Coalition formation and potential games[a]

Marco Slikker[b]

September 13, 1999

Abstract

In this paper we study the formation of coalition structures in situations described by a cooperative game. Players choose independently which coalition they want to join. The payoffs to the players are determined by an allocation rule on the underlying game and the coalition structure that results from the strategies of the players according to some formation rule. We study two well-known coalition structure formation rules. We show that for both formation rules there exists a unique component efficient allocation rule that results in a potential game and study the coalition structures resulting from potential maximizing strategy profiles.

JOURNAL OF ECONOMIC LITERATURE classification numbers: C71, C72.

KEYWORDS: cooperative game, coalition formation, potential game, potential maximizer.

[a]The author thanks Henk Norde, Anne van den Nouweland, and Stef Tijs for useful suggestions and comments.

[b]Department of Econometrics and CentER, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands. E-mail: M.Slikker@kub.nl

1 Introduction

In this paper, we study the formation of coalition structures in case the underlying economic possibilities of the players are represented by a cooperative game. The process of coalition formation and payoff division will be modeled by means of a game in strategic form, which can be seen as a two-stage game. In the first stage, players reveal their preferred coalition and some coalition formation rule then determines the coalition structure. In the second stage, the payoffs to the players are determined by an exogenously given allocation rule for cooperative games with coalition structures.

We study two descriptions of the first stage, one reflects a stringent coalition formation rule and the other a less stringent rule. Both assume that each player chooses a coalition he wants to join. According to the stringent formation rule, two players end up in the same coalition if they choose the same coalition and all players in their preferred coalition also choose this coalition, whereas the less stringent formation rule only requires that they choose the same coalition.

The second stage, the stage of payoff division, is modeled by means of an exogenous allocation rule for cooperative games with a fixed coalition structure. The payoffs to the players are determined by applying the allocation rule to the underlying cooperative game and the coalition structure that was formed in the first stage. *Aumann* and *Drèze* (1974) study such allocation rules, mainly extensions of well-known allocation rules for cooperative games to the setting of cooperative games with a coalition structure. They impose component efficiency, which states that each element in a coalition structure should divide the total payoffs accruing to this coalition among its members. We will impose component efficiency as well.

These two-stage models of coalition formation might result in multiple equilibria. In this paper we will not interfere in the discussion on equilibrium selection. Rather, we study coalition formation models that are potential games (cf. *Monderer* and *Shapley* (1996)). Potential games are games in strategic form with a natural and generally accepted equilibrium refinement, the potential maximizer.

The goal of this paper is two-fold. First, for the two specifications of the coalition formation rule described above, we study which component efficient allocation rules result in the coalition formation game being a potential game. Secondly, if the coalition formation rule and the allocation rule are such that the coalition formation game is a potential game, we study which coalition structures result according to the potential maximizer.

We will show that if we adopt the stringent formation rule in the first stage and we impose component efficiency on the allocation rule in the second stage, then there is a unique allocation rule that results in a potential game. This allocation rule divides for each element of the coalition structure the gains over stand-alone values equally among the players in this coalition.

If we adopt the less stringent coalition formation rule rather than the stringent one and impose component efficiency on the allocation rule in the second stage, then we also find that there is a unique allocation rule that results in a potential game. This allocation rule is the extension of the Shapley value to cooperative games with coalition structures as suggested by *Aumann* and *Drèze* (1974). Moreover, we show that with this allocation rule, if the underlying game is superadditive then the strategy profile resulting in the grand coalition is a potential maximizing strategy profile, and all potential maximizing strategy profiles result in the same payoffs.

Von Neumann and *Morgenstern* (1944) already present a non-cooperative two-stage game of coalition formation in (superadditive) cooperative games. The first stage is described by the stringent formation rule. *Von Neumann* and *Morgenstern* (1944) assume that if a coalition is formed, in the second stage it equally divides the value of this coalition among its members.

An innovative approach to coalition structure formation is presented in *Hart* and *Kurz* (1983). In contrast with the models described so far, *Hart* and *Kurz* (1983) assume that coalitions only form for the sake of bargaining over the division of the value of the grand coalition. For a specific coalition structure they employ the Shapley value in two negotiation stages, first between coalitions, and then within coalitions. The resulting allocation rule is called the coalitional Shapley value and coincides with the value for games with a priori unions of *Owen* (1977). *Hart* and *Kurz* (1983) then analyze a two-stage model of coalition formation. They study two descriptions of the first stage, the stringent coalition formation rule and the less stringent rule. In the second stage, the players receive the coalitional Shapley value of the underlying game with the coalition structure of the first stage.

Meca-Martinez et al. (1998) also study a two-stage model of coalition structure formation. The first stage of their model is described by the stringent coalition formation rule. Unlike *Hart* and *Kurz* (1983), *Meca-Martinez et al.* (1998) do not fix an allocation rule for the second stage a priori. Rather, they study conditions on the allocation rule to ensure that if the underlying game is convex, then the grand coalition results from a strong Nash equilibrium of the corresponding game in strategic form. In this paper we show that the results shown by *Meca-Martinez et al.* (1998) also hold for the less stringent formation rule.

In the last few years several papers have studied two-stage cooperation structure formation models and their relation to non-cooperative potential games. *Monderer* and *Shapley* (1996) study a participation game. Players choose whether or not to participate in some underlying cooperative game. Subsequently, a non-participating player receives some stand-alone value, whereas a participating player receives a payoff according to an allocation rule applied on the subgame of participating players. *Monderer* and *Shapley* (1996) restrict themselves to allocation rules that divide the value of the coalition of participating players among these players. They show that if such an allocation rule results in a potential game, then this allocation rule coincides with the Shapley value.

Related cooperation structure formation models are studied by *Qin* (1996) and *Slikker et al.* (1999). *Qin* (1996) studies a model describing the formation of bilateral interaction links, introduced by *Myerson* (1991). *Qin* (1996) shows that under an efficiency requirement there is only one allocation rule that results in a potential game and that is the Myerson value (cf. *Myerson* (1977)). Furthermore, it is shown that if the underlying game is superadditive, then the strategy profile that results in the full cooperation structure is a potential maximizing strategy profile, and every potential maximizing strategy profile results in the same payoffs as the full cooperation structure. *Slikker et al.* (1999) find similar results for the formation of conferences, where conferences are subsets of players, representing the possibilities of direct negotiations between the players.

The plan of this paper is as follows. Section 2 contains preliminaries and a characterization of the extension of the Shapley value to cooperative games with coalition structures as introduced by *Aumann* and *Drèze* (1974). In section 3 we describe the model of coalition structure formation and two coalition formation rules. We show that the results of *Meca-Martinez et al.* (1998) also hold if the less stringent coalition formation rule is employed rather than the the stringent coalition formation rule. In section 4 we show that for both descriptions of the first stage it holds that under a feasibility requirement there is only one allocation rule that results in a potential game: the value of *Aumann* and *Drèze* (1974) for the less stringent formation rule and an allocation rule that coincides with the equal division rule of *Von Neumann* and *Morgenstern* (1944) on the class of zero-normalized games for the stringent formation rule. In section 5 we show that if the less stringent formation rule and the value of *Aumann* and *Drèze* (1974) are employed, then the strategy profile resulting in the grand coalition maximizes the potential and that every strategy profile that maximizes the potential, results in the same payoffs as are obtained if the grand coalition is formed. We conclude in section 6.

2 Preliminaries

A cooperative game is a pair (N, v), where $N = \{1, \ldots, n\}$ denotes the set of players and $v : 2^N \to \mathbb{R}$ the characteristic function, with $v(\emptyset) = 0$. If no confusion can arise we sometimes identify a game with its characteristic function. The set of all cooperative games with player set N is denoted by TU^N. A cooperative game (N, v) is *zero-normalized* if for all $i \in N$ it holds that $v(\{i\}) = 0$. A cooperative game (N, v) is *superadditive* if for all $T_1 \subseteq N$ and all $T_2 \subseteq N \backslash T_1$ it holds that[1]

$$v(T_1 \cup T_2) \geq v(T_1) + v(T_2).$$

Hence, a game is superadditive if the value of the union of two disjoint coalitions (weakly) exceeds the sum of the values of these coalitions. A cooperative game (N, v) is *convex* if for all $i \in N$ and all $T_1 \subseteq T_2 \subseteq N$ with $i \in T_1$ it holds that

$$v(T_1) - v(T_1 \backslash \{i\}) \leq v(T_2) - v(T_2 \backslash \{i\}).$$

So, a game is convex if the marginal contribution of a player to a coalition is (weakly) less than his marginal contribution to a superset of that coalition.

The *subgame* $(S, v_{|S})$ corresponding to a game (N, v) with $S \subseteq N$ is determined by $v_{|S}(T) = v(T)$ for all $T \subseteq S$. The *unanimity game* (N, u_R) is the game with $u_R(S) = 1$ if $R \subseteq S$ and $u_R(S) = 0$ otherwise (see *Shapley* (1953)). Every game (N, v) can be written as a unique linear combination of unanimity games, i.e., $v = \sum_{R \subseteq N} \lambda_R(v) u_R$. In case there is no ambiguity about the underlying game we simply write λ_R instead of $\lambda_R(v)$. The Shapley value Φ of a game (cf. *Shapley* (1953)) is now easily described by

$$\Phi_i(N, v) = \sum_{R \subseteq N,\ i \in R} \frac{\lambda_R}{|R|} \text{ for all } i \in N.$$

The Shapley value is the unique allocation rule that satisfies *efficiency*, i.e., $\sum_{i \in N} \Phi_i(N, v) = v(N)$, and *balanced contributions*, i.e., $\Phi_i(N, v) - \Phi_i(N \backslash \{j\}, v_{|N \backslash \{j\}}) = \Phi_j(N, v) - \Phi_j(N \backslash \{i\}, v_{|N \backslash \{i\}})$ for all $i, j \in N$ with $i \neq j$ (see *Myerson* (1980)).

A player $i \in N$ is a *dummy player* in the game (N, v) if $v(T \cup \{i\}) = v(T) + v(\{i\})$ for all $T \subseteq N \backslash \{i\}$.

A game with a coalition structure is a triple $(N, v, \mathcal{B})$, where (N, v) is a cooperative game and $\mathcal{B}$ a partition of N. The set of all partitions of N is denoted by Π^N. For notational convenience we denote for all $\mathcal{B} = \{B_1, \ldots, B_m\} \in \Pi^N$, all $k \in \{1, \ldots, m\}$, and all $i \in B_k$:

$$\mathcal{B} - i = \{B_1, \ldots, B_{k-1}, B_k \backslash \{i\}, \{i\}, B_{k+1}, \ldots, B_m\}.$$

[1] $T \subseteq N$ denotes that T is a subset of N, $T \subset N$ denotes that T is a strict subset of N.

An allocation rule γ for cooperative games with coalition structures is a function that assigns to every triple $(N, v, \mathcal{B})$ a vector $\gamma(N, v, \mathcal{B}) \in \mathbb{R}^N$. In case there is no ambiguity on the underlying game, we will simply write $\gamma(\mathcal{B})$ instead of $\gamma(N, v, \mathcal{B})$. *Aumann* and *Drèze* (1974) studied cooperative games with coalition structures and allocation rules for these situations. Among other things they studied the allocation rule Φ^{AD} that attributes to player $i \in B_k \in \mathcal{B}$ the Shapley value Φ for player i of the game restricted to partition element B_k, i.e., $\Phi_i^{AD}(N, v, \mathcal{B}) = \Phi(B_k, v_{|B_k})$. We will refer to Φ^{AD} as *the value of Aumann and Drèze*. We characterize the value of Aumann and Drèze by two properties, component efficiency and component restricted balanced contributions. Consider these properties for an allocation rule γ:

Component Efficiency (CE) For every cooperative game (N, v) and every partition $\mathcal{B} = \{B_1, \ldots, B_m\}$ of N it holds for all $k \in \{1, \ldots, m\}$ that

$$\sum_{i \in B_k} \gamma_i(N, v, \mathcal{B}) = v(B_k).$$

Component Restricted Balanced Contributions (CRBC) For every cooperative game (N, v), every partition $\mathcal{B} = \{B_1, \ldots, B_m\}$ of N , every $k \in \{1, \ldots, m\}$, and all $i, j \in B_k$ it holds that

$$\gamma_i(N, v, \mathcal{B}) - \gamma_i(N, v, \mathcal{B} - j) = \gamma_j(N, v, \mathcal{B}) - \gamma_j(N, v, \mathcal{B} - i).$$

Theorem 2.1 The value of *Aumann* and *Drèze* (1974) is the unique allocation rule for cooperative games with coalition structures satisfying (CE) and (CRBC).

Proof: The Shapley value satisfies efficiency and balanced contributions. Since the value of Aumann and Drèze for the players in a partition element coincides with the Shapley value of the game restricted to the players of this element it follows by efficiency and balanced contributions of the Shapley value that the value of Aumann and Drèze satisfies component efficiency and component restricted balanced contributions.

Let γ be an allocation rule that satisfies component efficiency and component restricted balanced contributions. Let (N, v) be a cooperative game. We will show that

$$\gamma_i(N, v, \mathcal{B}) = \Phi_i^{AD}(N, v, \mathcal{B}), \text{ for all } \mathcal{B} \in \Pi^N, \text{ all } B \in \mathcal{B}, \text{ and all } i \in B.$$

The proof will be by induction to $|B|$. Obviously, for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$ with $|B| = 1$ it follows by component efficiency that

$$\gamma_i(N, v, \mathcal{B}) = v(\{i\}) = \Phi_i^{AD}(N, v, \mathcal{B}).$$

Let $p \in \mathbb{N}$, $p \geq 2$. Assume that it holds for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$ with $|B| \leq p-1$, and all $i \in B$ that $\gamma_i(N,v,\mathcal{B}) = \Phi_i^{AD}(N,v,\mathcal{B})$. We will prove that $\gamma_i(N,v,\mathcal{B}) = \Phi_i^{AD}(N,v,\mathcal{B})$ for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$ with $|B| = p$, and all $i \in B$.

Let $\mathcal{B} \in \Pi^N$ and $B \in \mathcal{B}$ such that $|B| = p$. Let $i \in B$, then for all $j \in B\backslash\{i\}$ it holds that

$$\begin{aligned} \gamma_j(N,v,\mathcal{B}) - \gamma_i(N,v,\mathcal{B}) &= \gamma_j(N,v,\mathcal{B}-i) - \gamma_i(N,v,\mathcal{B}-j) \\ &= \Phi_j^{AD}(N,v,\mathcal{B}-i) - \Phi_i^{AD}(N,v,\mathcal{B}-j) \\ &= \Phi_j^{AD}(N,v,\mathcal{B}) - \Phi_i^{AD}(N,v,\mathcal{B}), \end{aligned}$$

where the first and third equality follow by component restricted balanced contributions of γ and Φ^{AD} respectively. The second equality follows by the induction hypothesis. So, for all $j \in B\backslash\{i\}$:

$$\gamma_j(N,v,\mathcal{B}) - \Phi_j^{AD}(N,v,\mathcal{B}) = \gamma_i(N,v,\mathcal{B}) - \Phi_i^{AD}(N,v,\mathcal{B}). \tag{1}$$

Hence,

$$\sum_{j \in B} \left[\gamma_j(N,v,\mathcal{B}) - \Phi_j^{AD}(N,v,\mathcal{B})\right] = |B| \left[\gamma_i(N,v,\mathcal{B}) - \Phi_i^{AD}(N,v,\mathcal{B})\right].$$

Component efficiency of γ and Φ^{AD} then implies that $\gamma_i(N,v,\mathcal{B}) = \Phi_i^{AD}(N,v,\mathcal{B})$.

This completes the proof. □

Potential games associated with cooperative games, were introduced by *Hart* and *Mas-Colell* (1989). They define a potential P as a map on the set of all cooperative games. *Hart* and *Mas-Colell* (1989) remark that for a specific TU-game (N,v) one can view $P^{HM}_{(N,v)}(S) = P(S, v_{|S})$ as a TU-game as well (see remark 2.8 of *Hart* and *Mas-Colell* (1989)). Furthermore, they show that $P^{HM}_{(N,v)} = \sum_{R \subseteq N} \frac{\lambda_R(v)}{|R|} u_R$. If there is no ambiguity about the underlying game we will simply write P^{HM} instead of $P^{HM}_{(N,v)}$. Note that a cooperative game completely determines its associated potential game and vice versa. For convenience we will sometimes refer to an associated potential game without specifying the underlying cooperative game. Finally, to avoid confusion with non-cooperative potential games, we will refer to potential games associated with cooperative games as HM-potential games.

A game in strategic form will be denoted by $\Gamma = (N; (X_i)_{i \in N}; (\pi_i)_{i \in N})$, where $N = \{1, \ldots, n\}$ denotes the player set, X_i the strategy space of player $i \in N$, and $\pi = (\pi_i)_{i \in N}$ the payoff function which assigns to every strategy-tuple $x = (x_i)_{i \in N} \in \prod_{i \in N} X_i = X$ a vector in $\mathbb{R}^N$. For notational convenience we write $x_{-i} = (x_l)_{l \in N \setminus \{i\}}$, $x_{-ij} = (x_l)_{l \in N \setminus \{i,j\}}$, and $x_R = (x_l)_{l \in R}$.

Monderer and *Shapley* (1996) formally defined the class of non-cooperative potential games. A function $P : \prod_{i \in N} X_i \to \mathbb{R}$ is called a *potential* for Γ if for every $i \in N$, every $x \in X$, and every $t_i \in X_i$ it holds that

$$\pi_i(x_i, x_{-i}) - \pi_i(t_i, x_{-i}) = P(x_i, x_{-i}) - P(t_i, x_{-i}). \tag{2}$$

The game Γ is called a potential game if it admits a potential.

The following set of collections of cooperative games forms the basis for a representation theorem of potential games.

$$\mathcal{G}_{N,X} := \Big\{ \{(N, v_x)\}_{x \in X} \in (TU^N)^X \mid v_x(R) = v_t(R) \text{ if } x_R = t_R \text{ for all } x, t \in X,\ R \subseteq N \Big\}. \tag{3}$$

This representation theorem (cf. *Ui* (1996)) describes a relation between non-cooperative potential games and Shapley values of cooperative games.

Theorem 2.2 Let $\Gamma = (N; (X_i)_{i \in N}; (\pi_i)_{i \in N})$ be a game in strategic form. Γ is a potential game if and only if there exists $\{(N, v_x)\}_{x \in X} \in \mathcal{G}_{N,X}$ such that

$$\pi_i(x) = \Phi_i(v_x) \text{ for all } i \in N \text{ and all } x \in X. \tag{4}$$

Proof: See *Ui* (1996). □

3 A model of coalition formation

In this section we will describe two models of coalition formation. We describe the model of coalition formation of *Meca-Martinez et al.* (1998) and a slight modification of this model. We will argue that this modification does not affect their results.

Both models of coalition formation we analyze in this section can be seen as two-stage models. We assume that a cooperative game (N, v) is exogenously given. In the first stage, each player announces the coalition he wants to join. Depending on the announcements of the players a coalition structure results. In the second stage, players negotiate over the division of the surplus, given the coalition structure of the first stage. This stage is modeled by means of an allocation rule.

Firstly, we will describe the model of coalition formation that is a slight modification of the model of coalition formation of *Meca-Martinez et al.* (1998). Formally, this model of coalition formation is given by $\Gamma(N,v,\gamma) = (N;(X_i)_{i\in N};(f_i^\gamma)_{i\in N})$ where for all $i \in N$

$$X_i = \{T \subseteq N \mid i \in T\}$$

represents the strategy space of player i. A strategy of a player is interpreted as the partition element this player wants to be in. A strategy profile $x = (x_1,\ldots,x_n)$ induces a cooperation structure $\mathcal{B}(x) = \{B_1,\ldots,B_m\}$ where players i and j are in the same partition element if and only if $x_i = x_j$, i.e., if they prefer the same partition element. Note that this condition can only be satisfied if $i \in x_j$ and $j \in x_i$. The payoff function $f^\gamma = (f_i^\gamma)_{i\in N}$ is then defined as the allocation rule γ applied to the cooperation structure that results,

$$f^\gamma(x) = \gamma(N,v,\mathcal{B}(x)).$$

In case there is no ambiguity on the underlying game we will simply write $\Gamma(\gamma)$ instead of $\Gamma(N,v,\gamma)$. For notational convenience we define for all $T \subseteq N$ and all $x_T \in \prod_{i\in T} X_i$, $\mathcal{B}(x_T)$ as the partition of T where players i and j are in the same partition element if and only if $x_i = x_j$.

Meca-Martinez et al. (1998) model the formation of a coalition slightly different. In their model a strategy profile $x = (x_1,\ldots,x_n)$ induces a cooperation structure $\mathcal{B}^M(x) = \{B_1,\ldots,B_m\}$ where player i ends up in coalition x_i if and only if $x_j = x_i$ for all $j \in x_i$, i.e., all players in the coalition prefered by player i prefer that coalition. If $x_j \neq x_i$ for some $j \in x_i$ player i ends up isolated. The formulation of *Meca-Martinez et al.* (1998) implies that a player ends up either isolated or in the coalition that he chose. In our formulation he can also end up in a subset of the coalition that he chose. For notational convenience we define for all $T \subseteq N$ and all $x_T \in \prod_{i\in T} X_i$, $\mathcal{B}^M(x_T)$ as the partition of T where player i is in partition element x_i if $x_i \subseteq T$ and $x_j = x_i$ for all $j \in x_i$. Otherwise, player i ends up isolated.

The difference between the two models of coalition structure formation is illustrated in the following example.

Example 3.1 Let (N,v) be a 3-person cooperative game and γ some allocation rule for cooperative games with coalition structures. Consider $\Gamma(N,v,\gamma)$ and assume the players have chosen the following strategies: $x_1 = N$, $x_2 = N$, and $x_3 = \{2,3\}$. Then $\mathcal{B}(x) = \{\{1,2\},\{3\}\}$ since $x_1 = x_2 \neq x_3$. However, in the model of *Meca-Martinez et al.* (1998) the resulting coalition structure is $\mathcal{B}^M(x) = \{\{1\},\{2\},\{3\}\}$ since $3 \in x_1 = x_2$ and $x_3 \neq x_1 = x_2$. Note that in the model of *Meca-Martinez et al.* (1998) player 3 can influence whether players 1 and 2 end up in the same partition element. Player 3 does not have this influence in our model $\Gamma(N, v, \gamma)$.

We will refer to the model of *Meca-Martinez et al.* (1998) with underlying cooperative game (N, v) and allocation rule γ by $\Gamma^M(N, v, \gamma)$. The distinction between the model of *Meca-Martinez et al.* (1998) and our model is exactly the same as the distinction between models γ and δ of *Hart* and *Kurz* (1983). In fact, the only difference between model γ of *Hart* and *Kurz* (1983) and the model of *Meca-Martinez et al.* (1998) is that they analyze different allocation rules. The difference between model δ of *Hart* and *Kurz* (1983) and our model is of a similar nature.

We will show that the results of *Meca-Martinez et al.* (1998) also hold for our model. The allocation rules they study have the property that in a game with coalition structure $(N, v, \mathcal{B})$ the payoff for player $i \in B \in \mathcal{B}$ depends only on $(v(S))_{S \subseteq B}$. We will call such an allocation rule a *component restricted allocation rule*.[2] Consider the following properties for a component restricted allocation rule γ:

Weak Monotonicity (WM) For all cooperative games with coalition structures $(N, v, \mathcal{B})$ and $(N, w, \mathcal{B})$ it holds for all $B \in \mathcal{B}$ that if

$$v(S \cup \{i\}) - v(S) \geq w(S \cup \{i\}) - w(S) \text{ for all } i \in B \text{ and all } S \subseteq B \backslash \{i\}$$

then $\gamma_i(N, v, \mathcal{B}) \geq \gamma_i(N, w, \mathcal{B})$ for all $i \in B$.

Dummy Out (DO) For all cooperative games (N, v) and every partition $\mathcal{B} = \{B_1, \ldots, B_m\}$ of N it holds for every $k \in \{1, \ldots, m\}$ and every $i \in B_k$ which is a dummy player in the game $(B_k, v_{|B_k})$ that

$$\gamma(N, v, \mathcal{B}) = \gamma(N, v, \mathcal{B} - i).$$

The following lemma corresponds to lemma 1 of *Meca-Martinez et al.* (1998).

Lemma 3.1 Let (N, v) be a convex game and γ a component restricted allocation rule satisfying Weak Monotonicity and Dummy Out. Then for every partition $\mathcal{B} = \{B_1, \ldots, B_m\}$ of N, every $k \in \{1, \ldots, m\}$, and every $S \subseteq B_k$ it holds that for all $i \in S$

$$\gamma_i(N, v, \mathcal{B}) \geq \gamma_i(N, v, \{B_1, \ldots, B_{k-1}, S, \{j\}_{j \in B_k \backslash S}, B_{k+1}, \ldots, B_m\}).$$

[2] *Meca-Martinez et al.* (1998) simply look at the game $(B, v_{|B})$ and allocation rules for cooperative games. This corresponds to restricting to component restricted allocation rules. We have changed the properties accordingly.

Proof: Let $\mathcal{B} = \{B_1, \ldots, B_m\} \in \Pi^N$, $k \in \{1, \ldots, m\}$, and let $S \subseteq B_k$. Define

$$v^S(T) = v(S \cap T) \text{ for all } T \subseteq N;$$

$$\hat{v}^S(T) = v^S(T) + \sum_{i \in T \setminus S} v(\{i\}) \text{ for all } T \subseteq N.$$

Now, since v is convex,

$$v(T \cup \{i\}) - v(T) \geq v^S(T \cup \{i\}) - v^S(T) = \hat{v}^S(T \cup \{i\}) - \hat{v}^S(T) \quad (5)$$

for all $i \in S$ and all $T \subseteq B_k \setminus \{i\}$. Also, for all $i \in B_k \setminus S$ and all $T \subseteq B_k \setminus \{i\}$,

$$v(T \cup \{i\}) - v(T) \geq v(\{i\}) = \hat{v}^S(T \cup \{i\}) - \hat{v}^S(T). \quad (6)$$

We conclude from (5) and (6) that

$$v(T \cup \{i\}) - v(T) \geq \hat{v}^S(T \cup \{i\}) - \hat{v}^S(T)$$

for all $i \in B_k$ and all $T \subseteq B_k \setminus \{i\}$. Then, since γ is a component restricted allocation rule that satisfies (WM), it follows that $\gamma_i(N, v, \mathcal{B}) \geq \gamma_i(N, \hat{v}^S, \mathcal{B})$ for all $i \in B_k$. Note that for all $j \in B_k \setminus S$, j is a dummy player in $(N, \hat{v}^S)$. Hence, j is a dummy player in $(T, \hat{v}^S_{|T})$ for all $T \subseteq N$ with $j \in T$. Specifically, j is a dummy player in $(T, \hat{v}^S_{|T})$ for all $T = S \cup U$, $U \subseteq B_k \setminus S$ with $j \in U$. By repeated application of (DO) for all $j \in B_k \setminus S$ it follows that $\gamma_i(N, \hat{v}^S, \mathcal{B}) = \gamma_i(N, \hat{v}^S, \{B_1, \ldots, B_{k-1}, S, \{j\}_{j \in B_k \setminus S}, B_{k+1}, \ldots, B_m\})$ for all $i \in S$. Now, since $(\hat{v}^S)_{|S} = v_{|S}$ and γ is a component restricted allocation rule, it follows that $\gamma_i(N, \hat{v}^S, \{B_1, \ldots, B_{k-1}, S, \{j\}_{j \in B_k \setminus S}, B_{k+1}, \ldots, B_m\}) = \gamma_i(N, v, \{B_1, \ldots, B_{k-1}, S, \{j\}_{j \in B_k \setminus S}, B_{k+1}, \ldots, B_m\})$ for all $i \in S$.

We conclude that $\gamma_i(N, v, \mathcal{B}) \geq \gamma_i(N, v, \{B_1, \ldots, B_{k-1}, S, \{j\}_{j \in B_k \setminus S}, B_{k+1}, \ldots, B_m\})$ for all $i \in S$. □

It is now straightforward to show that theorems 1 and 2 of *Meca-Martinez et al.* (1998) also hold for our model. Recall that $x \in X$ is a strong Nash equilibrium of $\Gamma(\gamma)$ if there is no coalition $T \subseteq N$ and strategy profile $\hat{x}_T$ such that $f_i^\gamma(\hat{x}_T, x_{N \setminus T}) \geq f_i^\gamma(x)$ for all $i \in T$, with the inequality being strict for at least one player $i \in T$. We denote the set of strong Nash equilibria by $\text{SNE}(\Gamma(\gamma))$. Furthermore, we define the set of coalition structures that result according to strong Nash equilibria (strong Nash equilibrium partitions):

$$\text{SNEP}(\Gamma(\gamma)) := \{\mathcal{B} \in \Pi^N \mid \exists x \in \text{SNE}(\Gamma(\gamma)) : \mathcal{B}(x) = \mathcal{B}\}.$$

The following theorem corresponds to theorems 1 and 2 in *Meca-Martinez et al.* (1998).

Theorem 3.1 Let (N, v) be a convex game and γ a component restricted allocation rule satisfying (WM) and (DO). Then $\{N\} \in \text{SNEP}(\Gamma(\gamma))$ and for all $\mathcal{B} \in \text{SNEP}(\Gamma(\gamma))$ it holds that $\gamma(N, v, \mathcal{B}) = \gamma(N, v, \{N\})$.

Proof: Let $x \in X$ be such that $\gamma(N, v, \mathcal{B}(x)) = \gamma(N, v, \{N\})$. Let $t \in X$, $i \in N$, and $B \in \mathcal{B}(t)$ with $i \in B$. Since γ is a component restricted allocation rule it follows that

$$\gamma_i(N, v, \mathcal{B}(t)) = \gamma_i(N, v, \{B, \{j\}_{j \in N \setminus B}\}) \leq \gamma_i(N, v, \{N\}) = \gamma_i(N, v, \mathcal{B}(x)), \qquad (7)$$

where the inequality follows by lemma 3.1. We conclude that $x \in \text{SNE}(\Gamma(\gamma))$.

Let $x \in X$ be such that $\gamma(N, v, \mathcal{B}(x)) \neq \gamma(N, v, \{N\})$. By (7) it follows that for all $i \in N$ it holds that $\gamma_i(N, v, \mathcal{B}(x)) \leq \gamma_i(N, v, \{N\})$. Hence, the deviation to $t_i = N$ for all $i \in N$ weakly improves the payoff for all players with a strict improvement for at least one player. So, $x \notin \text{SNE}(\Gamma(\gamma))$.

This completes the proof. □

Note that in fact we prove a somewhat stronger result: $\mathcal{B} \in \text{SNEP}(\Gamma(\gamma))$ if only if $\gamma(N, v, \mathcal{B}) = \gamma(N, v, \{N\})$. A similar strengthening is possible in the original model of *Meca-Martinez et al.* (1998).

4 Potential games

In this section we study under what conditions on the allocation rule the two models of coalition formation result in a potential game. We will show that under an efficiency requirement our model of coalition formation is a potential game if and only if the value of *Aumann* and *Drèze* (1974) is used as an allocation rule, whereas the original model of *Meca-Martinez et al.* (1998) is a potential game if and only if an allocation rule that equally divides the gains over the sum of stand-alone values is used. Furthermore, we describe for the model of coalition formation of *Meca-Martinez et al.* (1998) the potential maximizing strategy profiles.

Firstly, we study $\Gamma(N, v, \gamma)$, the model of coalition formation with the less stringent formation rule. We show that the value of *Aumann* and *Drèze* (1974) is the unique component efficient allocation rule that results in a coalition formation game that is a potential game. To accomplish this, we need two lemma's.

Lemma 4.1 Let γ be a component efficient allocation rule. Let (N, v) be a cooperative game. If the associated coalition formation game $\Gamma(N, v, \gamma)$ is a potential game then for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$, and all $i, j \in B$ it holds that

$$\gamma_i(\mathcal{B}) - \gamma_i(\mathcal{B} - j) = \gamma_j(\mathcal{B}) - \gamma_j(\mathcal{B} - i). \tag{8}$$

Proof: Let P be a potential for $\Gamma(N, v, \gamma)$. Let $\mathcal{B} \in \Pi^N$, $B \in \mathcal{B}$, and $i, j \in B$. Let x be a strategy profile that results in partition $\mathcal{B}$, i.e., $\mathcal{B}(x) = \mathcal{B}$. Define $t_i = \{i\}$ and $t_j = \{j\}$. Then

$$\begin{aligned} 0 &= P(x) - P(x_{-i}, t_i) + P(x_{-i}, t_i) - P(x_{-ij}, t_i, t_j) \\ &\quad + P(x_{-ij}, t_i, t_j) - P(x_{-j}, t_j) + P(x_{-j}, t_j) - P(x) \\ &= (\gamma_i(\mathcal{B}) - \gamma_i(\mathcal{B} - i)) + (\gamma_j(\mathcal{B} - i) - \gamma_j(\mathcal{B} - j - i)) \\ &\quad + (\gamma_i(\mathcal{B} - j - i) - \gamma_i(\mathcal{B} - j)) + (\gamma_j(\mathcal{B} - j) - \gamma_j(\mathcal{B})) \\ &= \gamma_i(\mathcal{B}) - v(\{i\}) + \gamma_j(\mathcal{B} - i) - v(\{j\}) + v(\{i\}) - \gamma_i(\mathcal{B} - j) + v(\{j\}) - \gamma_j(\mathcal{B}) \\ &= \gamma_i(\mathcal{B}) - \gamma_i(\mathcal{B} - j) - \gamma_j(\mathcal{B}) + \gamma_j(\mathcal{B} - i), \end{aligned} \tag{9}$$

where the second equality follows by definition of a potential and the third equality follows since $\gamma_i(\mathcal{B} - i) = \gamma_i(\mathcal{B} - i - j) = v(\{i\})$ by component efficiency. Equation (9) implies equation (8). This completes the proof. □

In the following lemma we show that if the value of *Aumann* and *Drèze* (1974) is applied as an allocation rule then the coalition formation game is a potential game.

Lemma 4.2 Let (N, v) be a cooperative game. The coalition formation game $\Gamma(N, v, \Phi^{AD})$ is a potential game.

Proof: For all $x \in X$ and all $R \subseteq N$ define

$$v_x(R) := \sum_{B \in \mathcal{B}(x)} v(B \cap R).$$

Then $(N, v_x)_{x \in X} \in \mathcal{G}_{N,X}$ since $v_x(R)$ depends only on x_R. Furthermore, $\Phi(N, v_x) = \Phi^{AD}(N, v, \mathcal{B}(x))$ for all $x \in X$. This follows directly by noting that Φ^{AD} can be found by computing the Shapley value for the subgames restricted to the partition elements. Now, theorem 2.2 completes the proof. □

Combining the lemmas above we can prove that the value of *Aumann* and *Drèze* (1974) is the unique allocation rule that results in a potential game.

Theorem 4.1 Let (N, v) be a cooperative game. Let γ be a component efficient allocation rule. The coalition formation game $\Gamma(N, v, \gamma)$ is a potential game if and only if γ coincides with the value of *Aumann* and *Drèze* (1974) for all partitions of N.

Proof: Suppose that the coalition formation game $\Gamma(\gamma)$ is a potential game. Lemma 4.1 implies that for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$, and all $i, j \in B$ equation (8) holds. It can then be shown analogously to the proof of theorem 2.1 that γ coincides with the value of *Aumann* and *Drèze* (1974) for all partitions of N.[3]

The reverse statement follows by lemma 4.2. □

The following example shows that the value of *Aumann* and *Drèze* (1974) does not result in a potential game in the model of *Meca-Martinez et al.* (1998).

Example 4.1 Consider the TU-game (N, v) with $N = \{1, 2, 3\}$ and

$$v(S) = \begin{cases} 0 & \text{if } |S| \leq 1; \\ 40 & \text{if } S = \{1,2\}; \\ 50 & \text{if } S = \{1,3\}; \\ 60 & \text{if } S = \{2,3\}; \\ 72 & \text{if } S = N. \end{cases} \tag{10}$$

Suppose that player 3 plays strategy $x_3 = \{1, 2, 3\}$. Then part of the payoff-matrix of $\Gamma^M(N, v, \Phi^{AD})$ is given below.

	$x_2 = \{2,3\}$	$t_2 = \{1,2,3\}$
$x_1 = \{1,3\}$	(0,0,0)	(0,0,0)
$t_1 = \{1,2,3\}$	(0,0,0)	(19,24,29)

If $\Gamma^M(N, v, \Phi^{AD})$ is a potential game it should hold that there exists a potential P such that

$$\begin{aligned} 0 &= P(x) - P(t_1, x_2, x_3) + P(t_1, x_2, x_3) - P(t_1, t_2, x_3) \\ &\quad + P(t_1, t_2, x_3) - P(x_1, t_2, x_3) + P(x_1, t_2, x_3) - P(x) \\ &= \Big(\Phi_1^{AD}(\mathcal{B}^M(x)) - \Phi_1^{AD}(\mathcal{B}^M(t_1, x_2, x_3))\Big) \\ &\quad + \Big(\Phi_2^{AD}(\mathcal{B}^M(t_1, x_2, x_3)) - \Phi_2^{AD}(\mathcal{B}^M(t_1, t_2, x_3))\Big) \end{aligned}$$

[3]It only follows that γ satisfies (CRBC) for the game (N, v). Therefore, we cannot use theorem 2.1 directly, since this requires (CRBC) for all cooperative games. Careful reading of the proof, however, reveals that (CRBC) for (N, v) is enough.

$$\begin{aligned}
&\quad +\Big(\Phi_1^{AD}(\mathcal{B}^M(t_1,t_2,x_3)) - \Phi_1^{AD}(\mathcal{B}^M(x_1,t_2,x_3))\Big)\\
&\quad +\Big(\Phi_2^{AD}(\mathcal{B}^M(x_1,t_2,x_3)) - \Phi_2^{AD}(\mathcal{B}^M(x))\Big)\\
&= (0-0)+(0-24)+(19-0)+(0-0)\\
&\neq 0, \qquad (11)
\end{aligned}$$

where the second equality follows by definition (2) of a potential and the third equality by the payoffs above. We conclude that $\Gamma^M(N, v, \Phi^{AD})$ is not a potential game.

In the following theorem we show that the model of *Meca-Martinez et al.* (1998) is a potential game if and only if every coalition divides the surplus of the coalition over the sum of stand-alone values equally among the players in this coalition.

Theorem 4.2 Let (N, v) be a cooperative game. Let γ be a component efficient allocation rule. The coalition formation game $\Gamma^M(N, v, \gamma)$ is a potential game if and only if for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$, and all $i \in B$ it holds that $\gamma_i(\mathcal{B}) = v(\{i\}) + \frac{v(B)-\sum_{j\in B} v(\{j\})}{|B|}$.

Proof: First we show the only-if-part. Assume that $\Gamma^M(N, v, \gamma)$ is a potential game with associated potential P. We will show that for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$, and all $i \in B$ it holds that

$$\gamma_i(\mathcal{B}) = v(\{i\}) + \frac{v(B) - \sum_{j\in B} v(\{j\})}{|B|} \qquad (12)$$

Obviously, by component efficiency (12) holds for all $\mathcal{B} \in \Pi^N$, all $B \in \mathcal{B}$ with $|B| = 1$, and $i \in B$. Let $p \geq 2$. We will show that (12) holds for all $\mathcal{B} \in \Pi^N$, all B with $|B| = p$, and all $i \in B$. Let $\mathcal{B} \in \Pi^N$ and $B \in \mathcal{B}$ such that $|B| = p$. Let $x \in X$ be a strategy profile that results in partition $\mathcal{B}$, i.e., $\mathcal{B}^M(x) = \mathcal{B}$. Let $i, j \in B$ and define $t_i = \{i\}$ and $t_j = \{j\}$. Then

$$\begin{aligned}
0 &= P(x) - P(x_{-i},t_i) + P(x_{-i},t_i) - P(x_{-ij},t_i,t_j) + P(x_{-ij},t_i,t_j) - P(x_{-j},t_j)\\
&\quad +P(x_{-j},t_j) - P(x)\\
&= (\gamma_i(\mathcal{B}) - v(\{i\})) + (v(\{j\}) - v(\{j\})) + (v(\{i\}) - v(\{i\}))\\
&\quad +(v(\{j\}) - \gamma_j(\mathcal{B})). \qquad (13)
\end{aligned}$$

The second equality follows by definition of a potential P. We conclude that $\gamma_i(\mathcal{B}) - v(\{i\}) = \gamma_j(\mathcal{B}) - v(\{j\})$. Since i and j were chosen arbitrarily in B it follows by component efficiency that for all $i \in B$

$$\gamma_i(\mathcal{B}) = v(\{i\}) + \frac{v(B) - \sum_{j\in B} v(\{j\})}{|B|}.$$

This completes the only-if-part.

It remains to show the if-part. Let γ be the allocation rule determined by

$$\gamma_i(\mathcal{B}) = v(\{i\}) + \frac{v(B) - \sum_{j \in B} v(\{j\})}{|B|}, \text{ for all } \mathcal{B} \in \Pi^N, \text{ all } B \in \mathcal{B}, \text{ and all } i \in B.$$

Define for all $x \in X$:

$$v_x = \sum_{i \in N} v(\{i\}) u_{\{i\}} + \sum_{B \in \mathcal{B}^M(x)} \left[v(B) - \sum_{i \in B} v(\{i\}) \right] u_B.$$

Hence, for all $x \in X$ and all $T \subseteq N$

$$\begin{aligned} v_x(T) &= \sum_{i \in T} v(\{i\}) + \sum_{B \in \mathcal{B}^M(x), B \subseteq T} \left[v(B) - \sum_{i \in B} v(\{i\}) \right] \\ &= \sum_{B \in \mathcal{B}^M(x), B \subseteq T} v(B) + \sum_{B \in \mathcal{B}^M(x), B \not\subseteq T} \sum_{i \in T \cap B} v(\{i\}) \\ &= \sum_{B \in \mathcal{B}^M(x_T)} v(B). \end{aligned}$$

So, the value of coalition $T \subseteq N$ in the game corresponding to strategy profile $x \in X$ depends only on the strategies of the players in coalition T and hence,

$$\{(N, v_x)\}_{x \in X} \in \mathcal{G}_{N,X}.$$

Furthermore, for all $x \in X$

$$\Phi_i(N, v_x) = v(\{i\}) + \frac{v(B) - \sum_{j \in B} v(\{j\})}{|B|} \text{ for all } B \in \mathcal{B}^M(x) \text{ and all } i \in B.$$

By theorem 2.2 it follows that $\Gamma^M(N, v, \gamma)$ is a potential game.

This completes the if-part. □

For a zero-normalized game this implies that the model of *Meca-Martinez et al.* (1998) is a potential game if and only if the value of each partition element is divided equally among its members. Since for zero-normalized games the model of *Meca-Martinez et al.* (1998) with this allocation rule coincides with the model of coalition formation of *Von Neumann* and *Morgenstern* (1944), we conclude that the model of *Von Neumann* and *Morgenstern* (1944) is a potential game if the underlying game is zero-normalized. It can be checked that for zero-normalized games a potential is given by $P(x) = \sum_{B \in \mathcal{B}^M(x)} \frac{v(B)}{|B|}$ for all $x \in X$. Analyzing this associated potential function then implies that according to the potential maximizer, a strategy x will be chosen by the players such that

$$\sum_{B \in \mathcal{B}^M(x)} \frac{v(B)}{|B|} = \max_{\mathcal{B} \in \Pi^N} \sum_{B \in \mathcal{B}} \frac{v(B)}{|B|}.$$

The potential maximizer selects a coalition structure that maximizes the sum over all partition elements of the payoffs each player in a partition element receives.

5 Potential Maximizers

In this section we will consider potential maximizing strategies in the coalition formation game $\Gamma(N, v, \Phi^{AD})$ with (N, v) a superadditive underlying cooperative game. We will show that the cooperation structure with all players in one component results from a potential maximizing strategy profile. Subsequently, we will show that every potential maximizing strategy profile results in a cooperation structure that is payoff equivalent to this structure.

Before we can show that cooperation structure $\{N\}$ results from a potential maximizing strategy profile we need some results on cooperative HM-potential games, cf. *Hart* and *Mas-Colell* (1989).

Theorem 5.1 Let (N, v) be a superadditive game. Then the associated HM-potential game (N, P^{HM}) is also superadditive.

Proof: *Hart* and *Mas-Colell* (1989) showed that (N, P^{HM}) satisfies $P^{HM}(\emptyset) = 0$ and

$$P^{HM}(S) = \frac{1}{|S|}\left[v(S) + \sum_{k \in S} P^{HM}(S \backslash \{k\})\right] \text{ for all } S \subseteq N,\ S \neq \emptyset. \tag{14}$$

We have to show that for all $S \subseteq N$ and all $T \subseteq N \backslash S$ it holds that

$$P^{HM}(S \cup T) \geq P^{HM}(S) + P^{HM}(T). \tag{15}$$

The proof will be by induction to the number of elements in $S \cup T$. Obviously, (15) holds for all S, T with $|S \cup T| = 0$, since $P^{HM}(\emptyset) = 0$. Let $p \geq 1$. Assume that (15) holds for all S, T with $|S \cup T| \leq p - 1$. We will show that (15) holds for all S, T with $|S \cup T| = p$. Let $S \subseteq N$, $T \subseteq N \backslash S$ such that $|S \cup T| = p$. Then

$$\begin{aligned} P^{HM}(S \cup T) &= \frac{1}{|S \cup T|}\left[v(S \cup T) + \sum_{k \in S \cup T} P^{HM}((S \cup T) \backslash \{k\})\right] \\ &\geq \frac{1}{|S \cup T|}\Bigg[v(S) + v(T) + \sum_{k \in S}[P^{HM}(S \backslash \{k\}) + P^{HM}(T)] + \\ &\qquad \sum_{k \in T}[P^{HM}(S) + P^{HM}(T \backslash \{k\})]\Bigg] \\ &= \frac{1}{|S \cup T|}\Bigg[v(S) + \sum_{k \in S} P^{HM}(S \backslash \{k\}) + v(T) + \sum_{k \in T} P^{HM}(T \backslash \{k\}) + \\ &\qquad |S| P^{HM}(T) + |T| P^{HM}(S)\Bigg] \\ &= \frac{1}{|S \cup T|}\left[|S| P^{HM}(S) + |T| P^{HM}(T) + |S| P^{HM}(T) + |T| P^{HM}(S)\right] \\ &= P^{HM}(S) + P^{HM}(T), \end{aligned} \tag{16}$$

where the inequality follows from superadditivity of (N, v) and the induction hypothesis. The third equality follows by equation (14). This completes the proof. □

The following lemma shows that a potential for $\Gamma(\Phi^{AD})$ can be given in terms of the HM-potential game associated with the underlying cooperative game.

Lemma 5.1 Let (N, v) be a cooperative game. Then P with

$$P(x) = \sum_{B \in \mathcal{B}(x)} P^{HM}_{(N,v)}(B) \text{ for all } x \in X \tag{17}$$

is a potential for the coalition formation game $\Gamma(N, v, \Phi^{AD})$.

Proof: We have to show that P is a potential. Therefore, consider $x \in X$, $i \in N$ and $u_i \in X_i$ with $x_i \neq u_i$. It suffices to check that $P(x) - P(x_{-i}, u_i) = \Phi_i^{AD}(\mathcal{B}(x)) - \Phi_i^{AD}(\mathcal{B}(x_{-i}, u_i))$. Denote by B_1^* and B_2^* the partition elements player i belongs to according to x and (x_{-i}, u_i) respectively. Then it follows from $x_i \neq u_i$ that $B_1^* \cap B_2^* = \{i\}$. Since $B_2^* \backslash \{i\} \in \mathcal{B}(x)$ or $B_2^* \backslash \{i\} = \emptyset$, and $B_1^* \backslash \{i\} \in \mathcal{B}(x_{-i}, u_i)$ or $B_1^* \backslash \{i\} = \emptyset$ it holds that[4]

$$\begin{aligned}
P(x) - P(x_{-i}, u_i) &= \sum_{B \in \mathcal{B}(x)} P^{HM}(B) - \sum_{B \in \mathcal{B}(x_{-i}, u_i)} P^{HM}(B) \\
&= \sum_{S \subseteq B_1^*} \frac{\lambda_S(v)}{|S|} + \sum_{S \subseteq B_2^* \backslash \{i\}} \frac{\lambda_S(v)}{|S|} + \sum_{B \in \mathcal{B}(x) \backslash \{B_1^*, B_2^* \backslash \{i\}\}} P^{HM}(B) \\
&\quad - \sum_{S \subseteq B_1^* \backslash \{i\}} \frac{\lambda_S(v)}{|S|} - \sum_{S \subseteq B_2^*} \frac{\lambda_S(v)}{|S|} - \sum_{B \in \mathcal{B}(x_{-i}, u_i) \backslash \{B_2^*, B_1^* \backslash \{i\}\}} P^{HM}(B) \\
&= \sum_{S \subseteq B_1^*} \frac{\lambda_S(v)}{|S|} - \sum_{S \subseteq B_1^* \backslash \{i\}} \frac{\lambda_S(v)}{|S|} + \sum_{S \subseteq B_2^* \backslash \{i\}} \frac{\lambda_S(v)}{|S|} - \sum_{S \subseteq B_2^*} \frac{\lambda_S(v)}{|S|} \\
&= \sum_{S \subseteq B_1^*, i \in S} \frac{\lambda_S(v)}{|S|} - \sum_{S \subseteq B_2^*, i \in S} \frac{\lambda_S(v)}{|S|} \\
&= \Phi_i^{AD}(N, v, \mathcal{B}(x)) - \Phi_i^{AD}(N, v, \mathcal{B}(x_{-i}, u_i)),
\end{aligned}$$

where the third equality holds since $\mathcal{B}(x) \backslash \{B_1^*, B_2^* \backslash \{i\}\} = \mathcal{B}(x_{-i}, u_i) \backslash \{B_2^*, B_1^* \backslash \{i\}\}$.

This completes the proof. □

For every player $i \in N$ we denote $\bar{x}_i = N$. Using the results above we can prove that the strategy profile $\bar{x} = (\bar{x}_i)_{i \in N}$ is a potential maximizing strategy profile.

[4] Recall that $P^{HM}_{(N,v)} = \sum_{R \subseteq N} \frac{\lambda_R(v)}{|R|} u_R$.

Theorem 5.2 Let (N, v) be a superadditive game and $\Gamma(\Phi^{AD})$ the associated coalition formation game with potential P. Then $\bar{x} \in \text{argmax}\, P$.

Proof: From lemma 5.1 it follows that a potential is given by

$$P(x) = \sum_{B \in \mathcal{B}(x)} P^{HM}(B) \text{ for all } x \in X.$$

Let $x \in X$. It follows from theorem 5.1 that

$$P(x) = \sum_{B \in \mathcal{B}(x)} P^{HM}(B) \leq P^{HM}(N) = P(\bar{x}).$$

Monderer and *Shapley* (1996) show that the set of potential maximizing strategy profiles does not depend on the choice of a particular potential. Hence, $\bar{x} \in \text{argmax}\, P'$ for every potential P'.

This completes the proof. □

Before we can show that every potential maximizing strategy profile results in the same payoffs as the strategy profile resulting in the unique component N, we need another lemma.

Lemma 5.2 Let (N, v) be a superadditive game. Then for all $S \subseteq N$ and all $T \subseteq N \backslash S$ it holds that

$$P^{HM}(S \cup T) = P^{HM}(S) + P^{HM}(T) \quad \Rightarrow v(U) = v(S \cap U) + v(T \cap U) \;\forall U \subseteq S \cup T. \quad (18)$$

Proof: First note that it follows from theorem 5.1 that (N, P^{HM}) is superadditive. We proceed by induction to the number of elements in $S \cup T$. Obviously, if $|S \cup T| = 0$ then (18) holds. Let $p \geq 1$. Suppose that (18) holds for all S, T with $|S \cup T| \leq p - 1$. We will show that (18) holds for all S, T with $|S \cup T| = p$. Let $S \subseteq N$ and $T \subseteq N \backslash S$ with $|S \cup T| = p$. Suppose $P^{HM}(S \cup T) = P^{HM}(S) + P^{HM}(T)$. Then since the inequality in equation (16) must hold with equality it follows that

$$v(S \cup T) = v(S) + v(T); \quad (19)$$

$$P^{HM}((S \cup T) \backslash \{k\}) = P^{HM}(S \backslash \{k\}) + P^{HM}(T) \;\forall k \in S; \quad (20)$$

$$P^{HM}((S \cup T) \backslash \{k\}) = P^{HM}(S) + P^{HM}(T \backslash \{k\}) \;\forall k \in T. \quad (21)$$

From the induction hypothesis and equations (20) and (21) it follows that for all $k \in S \cup T$

$$v(U) = v(S \cap U) + v(T \cap U) \;\forall U \subseteq (S \cup T) \backslash \{k\}. \quad (22)$$

Combining equations (19) and (22) completes the proof. □

Using the lemma above we can prove the following theorem.

Theorem 5.3 Let (N, v) be a superadditive game and $\Gamma(\Phi^{AD})$ the associated coalition formation game with potential P. Let $x \in \text{argmax}\, P$. Then $\Phi^{AD}(\mathcal{B}(x)) = \Phi^{AD}(\{N\})$.

Proof: Denote $\mathcal{B}(x) = (B_1, \ldots, B_m)$. If $\Gamma(\Phi^{AD})$ is a potential game then a potential is given by equation (17). Since the set of potential maximizing strategy profiles is independent of the specific potential, we can assume without loss of generality that P is given by equation (17). Then

$$\begin{aligned} P(\bar{x}) = P^{HM}(N) &\geq P^{HM}(\cup_{k=1}^{m-1} B_k) + P^{HM}(B_m) \\ &\geq \ldots \geq P^{HM}(B_1 \cup B_2) + \sum_{k=3}^{m} P^{HM}(B_k) \\ &\geq \sum_{k=1}^{m} P^{HM}(B_k) = P(x), \end{aligned} \tag{23}$$

where the equalities follow by lemma 5.1 and the inequalities follow by theorem 5.1. Since $x \in \text{argmax}\, P$ all inequalities hold with equality.

Let $U \subseteq N$ then, using lemma 5.2, the following equalities are implied by the corresponding equalities in equation (23).

$$\begin{aligned} v(U) &= v(U \cap (\cup_{k=1}^{m-1} B_k)) + v(U \cap B_m) \\ &= \ldots = v(U \cap (B_1 \cup B_2)) + \sum_{k=3}^{m} v(U \cap B_k) \\ &= \sum_{k=1}^{m} v(U \cap B_k). \end{aligned} \tag{24}$$

So, for all $U \subseteq N$

$$v(U) = \sum_{B \in \mathcal{B}(x)} v(B \cap U). \tag{25}$$

Equation (25) implies that $\Phi^{AD}(\mathcal{B}(x)) = \Phi^{AD}(\{N\})$. This completes the proof. □

In the following example we will show that not every strategy profile that results in the same payoffs as the full cooperation structure is potential maximizing.

Example 5.1 Let (N, v) be a 4-person superadditive cooperative game with $N = \{1, 2, 3, 4\}$ and

$$v(S) = \begin{cases} 0 & \text{if } |S| \leq 1; \\ 2 & \text{if } |S| = 2; \\ 3 & \text{if } |S| = 3; \\ 4 & \text{if } S = N. \end{cases} \tag{26}$$

Then some straightforward calculations show that $\Phi_i^{AD}(\{N\}) = 1$ for all $i \in N$ and $\Phi_i^{AD}((\{1, 2\}, \{3, 4\})) = 1$ for all $i \in N$. Since $v = \sum_{S:|S|=2} 2u_S - \sum_{S:|S|=3} 3u_S + 4u_N$ it follows that $P^{HM} = \sum_{S:|S|=2} u_S - \sum_{S:|S|=3} u_S + u_N$, i.e.,

$$P^{HM}(S) = \begin{cases} 0 & \text{if } |S| \leq 1; \\ 1 & \text{if } |S| = 2; \\ 2 & \text{if } |S| = 3; \\ 3 & \text{if } S = N. \end{cases} \tag{27}$$

Hence, $P^{HM}(N) > P^{HM}(\{1, 2\}) + P^{HM}(\{3, 4\})$. Then by lemma 5.1 it follows that $P(\bar{x}) > P(x)$, with $x = (\{1, 2\}, \{1, 2\}, \{3, 4\}, \{3, 4\})$, showing that not every strategy profile that results in the same payoffs as the full cooperation structure maximizes the potential function.

6 Conclusions

In this paper we have studied two models of coalition formation. The models differ only in the formation rule. We showed that the results of *Meca-Martinez et al.* (1998) also hold for our model. These results deal with sufficient conditions on an allocation rule to ensure that the grand coalition results from a strong Nash equilibrium when the original TU-game is convex.

Subsequently, we showed that under an efficiency requirement, our model of coalition formation is a potential game if and only if the value of *Aumann* and *Drèze* (1974) is used as an allocation rule. The model of coalition formation of *Meca-Martinez et al.* (1998) is a potential game if and only if an allocation rule that equally divides the surplus over the sum of stand-alone values is used.

Finally, we showed that if the underlying cooperative game is superadditive then the potential maximizer in our model with the value of *Aumann* and *Drèze* (1974) used as an allocation rule points towards the formation of the grand coalition. This result is in line with results of *Qin* (1996) and *Slikker et al.* (1999) who deal with link formation and conference formation respectively. However, all these results are sensitive to the superadditivity assumption and the equilibrium refinement chosen.

References

Aumann, R. and J. Drèze (1974). *Cooperative games with coalition structures.* International Journal of Game Theory 3, 217-237.

Hart, S. and M. Kurz (1983). *Endogenous Formation of Coalitions.* Econometrica 51, 1047-1061.

Hart, S. and A. Mas-Colell (1989). *Potential, Value and Consistency.* Econometrica 57, 589-614.

Meca-Martínez, A., J. Sánchez-Soriano, I. García-Jurado, and S. Tijs (1998). *Strong equilibria in claim games corresponding to convex games.* International Journal of Game Theory 27, 211-217.

Myerson, R. (1977). *Graphs and cooperation in games.* Mathematics of Operations Research 2, 225-229.

Myerson, R. (1980). *Conference structures and fair allocation rules.* International Journal of Game Theory 9, 169-182.

Myerson, R. (1991). *Game Theory: Analysis of Conflict.* Harvard University Press, Cambridge, Massachusetts.

Monderer, D. and L. Shapley (1996). *Potential games.* Games and Economic Behavior 14, 124-143.

Owen, G. (1977). *Values of games with a priori unions.* In: Essays in Mathematical Economics and Game Theory (Eds. R. Henn and O. Moeschlin), Springer Verlag, Berlin, Germany, 76-88.

Qin, C. (1996). *Endogenous formation of cooperation structures.* Journal of Economic Theory 69, 218-226.

Shapley, L. (1953). *A value for n-person games.* Contributions to the theory of games II (Eds. A. Tucker and H. Kuhn), 307-317.

Slikker, M., B. Dutta, A. van den Nouweland and S. Tijs (1999). *Potential maximizers and network formation.* Mathematical Social Sciences (to appear).

Ui, T. (1996). *A Shapley value representation of potential games.* Working paper, Department of Economics, Stanford University.

Von Neumann, J. and O. Morgenstern (1944). *Theory of games and economic behavior.* Princeton University Press, Princeton, New Jersey.