \mathbf{X} = X^n$ equals

$$\hat{P}_C(\mathbf{x}) = \begin{cases} -c_i(x_i) + b_i x_i + b_i \sum_{l \neq i} x_l \text{ if } C = \{i\} \text{ with } i \in \hat{S}_\star(\mathbf{m}), \\ \sum_{i \in C}(-c_i(x_i) + b(C)x_i) + b(C)\sum_{l \notin C} x_l \text{ if } C = \hat{S}(\mathbf{m}), \end{cases}$$

the effective physical actions are

$$\hat{x}_i(\mathbf{m}) = \begin{cases} \underline{x}_i \ (i \in \hat{S}_\star(\mathbf{m})), \\ \overline{x}_i(\mathbf{m}) \ (i \in \hat{S}(\mathbf{m})), \end{cases}$$

where $\underline{x}_i = c_i'^{<-1>}(b_i)$ and $\overline{x}_i(\mathbf{m}) = c_i'^{<-1>}(b(\hat{S}(\mathbf{m})))$ and the effective payoffs are

$$g_i(\mathbf{m}) = \begin{cases} -c_i(\underline{x}_i) + b_i(\sum_{l \in \hat{S}(\mathbf{m})} \overline{x}_l(\mathbf{m}) + \sum_{l \in \hat{S}_\star(\mathbf{m})} \underline{x}_l) \text{ if } m_i = 0, \\ -c_i(\overline{x}_i(\mathbf{m})) + b_i(\sum_{l \in \hat{S}(\mathbf{m})} \overline{x}_l(\mathbf{m}) + \sum_{l \in \hat{S}_\star(\mathbf{m})} \underline{x}_l) \text{ if } m_i = 1. \end{cases}$$

Although the further analysis also is analogous to that in Example 1, we shall proceed from here on by referring to the theory developed in the meanwhile.

In order to simplify, further suppose that $c_1 = \cdots = c_n =: c$ *and* $b_1 = \cdots = b_n =: b$*. As the physical game is symmetric, the effective game G is so too by Theorem 2.*[30]

Now let us determine the stability function t. We already know that

$$t(0) = +\infty,\ t(1) = 0,\ t(n+1) = -\infty.$$

For the functions w_0 *and* w_1 *in (8) and (9) writing* $s = \#\hat{S}(\mathbf{m})$ *and* $\overline{x}(s) = c'^{<-1>}(sb)$ $(s \geq 1)$

$$w_0(s) = g_1(0, \mathbf{1}_s, \mathbf{0}) = -c(\overline{x}(1)) + b((n-s)\overline{x}(1) + s\overline{x}(s))\ (1 \leq s \leq n-1),$$

$$w_1(s) = g_1(\mathbf{1}_s, \mathbf{0}) = -c(\overline{x}(s)) + b((n-s)\overline{x}(1) + s\overline{x}(s))\ (1 \leq s \leq n).$$

So for $s \geq 2$*, we obtain* $t(s) = w_1(s) - w_0(s-1) = -c(\overline{x}(s)) + b((n-s)\overline{x}(1) + s\overline{x}(s)) - (-c(\overline{x}(1)) + b((n-(s-1))\overline{x}(1) + s\overline{x}(s-1))) = -c(\overline{x}(s)) + c(\overline{x}(1)) + bs\overline{x}(s) - b\overline{x}(1) - b(s-1)\overline{x}(s-1) = -c(\overline{x}(s)) + c(\overline{x}(1)) + b\Big(s(\overline{x}(s) - \overline{x}(s-1)) + \overline{x}(s-1) - \overline{x}(1)\Big)$*. Thus for* $2 \leq s \leq n$ *we get the formula*

$$t(s) = c(\overline{x}(1)) - c(\overline{x}(s)) + b\Big(s(\overline{x}(s) - \overline{x}(s-1)) + \overline{x}(s-1) - \overline{x}(1)\Big). \tag{16}$$

(For $n \geq 3$*) this formula implies* $t(2) = c(\overline{x}(1)) - c(\overline{x}(2)) + b(2\overline{x}(2) - 2\overline{x}(1))$ *and* $t(3) = c(\overline{x}(1)) - c(\overline{x}(3)) + b(3\overline{x}(3) - 2\overline{x}(2) - \overline{x}(1))$*.*

Further suppose $c(x) = \frac{a}{p}x^p$ *with* $a > 0$ *and* $p > 1$*. Then* $c'^{<-1>}(y) = (y/a)^{1/(p-1)}$*,* $\overline{x}(s) = s^{1/(p-1)}\overline{x}(1)$*,* $\overline{x}(1) = (b/a)^{1/(p-1)}$ *and* $c(\overline{x}(s)) = s^{p/(p-1)}\frac{b}{p}\overline{x}(1))$*. Therefore for* $2 \leq s \leq n$ *we find*

$$t(s) = b\overline{x}(1)\Big(\frac{1}{p}(1 - s^{p/(p-1)}) + s^{p/(p-1)} - 1 - (s-1)^{p/(p-1)}\Big).$$

Let us consider various cases. First suppose $n = 2$*. We have* $t(0) = \infty, t(1) = 0$*,* $t(3) = -\infty$ *and* $t(2) = \overline{x}(1)b(-\frac{3}{2} + 2^{\frac{1}{p-1}})$*. This implies* $t(2) \leq 0 \Leftrightarrow p \geq \log_{3/2} 3$ *and* $t(2) = 0 \Leftrightarrow p = \log_{3/2} 3$*. It follows that for* $n = 2$ *the equilibrium cartel coalition structures are exactly the cartel coalition structures with* 0 *and with* 1 *cooperators if* $p > \log_{3/2} 3$*, the cartel coalition structures with* $0, 1$ *and* 2 *cooperators if* $p = \log_{3/2} 3$ *and the cartel coalition structures with* 0 *or* 2 *cooperators if* $p < \log_{3/2} 3$*. Also for* $n = 2$ *the semi-strict equilibrium cartel coalition structures are exactly the cartel coalition structures with* 1 *cooperator if* $p > \log_{3/2} 3$ *and the cartel coalition structures with* 2 *cooperators if* $p \leq \log_{3/2} 3$*.*

A further analysis (with numerical simulations) shows that for $n \geq 3$ *and p an integer with* $p \geq 3$ *the equilibrium cartel coalition structures are exactly the cartel coalition structures with* 0 *and with* 2 *cooperators.* ⋄

Example 6. *Consider the cartel game with as physical game the Cournot-like public bad game with common action set* $X = [0, 1]$ *and with payoff functions*

$$f_i(\mathbf{x}) = \beta x_i - \delta_i \sum_{l=1}^{n} x_l$$

where $0 \leq \delta_i \leq 1 < \beta$*. Also suppose that* $\max_{i \in N} \delta_i = 1$ *and there does not exist* $\mathbf{m} \in M^n$ *with* $\beta = \sum_{l \in \hat{S}(\mathbf{m})} \delta_l$*.*

---

[30] Of course, this also follows from the explicit expression for $g_i(\mathbf{m})$.

Consider for $\mathbf{m} \in M^n$ *the coalitional game* $\Gamma_{R_c(\mathbf{m})}$. *In terms of the action profile* $\mathbf{x} \in X^n$ *the payoff function of meta player* $C \in R_c(\mathbf{m})$ *equals*

$$\hat{P}_C(\mathbf{x}) = \begin{cases} (\beta - \delta_i)x_i - \delta_i \sum_{l \neq i} x_l \text{ if } C = \{i\} \text{ with } i \in \hat{S}_\star(\mathbf{m}), \\ \sum_{i \in C}(\beta - \sum_{l \in C} \delta_l)x_i - (\sum_{l \in C} \delta_l) \sum_{l \notin C} x_l \text{ if } C = \hat{S}(\mathbf{m}). \end{cases}$$

It is clear that in this coalitional game each meta player $C = \{i\}$ *has* 1 *as a strictly dominant action. Additionally, meta player* $C = \hat{S}(\mathbf{m})$ *has a strictly dominant action: if* $\beta > \sum_{l \in \hat{S}(\mathbf{m})} \delta_l$, *then each of its members plays* 1 *and if* $\beta < \sum_{l \in \hat{S}(\mathbf{m})} \delta_l$, *then each of its members plays* 0. *Thus, this game has a unique Nash equilibrium, the two-stage coalition formation game is regular and for the with this equilibrium corresponding action profile* $\hat{\mathbf{x}}(\mathbf{m})$ *we have*

$$\hat{x}_i(\mathbf{m}) = \begin{cases} 1 \text{ if } i \in \hat{S}_\star(\mathbf{m}), \\ 1 \text{ if } i \in \hat{S}(\mathbf{m}) \text{ in case } \beta > \sum_{l \in \hat{S}(\mathbf{m})} \delta_l, \\ 0 \text{ if } i \in \hat{S}(\mathbf{m}) \text{ in case } \beta < \sum_{l \in \hat{S}(\mathbf{m})} \delta_l. \end{cases}$$

Now further suppose that $\delta_1 = \cdots = \delta_n = 1$ *and let us consider the effective game* G. *For its effective payoffs* $g_i(\mathbf{m}) = f_i(\hat{\mathbf{x}}(\mathbf{m}))$, *we obtain*

$$g_i(\mathbf{m}) = \begin{cases} \beta - n \text{ if } i \in \hat{S}_\star(\mathbf{m}) \text{ in case } \beta > \#\hat{S}(\mathbf{m}), \\ \beta - n + \#\hat{S}(\mathbf{m}) \text{ if } i \in \hat{S}_\star(\mathbf{m}) \text{ in case } \beta < \#\hat{S}(\mathbf{m}), \\ \beta - n \text{ if } i \in \hat{S}(\mathbf{m}) \text{ in case } \beta > \#\hat{S}(\mathbf{m}), \\ \#\hat{S}(\mathbf{m}) - n \text{ if } i \in \hat{S}(\mathbf{m}) \text{ in case } \beta < \#\hat{S}(\mathbf{m}). \end{cases}$$

Define the integer $s_\star$ *as follows*

$$\text{if } \beta < n : s_\star \text{ is the unique integer such that } \beta < s_\star < \beta + 1,$$

$$\text{if } \beta > n : s_\star = n.$$

Note that $2 \leq s_\star \leq n$. *Now, writing* $s = \#\hat{S}(\mathbf{m})$,

$$g_i(\mathbf{m}) = \begin{cases} \beta - n \text{ if } i \in \hat{S}_\star(\mathbf{m}) \text{ in case } s \leq s_\star - 1, \\ \beta - n + s \text{ if } i \in \hat{S}_\star(\mathbf{m}) \text{ in case } s \geq s_\star, \\ \beta - n \text{ if } i \in \hat{S}(\mathbf{m}) \text{ in case } s \leq s_\star - 1, \\ s - n \text{ if } i \in \hat{S}(\mathbf{m}) \text{ in case } s \geq s_\star. \end{cases}$$

Again the effective game is symmetric. Now let us determine the stability function t. *For the functions* w_0 *and* w_1 *in (8) and (9) we obtain*

$$w_0(s) = g_1(0, \mathbf{1}_s, \mathbf{0}) = \begin{cases} \beta - n \ (0 \leq s \leq s_\star - 1), \\ \beta - n + s \ \ (s_\star \leq s \leq n - 1), \end{cases}$$

$$w_1(s) = g_1(\mathbf{1}_s, \mathbf{0}) = \begin{cases} \beta - n \ (1 \leq s \leq s_\star - 1), \\ s - n \ (s_\star \leq s \leq n). \end{cases}$$

With (10) we obtain $t(0) = \infty$, $t(s) = (\beta - n) - (\beta - n) = 0 \ (1 \leq s \leq s_\star - 1)$, $t_{s_\star} = (s_\star - n) - (\beta - n) = s_\star - \beta > 0$, $t_s = (s - n) - (\beta - n + s - 1) = 1 - \beta < 0$, $t(n+1) = -\infty$. *It follows that the equilibrium coalition structures are exactly the cartel coalition structures with* $0, 1, \ldots, s_\star - 2$ *and* $s_\star$ *cooperators. The semi-strict equilibrium coalition structures are exactly the cartel coalition structures with* $s_\star$ *cooperators.* $\diamond$

Example 7. *Consider the cartel game with as physical game the Cournot-like public good game with common action set $X = \mathbb{R}_+$ and with payoff functions*

$$f_i(\mathbf{x}) = -\gamma_i x_i + b_i(\sum_{l=1}^{n} x_l)$$

where $0 < \gamma_1 < \cdots < \gamma_n$ and b_i continuously differentiable, strictly concave, and strictly increasing with

$$0 < b_i'(0) < \gamma_i \text{ and } \lim_{y \to \infty} b_i'(y) = 0.$$

We shall prove that for each semi-strict Nash equilibrium $\mathbf{m}^\star$ its associated physical action profile $\hat{\mathbf{x}}(\mathbf{m}^\star)$ is the action profile $\mathbf{0}$.

Consider, for $\mathbf{m} \in M^n$, the coalitional game $\Gamma_{R_c(\mathbf{m})}$. In terms of the physical action profile $\mathbf{x}$ the payoff of meta player C is

$$\hat{P}_C(\mathbf{x}) = \begin{cases} b_i(x_i + \sum_{l \neq i} x_l) - \gamma_i x_i \text{ if } C = \{i\} \text{ with } i \in \hat{S}_\star(\mathbf{m}), \\ \sum_{l \in C} b_l(\sum_{i \in C} x_i + \sum_{j \notin C} x_j) - \sum_{l \in C} \gamma_l x_l \text{ if } C = \hat{S}(\mathbf{m}). \end{cases}$$

It follows that each meta player $C = \{i\}$ has 0 as strictly dominant strategy. However, the best-reply correspondence of meta player $C = \hat{S}(\mathbf{m})$ depends on $\sum_{j \notin C} x_j$ which means that such a meta player does not have a strictly dominant strategy.

The above implies that $\mathbf{k} = (k_C)_{C \in R_c(\mathbf{m})}$ is a Nash equilibrium of $\Gamma_{R_c(\mathbf{m})}$ if and only if

$$k_i = 0 \ (i \in \hat{S}_\star(\mathbf{m}))$$

and $k_{\hat{S}(\mathbf{m})}$ is, with $Z := \prod_{l \in \hat{S}(\mathbf{m})} \mathbb{R}_+$, a solution of the concave programming problem

$$\text{MAX}_{\mathbf{k} \in Z} \sum_{l \in \hat{S}(\mathbf{m})} b_l(\sum_{r \in \hat{S}(\mathbf{m})} k_r) - \sum_{l \in \hat{S}(\mathbf{m})} \gamma_l k_l.$$

As Z is not open, special attention has to be given to analyse this problem with a (standard) Karush–Kuhn–Tucker like theorem. Concerning this, we note that b_i can be extended to an open interval J containing $\mathbb{R}_+$ on which b_i is continuously differentiable, strictly concave, and strictly increasing. The Karush–Kuhn–Tucker theorem guarantees that $k_{\hat{S}(\mathbf{m})}$ is a solution of the concave programming problem if and only if, writing $C = \hat{S}(\mathbf{m})$, there exists λ_j $(j \in C)$ such that

$$\text{for all } i \in C: \ \ k_i \geq 0;$$

$$\text{for all } j \in C: \ \ \lambda_j \geq 0;$$

$$\text{for all } i \in C: \sum_{l \in C} b_l'(\sum_{r \in C} k_r) - \gamma_i + \lambda_i \leq 0;$$

$$\text{for all } i \in C: k_i \Big(\sum_{l \in C} b_l'(\sum_{r \in C} k_r) - \gamma_i + \lambda_i \Big) = 0;$$

$$\text{for all } j \in C: \lambda_j k_j = 0.$$

This concave programming problem has a unique solution. In order to formulate this solution we define for a non-empty subset A of N, the function $h_A : \mathbb{R}_+ \to \mathbb{R}$ by $h_A := \sum_{l \in A} b_l'$. Note that h_A is strictly decreasing and that its image $h_A(\mathbb{R}_+) =]0, h_A(0)]$ equals $]0, \sum_{l \in A} b_l'(0)]$. The unique solution is:[31]

[31] Here is a proof. First we prove that if there is a solution; then, it is the one given by the three above case expressions.

$$k_i = \begin{cases} 0 & \textit{if there exists } j \in C \setminus \{i\} \textit{ with } \gamma_j < \gamma_i, \\ 0 & \textit{if for all } j \in C \setminus \{i\} \textit{ one has } \gamma_j > \gamma_i \wedge h_C(0) \le \gamma_i, \\ h_C^{<-1>}(\gamma_i) & \textit{if for all } j \in C \setminus \{i\} \textit{ one has } \gamma_j > \gamma_i \wedge h_C(0) > \gamma_i. \end{cases}$$

The conclusion is that the two-stage coalition formation game is regular and that its effective physical actions are

$$\hat{x}_i(\mathbf{m}) = \begin{cases} 0 & \textit{if } i \in \hat{S_\star}(\mathbf{m}), \\ 0 & \textit{if } i \in \hat{S}(\mathbf{m}) \textit{ and } \exists j \in \hat{S}(\mathbf{m}) \setminus \{i\} \textit{ with } \gamma_j < \gamma_i, \\ 0 & \textit{if } i \in \hat{S}(\mathbf{m}) \textit{ and } \forall j \in \hat{S}(\mathbf{m}) \setminus \{i\} \textit{one has } \gamma_j > \gamma_i \wedge h_{\hat{S}(\mathbf{m})}(0) \le \gamma_i, \\ h_{\hat{S}(\mathbf{m})}^{<-1>}(\gamma_i) & \textit{if } i \in \hat{S}(\mathbf{m}) \textit{ and } \forall j \in \hat{S}(\mathbf{m}) \setminus \{i\} \textit{ one has } \gamma_j > \gamma_i \wedge h_{\hat{S}(\mathbf{m})}(0) > \gamma_i. \end{cases}$$

Introducing for $\mathbf{m} \neq \mathbf{0}$, $i_{\mathbf{m}} := \min(\hat{S}(\mathbf{m}))$ *we can rewrite* $\hat{x}_i(\mathbf{m})$ *as follows*

$$\hat{x}_i(\mathbf{m}) = \begin{cases} 0 & \textit{if } i \in \hat{S_\star}(\mathbf{m}), \\ 0 & \textit{if } i \in \hat{S}(\mathbf{m}) \setminus i_{\mathbf{m}}, \\ 0 & \textit{if } i \in \hat{S}(\mathbf{m}) \wedge i = i_{\mathbf{m}} \wedge h_{\hat{S}(\mathbf{m})}(0) \le \gamma_{i_{\mathbf{m}}}, \\ h_{\hat{S}(\mathbf{m})}^{<-1>}(\gamma_{i_{\mathbf{m}}}) & \textit{if } i \in \hat{S}(\mathbf{m}) \wedge i = i_{\mathbf{m}} \wedge h_{\hat{S}(\mathbf{m})}(0) > \gamma_{i_{\mathbf{m}}}. \end{cases}$$

Note that we have $\sum_{l=1}^n \hat{x}_l(\mathbf{m}) = \hat{x}_{i_{\mathbf{m}}}(\mathbf{m})$ $(\mathbf{m} \neq \mathbf{0})$ *and* $\sum_{l=1}^n \hat{x}_l(\mathbf{0}) = 0$. *For the effective payoffs* $g_i(\mathbf{m}) = f_i(\hat{\mathbf{x}}(\mathbf{m}))$, *we obtain* $g_i(\mathbf{0}) = b_i(0)$ *and, for* $\mathbf{m} \neq \mathbf{0}$,

$$g_i(\mathbf{m}) = -\gamma_i \hat{x}_i(\mathbf{m}) + b_i(\hat{x}_{i_{\mathbf{m}}}(\mathbf{m}))$$

$$= \begin{cases} b_i(\hat{x}_{i_{\mathbf{m}}}(\mathbf{m})) & \textit{if } i \in \hat{S}_\star(\mathbf{m}), \\ b_i(\hat{x}_{i_{\mathbf{m}}}(\mathbf{m})) & \textit{if } i \in \hat{S}(\mathbf{m}) \wedge i \neq i_{\mathbf{m}}, \\ b_{i_{\mathbf{m}}}(\hat{x}_{i_{\mathbf{m}}}(\mathbf{m})) & \textit{if } i \in \hat{S}(\mathbf{m}) \wedge i = i_{\mathbf{m}} \wedge h_{\hat{S}(\mathbf{m})}(0) \le \gamma_{i_{\mathbf{m}}}, \\ -\gamma_{i_{\mathbf{m}}} \hat{x}_{i_{\mathbf{m}}}(\mathbf{m}) + b_{i_{\mathbf{m}}}(\hat{x}_{i_{\mathbf{m}}}(\mathbf{m})) & \textit{if } i \in \hat{S}(\mathbf{m}) \wedge i = i_{\mathbf{m}} \wedge h_{\hat{S}(\mathbf{m})}(0) > \gamma_{i_{\mathbf{m}}}. \end{cases}$$

The assumptions on b_i *imply that* $-\gamma_i x_i + b_i(x_i)$ *is a strictly decreasing function of* x_i. *With this we obtain for* $\mathbf{m} \neq \mathbf{0}$,

$$g_{i_{\mathbf{m}}}(\mathbf{m}) = \begin{cases} b_{i_{\mathbf{m}}}(0) & \textit{if } h_{\hat{S}(\mathbf{m})}(0) \le \gamma_{i_{\mathbf{m}}}, \\ < b_{i_{\mathbf{m}}}(0) & \textit{if } h_{\hat{S}(\mathbf{m})}(0) > \gamma_{i_{\mathbf{m}}}. \end{cases}$$

Note that for $\mathbf{m} \neq \mathbf{0}$ *we have* $h_{\hat{S}(\mathbf{m})}(0) = \sum_{l \in \hat{S}(\mathbf{m})} b_l'(0) = b_{i_{\mathbf{m}}}'(0) + \sum_{l \in \hat{S}(\mathbf{m}) \setminus i_{\mathbf{m}}} b_l'(0)$. *As* $b_{i_{\mathbf{m}}}'(0) < \gamma_{i_{\mathbf{m}}}$, *this implies*

$$h_{\hat{S}(\mathbf{m})}(0) > \gamma_{i_{\mathbf{m}}} \Rightarrow \#\hat{S}(\mathbf{m}) \ge 2.$$

Now we are ready to prove the desired result, that for each semi-strict Nash equilibrium $\mathbf{m}^\star$ *it holds that* $\hat{\mathbf{x}}(\mathbf{m}^\star) = \mathbf{0}$. *This we do by contradiction. So suppose* $\mathbf{m}^\star$ *is a semi-strict Nash equilibrium* $\mathbf{m}^\star$ *with* $\hat{\mathbf{x}}(\mathbf{m}^\star) \neq \mathbf{0}$.

First case: fix j with $\gamma_j < \gamma_i$. By contradiction suppose $k_i \neq 0$. Then $\lambda_i = 0$ and $h_C(\sum_{r \in C} k_r) = \gamma_i$. Also $h_C(\sum_{r \in C} k_r) \le \gamma_j - \lambda_j < \gamma_i - \lambda_j \le \gamma_i$, a contradiction.

Second case: suppose $h_C(0) \le \gamma_i$. If $\lambda_i \neq 0$, then $k_i = 0$. Now suppose $\lambda_i = 0$. By contradiction suppose $k_i \neq 0$. Then $h_C(\sum_{r \in C} k_r) - \gamma_i - \lambda_i = h_C(\sum_{r \in C} k_r) - \gamma_i < h_C(0) - \gamma_i \le 0$, and therefore, $k_i = 0$, a contradiction.

Third case: here $i = \min(C)$. The first case implies that $k_j = 0$ $(j \neq i)$. First we prove that $k_i \neq 0$. Well, $k_i = 0$ would imply $h_C(0) - \gamma_i + \lambda_i \le 0$, and therefore, $h_C(0) \le \gamma_i - \lambda_i \le \gamma_i$, a contradiction with $h_C(0) > \gamma_i$. As $k_i \neq 0$, we have $\lambda_i = 0$ and $h_C(k_i) - \gamma_i = 0$. Thus $k_i = h_C^{<-1>}(\gamma_i)$. Next, the given k_i indeed provides a solution as for these k_is there exists λ_j $(j \in C)$ such that the above five Karush–Kuhn–Tucker conditions are satisfied.

Then $\mathbf{m}^\star \neq \mathbf{0}$*. Fix* $i \in N$ *with* $\hat{x}_i(\mathbf{m}^\star) > 0$*. By the above* $i = i_{\mathbf{m}^\star}$ *and* $h_{\hat{S}(\mathbf{m}^\star)}(0) > \gamma_{i_{\mathbf{m}^\star}}$*. Noting that* $i_{\mathbf{m}^\star} \in \hat{S}(\mathbf{m}^\star)$ *and* $T_{i_{\mathbf{m}^\star}}(\mathbf{m}^\star) \neq \mathbf{0}$ *(as* $\#\hat{S}(\mathbf{m}^\star) \geq 2$*) we obtain* $g'_{i_{\mathbf{m}^\star}}(\mathbf{m}^\star) = g_{i_{\mathbf{m}^\star}}(T_{i_{\mathbf{m}^\star}}(\mathbf{m})) - g_{i_{\mathbf{m}^\star}}(\mathbf{m}^\star) = b_{i_{\mathbf{m}^\star}}(\hat{x}_{i_{T_{i_{\mathbf{m}^\star}}(\mathbf{m}^\star)}}(T_i(\mathbf{m}^\star))) - g_{i_{\mathbf{m}^\star}}(\mathbf{m}^\star) > b_{i_{\mathbf{m}^\star}}(\hat{x}_{i_{T_{i_{\mathbf{m}^\star}}(\mathbf{m}^\star)}}(T_{i_{\mathbf{m}^\star}}(\mathbf{m}^\star))) - b_{i_{\mathbf{m}^\star}}(0) \geq 0$*, which is impossible as* $\mathbf{m}^\star$ *cannot be a semi-strict equilibrium.* $\diamond$

Example 8. *Consider the cartel game with a physical game the Cournot-like public good game with common action set* $X = \{0, 1\}$ *and with payoff functions*

$$f_i(\mathbf{x}) = -cx_i + \beta_i \sum_{l=1}^{n} x_l$$

where $c > 1$ *and* $\beta_i \in \{\alpha, 1\}$ *with* $0 < \alpha < 1$*. We suppose that there does not exist integers* k_1, k_2 *with* $\alpha k_1 + k_2 = c$*. Let* N_1 *be the set of players i with* $\beta_i = \alpha$ *and* N_2 *be the set of players i with* $\beta_i = 1$*. We refer to the players in* N_t *as "type t players."*

We obtain for the payoff of meta player C in the coalitonal game $\Gamma_{R_c(\mathbf{m})}$ *in terms of the physical action profile* $\mathbf{x}$

$$\hat{P}_C(\mathbf{x}) = \begin{cases} -cx_i + \beta_i \sum_{l \in N} x_l & \text{if } C = \{i\} \text{ with } i \in \hat{S}_\star(\mathbf{m}), \\ -c\sum_{i \in C} x_i + (\sum_{i \in C} \beta_i)(\sum_{l \in C} x_l + \sum_{l \notin C} x_l) & \text{if } C = \hat{S}(\mathbf{m}). \end{cases}$$

From this follows that the coalitional game $\Gamma_{R_c(\mathbf{m})}$ *has a unique Nash equilibrium in strictly dominant strategies with effective physical actions*

$$\hat{x}_i(\mathbf{m}) = \begin{cases} 0 & \text{if } i \in \hat{S}_\star(\mathbf{m}), \\ 0 & \text{if } i \in \hat{S}(\mathbf{m}) \wedge \alpha k_1(\mathbf{m}) + k_2(\mathbf{m}) < c, \\ 1 & \text{if } i \in \hat{S}(\mathbf{m}) \wedge \alpha k_1(\mathbf{m}) + k_2(\mathbf{m}) > c. \end{cases}$$

where $k_t(\mathbf{m}) = \#\{i \in \hat{S}(\mathbf{m}) \mid i \in N_t\}$ $(t = 1, 2)$*. Thus the two-stage game is regular and the effective payoffs are given by*

$$g_i(\mathbf{m}) = \begin{cases} 0 & \text{if } i \in \hat{S}_\star(\mathbf{m}) \wedge \alpha k_1(\mathbf{m}) + k_2(\mathbf{m}) < c, \\ \beta_i(k_1(\mathbf{m}) + k_2(\mathbf{m})) & \text{if } i \in \hat{S}_\star(\mathbf{m}) \wedge \alpha k_1(\mathbf{m}) + k_2(\mathbf{m}) > c, \\ 0 & \text{if } i \in \hat{S}(\mathbf{m}) \wedge \alpha k_1(\mathbf{m}) + k_2(\mathbf{m}) < c, \\ -c + \beta_i(k_1(\mathbf{m}) + k_2(\mathbf{m})) & \text{if } i \in \hat{S}(\mathbf{m}) \wedge \alpha k_1(\mathbf{m}) + k_2(\mathbf{m}) > c. \end{cases}$$

Simply writing $k_1 = k_1(\mathbf{m})$ *and* $k_2 = k_2(\mathbf{m})$ *we obtain the following formulas for* $g'_i(\mathbf{m})$*.*

If $i \in \hat{S}_\star(\mathbf{m})$*, then*

$$g'_i(\mathbf{m}) = \begin{cases} 0 & \text{if } \alpha(k_1+1) + k_2 < c \wedge \alpha k_1 + k_2 < c \wedge i \in N_1, \\ \alpha(1 + k_1 + k_2) - c & \text{if } \alpha(k_1+1) + k_2 > c \wedge \alpha k_1 + k_2 < c \wedge i \in N_1, \\ (1 + k_1 + k_2) - c & \text{if } \alpha k_1 + k_2 + 1 > c \wedge \alpha k_1 + k_2 < c \wedge i \in N_2, \\ 0 & \text{if } \alpha k_1 + k_2 + 1 < c \wedge \alpha k_1 + k_2 < c \wedge i \in N_2, \\ \alpha - c & \text{if } \alpha k_1 + k_2 > c \wedge i \in N_1, \\ 1 - c & \text{if } \alpha k_1 + k_2 > c \wedge i \in N_2. \end{cases}$$

And if $i \in \hat{S}(\mathbf{m})$, *then*

$$g_i'(\mathbf{m}) = \begin{cases} 0 & \text{if } \alpha k_1 + k_2 < c \wedge i \in N_1, \\ 0 & \text{if } \alpha k_1 + k_2 < c \wedge i \in N_2, \\ c - \alpha & \text{if } \alpha(k_1 - 1) + k_2 > c \wedge \alpha k_1 + k_2 > c \wedge i \in N_1, \\ c - \alpha(k_1 + k_2) & \text{if } \alpha(k_1 - 1) + k_2 < c \wedge \alpha k_1 + k_2 > c \wedge i \in N_1, \\ c - 1 & \text{if } \alpha k_1 + k_2 - 1 > c \wedge \alpha k_1 + k_2 > c \wedge i \in N_2, \\ c - (k_1 + k_2) & \text{if } \alpha k_1 + k_2 - 1 < c \wedge \alpha k_1 + k_2 > c \wedge i \in N_2. \end{cases}$$

The conditions for a semi-strict Nash-equilibrium require that $g_i'(\mathbf{m}) \leq 0$ *for each cooperator i and* $g_i'(\mathbf{m}) < 0$ *for each non-cooperator i. From the possible cases for* $g_i'(\mathbf{m})$ *above, we see that for a non-cooperator* $g_i'(\mathbf{m}) < 0$ *is only possible for the cases* $2, 5, 6$ *in the above formula of* $g_i'(\mathbf{m})$. *Requiring* $g_i'(\mathbf{m}) \leq 0$ *for each cooperator, we see that this only possible for the cases* $1, 2, 4, 6$ *in the above formula of* $g_i'(\mathbf{m})$. *Combining these conditions (5 and 6 for non-cooperators with 4 and 6 for cooperators) it follows that for a membership profile* $\mathbf{m}$ *with*

$$0 \leq k_1(\mathbf{m}) \leq \#N_1 \ \wedge \ 0 \leq k_2(\mathbf{m}) \leq \#N_2,$$

a sufficient condition for being a semi-strict equilibrium is:

$$c + \alpha > \alpha k_1(\mathbf{m}) + k_2(\mathbf{m}) > c \ \wedge \ \alpha(k_1(\mathbf{m}) + k_2(\mathbf{m})) > c;$$

note that here it holds that $k_1(\mathbf{m}) \neq 0$. *But when looking for semi-strict Nash equilibria with* $k_1(\mathbf{m}) \neq 0$ *or* $k_2(\mathbf{m}) \neq 0$ *each of the following two conditions separately, are sufficient:*

$$c + \alpha > \alpha k_1(\mathbf{m}) > c \wedge k_2(\mathbf{m}) = 0;$$

$$c + 1 > k_2(\mathbf{m}) > c \wedge k_1(\mathbf{m}) = 0. \diamond$$

11. Things to Do (Instead of Conclusions)

> 'The time has come," the Walrus said, "To talk of many things: Of shoes—and ships—and sealing-wax—Of cabbages—and kings. And why the sea is boiling hot—And whether pigs have wings." (Lewis Carroll)

Although two-stage coalition formation games have been studied already for more than 30 years, their theoretical features have still not been sufficiently investigated. For example, to the best of our knowledge, there is no theoretical result about the existence of an effective equilibrium coalition structure for a regular cartel game that holds for a (sufficiently) abstract Cournot-like physical game in terms of its qualitative properties, although it is the simplest type of a two-stage coalition formation game. Only results for concrete (mostly symmetric) physical games with quite simple payoff functions are available. Therefore, all claims in the literature concerning such situations, for example, on the sizes of the coalitions in equilibrium coalition structures, are only supported by specific examples. In this section we make some suggestions how further progress can be made.

We see three reasons why such full results for two-stage coalition formation games are, presently, not available in the literature:

I. A lack of sufficient conditions in terms of the primitives of the physical game for each possible coalitional game to have a unique Nash equilibrium.
II. A lack of results concerning the qualitative properties of the effective payoffs in terms of the primitives of the physical game.
III. A lack of general results concerning the structure of the Nash equilibrium set of finite games, such as binary action games, that arise as effective games in the theory of the two-stage game

approach to coalition formation. In particular, for these effective games, a lack of general results on the validity of the "paradox of cooperation."

These problems may be the basis for a research program. We discuss them now in more detail.

Problem I. This problem concerns regularity of two stage coalition formation games. Regularity is very fundamental for the two-stage approach.[32] One needs sufficient conditions in terms of the primitives of the physical game for the two stage coalition formation game to be regular. As the ineffective coalition structure often (as in a cartel game) is a possible coalition structure, one has to understand in particular the Nash equilibria of the physical game. Let us mention here that, although Cournot-like games and in particular Cournot oligopolies belong to the most studied games in the literature, this problem is, as recent literature (e.g., [35,41,42]) shows, still an active research subject. Of course, for coalitional Nash equilibria the equilibrium uniqueness problem is even more complicated. The only abstract result for equilibrium uniqueness for coalitional equilibria in the literature for Cournot-like games is, as far as we know, Theorem 1, taken form [36].[33] It may be good to improve this theorem as follows, along various lines.

1. Theorem 1 provides sufficient conditions for each coalitional game to have a unique Nash equilibrium. However in a cartel game not all coalition structures are possible. So for a cartel game, improvements may be possible if we restrict ourselves to the subset of possible coalitional games.

2. Theorem 1 deals with compact action sets. As various games in the literature deal with situations where each player has $\mathbb{R}_+$ as action set, variants of Theorem 1(using some "effective compactness condition') for this situation are needed.

3. The results in Theorem 1 use assumptions that, loosely speaking, make that the actions are strategic substitutes (having downward sloping best replies). These assumptions are popular. However, in [20,43] it is argued that one should also give more attention to situations where the actions are strategic complements (and have upward sloping best replies). This could be helpful for the analysis of climate cooperation when countries have the ability to adapt to climate change.[34]

Problem II. In order to understand the qualitative properties of the payoffs of the effective game G in terms of the primitives of the physical game Γ, one first has to have some insight in the properties of the mapping that assigns the membership action profile (via the unique Nash equilibrium of the coalitional game $\Gamma_{R(\mathbf{m})}$) to the effective physical action profile $\mathbf{x}(\mathbf{m})$. Various initial, simple results concerning this mapping exist (see [46] and references therein). Next one has to obtain insight into the mapping that assigns to a membership action profile $\mathbf{m}$, for $i \in N$, the effective payoff $g_i(\mathbf{m})$.

An intuitive general result (with a quite technical proof) in this context is Theorem 2 in the case of a symmetric physical game and a symmetric membership rule: its first part states that the effective game G is symmetric and its second part specifies further qualitative properties of the effective payoff functions. In the literature, other properties of the effective game, such as super-additivity, are dealt with, but there are no results for these properties in terms of the primitives of the physical game (even if it is symmetric). Upon approaching these problems, we suggest the following.

1. A careful study of the concrete examples in the literature in order to find the crucial assumptions. A good starting point is the case of a cartel game with a symmetric physical game.

2. Providing sufficient conditions in terms of the primitives of the physical game for the effective game to be super-additive.

---

[32] Up to now, how to solve the two-stage game in the case it is not regular, has not been addressed in the literature.
[33] The technique for proving equilibrium uniqueness of of coalitional equilibria in [36] goes back to [31] and was generalised and refined in [32,34]. This technique relies on an analysis of first order conditions by means of so-called "marginal reductions."
[34] See, for example, [44,45].

3. A generalisation of a symmetric game, where players are "identical," is a game where there are two types of players (as in Example 8 in section 10). It would be very interesting to have a variant of Theorem 2 for such a game.

Problem III. Understanding the set of equilibrium coalition structures of a regular two-stage coalition formation game comes down to understanding of the Nash equilibrium set of its effective game G. This game is a finite game, and due to its construction has a special structure. For example, for the popular case of a cartel game, it is a binary action game. We have seen (Proposition 6, Theorem 6, Proposition 5) that a symmetric binary action game is an aggregative game, an exact potential game, a congestion game and has the deviation property D_1. Also (Theorem 5) a binary action game with the deviation property D_1 is a generalised ordinal potential game and has a semi-strict Nash equilibrium. Next steps in understanding the equilibrium set of the effective game in case of a regular cartel game could entail the following.

1. Give sufficient conditions (in terms of the primitives of the physical game) for the effective game to have single valued best-reply correspondences. And give sufficient conditions (in terms of the primitives of the physical game) for the effective game to have the deviation property D_1.

2. There exists nowadays, a considerable literature about aggregative games (see, for instance, [22,47]). However, in the present article this was not exploited. Especially it seems to be interesting to further study aggregative binary action games which allow for a potential function.

3. The structure of the Nash equilibrium set of the effective game ultimately determines the coalition size of the cooperators in equilibrium coalitional structures. Concerning this size it would be interesting to formalise the so-called paradox of cooperation (see [30] and references therein) and to analyse its claim. The "paradox" was first established by [6] for a public good game with quadratic costs and quadratic benefits in a Stackelberg setting (3-stage game). It refers to the finding that large coalitions are stable when the gains from cooperation are small. If gains are large, only small coalitions will be stable, if any. Later, similar features of stable coalitional equilibria were established in 2-stage games. For example in [48] evidence for the "paradox" was found from a large number of simulations.[35] Still the paradox is not well understood, let alone characterised more formally.[36]

4. In (16) in Example 5, $\overline{x}(s)$ and $c(\overline{x}(s))$ are strictly increasing functions of s. This implies that w_0 is a strictly increasing function of s. A further analysis of (16) would be interesting. Concerning this, we want to note that the cartel game in [49] is a variant of Example 5. The formula (16) applies to this variant: instead of our $c(x) = \frac{a}{p}x^p$, [49] deals with $c(x) = x + \frac{a}{p}x^p$. This slight modification leads to completely different equilibrium cartel coalition structures.

Needless to say, our list of problems is not exhaustive.

Author Contributions: P.v.M. had the idea for the paper and has written the first draft. All authors contributed jointly to the further development of the paper, its structure, the writing, and the examples. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Haeringer, G. Sur la Coopération dans les Jeux Non Coopératifs. *Revue d'Economie Industrielle* **2003**, *103*, 175–190.
2. Marini, M.A. *An Overview of Coalition and Network Formation for Economic Models*; Technical report, EMS WP and CREI Working Paper; Università di Roma Tre: Roma, Italy, 2007.

[35] Also see [49].
[36] In [50] the effect of coalition formation on payoffs and equilibrium coalition structures is decomposed into an internalisation effect, a cost-effectiveness effect, and a timing effect. The analysis, however, only applies to the Stackelberg setting.

3. d'Aspremont, C.; Jaquemin, A.; Gabszewicz, J.; Weymark, J. On the Stability of Collusive Price Leadership. *Can. J. Econ.* **1983**, *16*, 17–25.
4. Salant, S.W.; Switzer, S.; Reynolds, R.J. Losses from Horizontal Merger: The Effects of an Exogeneous Change in Industry Structure on Cournot-Nash Equilibtium. *Q. J. Econ.* **1983**, *98*, 185–199.
5. Carraro, C.; Siniscalco, D. Strategies for the International Protection of the Environment. *J. Publ. Econ.* **1993**, *52*, 309–328.
6. Barrett, S. Self-enforcing International Environmental Agreements. *Oxf. Econ. Pap.* **1994**, *46*, 878–894.
7. Carraro, C. *The Endogeneous Formation of Economic Coalitions*; Edward Elgar: Cheltenham, UK, 2003.
8. Benchekroun, H.; Van Long, N. Collaborative Environmental Management: A Review of the Literarature. *Int. Game Theory Rev.* **2012**, *14*, doi:10.1142/S0219198912400026.
9. Marrouch, W.; Chaudhuri, A.R. International Environmental Agreements: Doomed to Fail or Destined to Succeed? A Review of the Literature. *Int. Rev. Environ. Resour. Econ.* **2015**, *9*, 245–319.
10. Punt, M.; Wesseler, J. The Formation of GM-free and GM-Coasean Clubs: Will They Form and if so How Much Can They Achieve? *J. Agric. Econ.* **2018**, *69*, 413–438.
11. Alvarado-Quesada, I.; Weikard, H.P. International Cooperation on Biodiversity Conservation when Spatial Structures Matter. *Spat. Econ. Anal.* **2017**, *12*, 27–49.
12. Klepac, P.; Megiddo, I.; Grenfell, B.T.; Laxminarayan, R. Self-enforcing Regional Vaccination Agreements. *J. R. Soc. Interface* **2016**, *13*, 20150907. doi:10.1098/rsif.2015.0907
13. Hagen, A.; Kähler, L.; Eisenack, K. Transnational Environmental Agreements with Heterogeneous Actors. In *Economics of International Environmental Agreements: A Critical Approach*; Kayalica, M.O., Selim Çağatay, S., Mihçi, H., Eds.; Routledge: Abingdon, UK, 2017; pp. 79–96.
14. Ansink, E.; Weikard, H.P.; Withagen, C. International Environmental Agreements with Support. *J. Environ. Econ. Manag.* **2019**, *97*, 241–252.
15. Li, Q.; Fujita, T. Stabilizing International Environmental Agreements. In *International Development and the Environment. Sustainable Development Goals Series*; Hori, S., Takamura, Y., Fujita, T., Kanie, N., Eds.; Springer: Singapore, 2020.
16. Bloch, F. Endogenous Structures of Association in Oligopolies. *RAND J. Econ.* **1995**, *26*, 537–576.
17. Bloch, F. Non–cooperative Models of Coalition Formation in Games with Spillovers. In *Endogenous Formation of Economic Coalitions*; Carraro, C., Ed.; Edward Elgar: Cheltenham, UK, 2003; Chapter 2, pp. 35–79.
18. Finus, M. *Game Theory and International Environmental Cooperation*; Edward Elgar: Cheltenham, UK, 2001.
19. Weikard, H.P. Cartel Stability under an Optimal Sharing Rule. *Manch. Sch.* **2009**, *77*, 575–593.
20. Heugues, M. International Environmental Cooperation: A New Eye on the Greenhouse Gas Emissions' Control. *Ann. Oper. Res.* **2014**, *220*, 239–262.
21. Thoron, S. Formation of a Coalition-Proof Stable Cartel. *Revue Canadienne d'Economique* **1998**, *31*, 63–76.
22. Cornes, R.; Hartley, R. *Well-Behaved Aggregative Games*; Economic Discussion Paper May 24; School of Social Sciences, The University of Manchester: Manchester, UK, 2011.
23. Acemoglu, D.; Jensen, M. Aggregate Comparative Statics. *Games Econ. Behav.* **2013**, *81*, 27–49.
24. von Mouche, P.H.M. The Selten-Szidarovszky Technique: The Transformation Part. In *Recent Advances in Game Theory and Applications*; Petrosyan, L.A., Mazalov, V.V., Eds.; Birkhäuser: Cham, Switzerland, 2016; pp. 147–164.
25. Belleflamme, P. Stable Coalition Structures with Open Membership and Asymmetric Firms. *Games Econ. Behav.* **2000**, *30*, 1–21.
26. Ichiishi, T. A Social Equilibrium Existence Lemma. *Econometrica* **1981**, *49*, 369–377.
27. Laraki, R. *Coalitonal Nash Equilibria*; Technical Report 361, Cahier du Lamsade; Dauphine Université Paris: Paris, France, 2014.
28. Finus, M.; Rundshagen, B. Membership Rules and Stability of Coalition Structures in Positive Externality Games. *Soc. Choice Welf.* **2009**, *32*, 389–406.
29. Eyckmans, J.; Finus, M.; Mallozzi, L. *A New Class of Welfare Maximizing Stable Sharing Rules for Partition Function Games with Externatilities*; Technical Report 6/12; Bath Economic Research Papers; Department of Economics, University of Bath: Bath, UK, 2012.
30. Finus, M.; McGinty, M. The Anti-Paradox of Cooperation. *J. Econ. Behav. Organ.* **2019**, *157*, 541–559.
31. Corchón, L.C. *Theories of Imperfectly Competitive Markets*; 2nd ed.; Lecture Notes in Economics and Mathematical Systems; Springer: Berlin, Germany, 1996; Volume 442.

32. von Mouche, P.H.M. On Games with Constant Nash Sum. In *Contributions to Game Theory and Management*; Petrosjan, L.A., Zenkevich, N.A., Eds.; Graduate School of Management, St. Petersberg University: St. Petersburg, Russia, 2011; Volume IV, pp. 294–310.
33. Forgó, F. On the Existence of Nash-equilibrium in n-person Generalized Concave Games. In *Generalized Convexity*; Lecture Notes in Economics and Mathematical Systems; Komlósi, S., Rapscák, T., Schaible, S., Eds.; Springer: Berlin, Germany, 1994; Volume 405, pp. 53–61.
34. Folmer, H.; von Mouche, P.H.M. On a Less Known Nash Equilibrium Uniqueness Result. *J. Math. Sociol.* **2004**, *28*, 67–80.
35. von Mouche, P.H.M.; Sato, T. Cournot Equilibrium Uniqueness: At 0 Discontinuous Industry Revenue and Decreasing Price Flexibility. *Int. Game Theory Rev.* **2019**, *21*, doi:10.1142/S0219198919400103.
36. Finus, M.; von Mouche, P.H.M.; Rundshagen, B. On Uniqueness of Coalitional Equilibria. In *Contributions to Game Theory and Management*; Petrosjan, L.A., Zenkevich, N.A., Eds.; Graduate School of Management, St. Petersburg State University: St. Petersburg, Russia, 2014; Volume VII, pp. 51–60.
37. Cao, Z.; Yang, X. Symmetric Games Revisited. *Math. Soc. Sci.* **2018**, *95*, 9–18.
38. Iimura, T.; von Mouche, P.H.M.; Watanabe, T. Binary Action Games: Deviation Properties, Semi-Strict Equilibria and Potentials. *Discrete Appl. Math.* **2018**, *251*, 57–68.
39. Monderer, D.; Shapley, L. Potential Games. *Games Econ. Behav.* **1996**, *14*, 124–143.
40. Park, J. Potential Games with Incomplete Preferences. *J. Math. Econ.* **2015**, *61*, 58–66.
41. Ewerhart, C. Cournot Games with Biconcave Demand. *Games Econ. Behav.* **2014**, *85*, 37–47.
42. Folmer, H.; von Mouche, P.H.M. Nash Equilibria of Transboundary Pollution Games. In *Handbook of Research Methods and Applications in Environmental Studies*; Ruth, M., Ed.; Edward-Elgar: Cheltenham, UK, 2015; pp. 504–524.
43. Bayramoglu, B.; Finus, M.; Jacques, J.F. Climate Agreements in a Mitigation-adaptation Game. *J. Public Econ.* **2018**, *165*, 101–113.
44. Breton, M.; Sbragia, L. The Impact of Adaptation on the Stability of International Environmental Agreements. *Environ. Resour. Econ.* **2019**, *74*, doi:10.1007/s10640–019–00341–y.
45. Eisenack, K.; Kähler, L. Adaptation to Climate Change Can Support Unilateral Emission Reductions. *Oxf. Econ. Pap.* **2016**, *68*, 258—-278.
46. Hagen, A.; Eisenack, K. Climate Clubs Versus Single Coalitions: The Ambition of International Environmental Agreements. *Clim. Chang. Econ.* **2019**, *10*, 258—-278.
47. Jensen, M.K. Aggregative Games. In *Handbook of Game Theory and Industrial Organization*; Corchón, L., Marini, M.A., Eds.; Edward Elgar: New York, NY, USA, 2018; Volume I.
48. Bakalova, I.; Eyckmans, J. Simulating the Impact of Heterogeneity on Stability and Effectiveness of International Environmental Agreements. *Eur. J. Oper. Res.* **2019**, *277*, 1151–1162.
49. Rauscher, M. Stable International Environmental Agreements Reconsidered. *Games* **2019**, *10*, 47.
50. McGinty, M. *Leadership and Free-Riding: Decomposing and Explaining the Paradox of Cooperation in International Environmental Agreements*; Technical report; University of Wisconsin-Milwaukee: Milwaukee, WI, USA, 2019.