Soc Choice Welfare (2002) 19: 241–263

Social Choice and Welfare

Representation of effectivity functions in coalition proof Nash equilibrium: A complete characterization

Hans Keiding[1], Bezalel Peleg[2]

[1] Institute of Economics, University of Copenhagen, Studiestraede 6, 1455 Copenhagen K., Denmark (e-mail: Hans.Keiding@pop.oko.ku.dk)
[2] Institute of Mathematics and Center for Rationality and Interactive Decision Theory, Hebrew University of Jerusalem

Received: 24 June 1999/Accepted: 20 September 2000

Abstract. The concept of coalition proof Nash equilibrium was introduced by Bernheim et al. [5]. In the present paper, we consider the representation problem for coalition proof Nash equilibrium: For a given effectivity function, describing the power structure or the system of rights of coalitions in society, it is investigated whether there is a game form which gives rise to this effectivity function and which is such that for any preference assignment, there is a coalition proof Nash equilibrium.

It is shown that the effectivity functions which can be represented in coalition proof Nash equilibrium are exactly those which satisfy the well-known properties of maximality and superadditivity. As a corollary of the result, we obtain necessary conditions for implementation of a social choice correspondence in coalition proof Nash equilibrium which can be formulated in terms of the associated effectivity function.

1 Introduction

The theory of implementation is concerned with the construction of rules for choice of alternatives in a society such that the equilibrium behavior of the individuals results in choices satisfying certain, preassigned properties. The implementation problem in its classical formulation, as found e.g. in Hurwicz [10] and Maskin [11], starts with a social choice function or correspondence and consists in the design of a game form, such that for all conceivable assignments of preferences to individuals, the equilibrium outcome of the game coincides with the outcome prescribed by the social choice correspondence.

In the characterization of social choice rules which are implementable in cooperative equilibria, the concept of an effectivity function introduced by

Moulin and Peleg [12] turned out to play an important role. An effectivity function is a formal description of a power structure in a society; it describes for each coalition alternative subsets of alternatives such that the coalition can force outcome to belong to these subsets. There are obvious ways of associating effectivity functions with social choice correspondences and game forms, and these effectivity functions will coincide in some important cases.

It may actually be argued, that the effectivity function is a concept of considerable independent interest. Indeed, it may be interpreted as a specification of the rights of the coalitions in society; the fact that a coalition of individuals has a right to demand that society's choice belongs to a particular subset of the alternatives may be described by effectivity of the coalition for this subset. Thus, effectivity functions describe systems of rights in the sense of Gärdenfors [8] or constitutions (cf. Peleg [15]).

Given an effectivity function, the representation problem (for a particular solution concept) consists in finding a game form with which the given effectivity function is associated, and such that for any preference profile, there exists an equilibrium (of the type considered). If the representation problem has a solution, then the specifications of rights described by the effectivity function may indeed be resolved in society by equilibrium behavior within some given rules, namely those of the game form. As a by-product, one obtains a social choice rule, namely the equilibrum outcomes at any profile, which is implemented by the game form, and which in interesting cases has the same effectivity function. The work by Moulin and Peleg [12] on implementation in strong Nash equilibrium may be restated in these terms: If an effectivity function satisfies the properties of maximality (see Sect. 4 for a definition) together with another one called stability, then it has a representation in strong Nash equilibrium.

In the present paper, we consider the representation problem in another and weaker equilibrium, namely that of coalition proof Nash equilibrium introduced by Bernheim et al. [5]. Loosely speaking, a choice of strategies by the individuals is a coalition proof Nash equilibrium if no coalition can find another strategy which gives a better outcome for its members given the choices of the others, provided that none of its subcoalitions defect from the new strategies in order to achieve something still better.

The above naive description does not quite capture the essence of the definition; the notion of a defection is not made clear. To do that, one has to define coalitional improvements in a recursive way as it will be done in Sect. 2 below. Note that although coalition proof Nash equilibrium might be considered as a cooperative solution concept, it is non-cooperative in its nature; it allows only for such coalitional actions which are self-enforcable since it is never in the interest of a subcoalition to defect.

In the present paper, we give a characterization of the effectivity functions which have a representation in coalition proof Nash equilibrium. It is shown that the properties of maximality, and superadditivity are both necessary and sufficient conditions for an effectivity function to be represented in coalition proof Nash equilibrium. These conditions on the effectivity function, which imply that it may be considered as the power structure inherent in a coalition proof Nash implementable social choice correspondence, are strictly weaker than the corresponding conditions related to strong Nash equilibrium as treated in Moulin and Peleg [12]. This is as it should be, since strong Nash equilibrium is a stronger concept than coalitional proof Nash equilibrium. But it is important to notice that the conditions are really quite weak; thus, representation in coalition proof Nash equilibrium is something which must obtain very generally.

In this work, we consider the effectivity function as the primitive concept of the analysis of implementation. The effectivity function is a description of power structure which does not exploit the notion of preference profiles. However, several social choice correspondences may give rise to the same effectivity function, being different representations of the same power structure. What we show is that if the effectivity function satisfies the three conditions of superadditivity, monotonicity, and maximality, then at least one of the social choice correspondences which represent the effectivity function is coalition proof Nash implementable. The effectivity function characterization does not therefore give us a method of checking implementability on the particular social choice correspondence, but it allows us the determine whether there is an equivalent social choice correspondence – equivalence being defined in terms of equality of underlying power structure – which is implementable.

The paper is organized as follows: In Sect. 2, we give the definitions of the necessary game theoretical concepts, including that of coalition proof Nash equilibrium. In Sect. 3 we define implementation in coalition proof Nash equilibrium, and in Sect. 4, effectivity functions are introduced and discussed, and we establish the first (necessity) part of our characterization result. In Sect. 5, for a given effectivity function satisfying the properties of maximality, monotonicity, and superadditivity, we consider a notion of cooperative solution which in a certain sense generalizes the core of an effectivity function as discussed in Moulin and Peleg [12]. This solution concept is used in Sect. 6 to define a particular game form. In Sect. 7, this game form is shown to represent the original effectivity function, thereby establishing the sufficiency part of our characterization theorem. A final Sect. 8 contains some concluding remarks as well as a discussion of other contributions to the literature and their relation to the present work.

2 Coalition proof Nash equilibria

In this section, we recall the definition of coalition proof Nash equilibria (Bernheim et al. [5]) and discuss some simple properties of this solution concept.

Let $N = \{1, \ldots, n\}$ with $n \geq 2$ be a set of *players*. A *coalition* is a nonempty subset of N. We denote by 2^N the set of all coalitions. If $S \in 2^N$ and for each $i \in S$, D^i is a nonempty set, then we denote by $\prod_{i \in S} D^i = D^S$ the Cartesian product of the sets D^i.

Definition 2.1. *An n-person game in strategic form is a 2n-tuple*

$$\Gamma = (\Sigma^1, \ldots, \Sigma^n; h^1, \ldots, h^n),$$

where Σ^i is a nonempty set for every $i \in N$, and h^i is a function from Σ^N to $\mathbf{R}$ *for all $i \in N$.*

Let $\Gamma = (\Sigma^1, \ldots, \Sigma^n; h^1, \ldots, h^n)$ be an n-person game in strategic form. Then Σ^i is the set of *strategies* of player i, $i \in N$, and h^i is i's *payoff function* for every $i \in N$.

We now recall the definitions of Nash and strong Nash equilibria.

Definition 2.2. *Let $\Gamma = (\Sigma^1, \ldots, \Sigma^n; h^1, \ldots, h^n)$ be an n-person game, let $\sigma^N \in \Sigma^N$, and let $S \in 2^N$. Then $\tau^S \in \Sigma^S$ is an improvement of S upon σ^N if*

$$h^i(\tau^S, \sigma^{N \setminus S}) > h^i(\sigma^N)$$

for all $i \in S$.

Definition 2.3. *Let $\Gamma = (\Sigma^1, \ldots, \Sigma^n; h^1, \ldots, h^n)$ be an n-person game and let $\sigma^N \in \Sigma^N$. Then σ^N is a Nash equilibrium (NE) of Γ if no $i \in N$ has an improvement upon σ^N.*

Definition 2.4. *Let $\Gamma = (\Sigma^1, \ldots, \Sigma^n; h^1, \ldots, h^n)$ be an n-person game and let $\sigma^N \in \Sigma^N$. Then σ^N is a strong Nash equilibrium (SNE) of Γ if no $S \in 2^N$ has an improvement upon σ^N.*

The foregoing definitions are entirely standard, but now we proceed to consider another solution concept, namely that of coalition-proof Nash equilibrium (Bernheim et al. [5]). We start with an informal discussion:

Let $\Gamma = (\Sigma^1, \ldots, \Sigma^n; h^1, \ldots, h^n)$ be a game in strategic form and let $\sigma^N \in \Sigma^N$. An improvement τ^S upon σ^N is "self-enforcing" or "self-supporting" if no subcoalition T of S has an incentive to deviate from it. Clearly, if S is a one-player coalition then τ^S is self-supporting. However, improvements of larger coalitions may not be self-supporting as can be seen from the following example.

Example 2.5. Let $\Gamma = (\Sigma^1, \Sigma^2; h^1, h^2)$, where $\Sigma^1 = \{\sigma_1^1, \sigma_2^1\}$, $\Sigma^2 = \{\sigma_1^2, \sigma_2^2, \sigma_3^2\}$, and h^1 and h^2 are given by the following matrix:

	σ_1^2	σ_2^2	σ_3^2
σ_1^1	$(1, 1)$	$(3, 0)$	$(1, 0)$
σ_2^1	$(0, 0)$	$(2, 2)$	$(0, 0)$

First, let $\sigma^N = (\sigma_2^1, \sigma_3^2)$. Then $\tau^N = (\sigma_1^1, \sigma_1^2)$ is a self-supporting improvement of the coalition N upon σ^N. Indeed, τ^N is a NE. Now, let $\mu^N = (\sigma_2^1, \sigma_2^2)$. Then μ^N is an improvement upon τ^N. However, μ^N is not self-enforcing. Indeed, player 1 benefits by deviating.

The notion of a self-supporting improvement is made precise by the following definition.

Definition 2.6. *Let $\Gamma = (\Sigma^1, \dots, \Sigma^n; h^1, \dots, h^n)$ be a game in strategic form, let $\sigma^N \in \Sigma^N$, and let $S \in 2^N$. An internally consistent improvement (ICI) of S upon σ^N is defined by induction on $|S|$, the number of members of S, as follows: (i) If $|S| = 1$, that is $S = \{i\}$ for some $i \in N$, then $\tau^i \in \Sigma^i$ is an ICI of S upon σ^N if $h^i(\tau^i, \sigma^{N\setminus\{i\}}) > h^i(\sigma^N)$ (i.e. if τ^i is an improvement upon σ^N). (ii) If $|S| > 1$ then $\tau^S \in \Sigma^S$ is an ICI of S upon σ^N if (a) τ^S is an improvement of S upon σ^N (see Definition 2.2), and (b) if $T \subset S$ and $|T| < |S|$ then T has no ICI upon $(\tau^S, \sigma^{N\setminus S})$.*

Now we can define the equilibrium concept which is central in this paper:

Definition 2.7. *Let $\Gamma = (\Sigma^1, \dots, \Sigma^n; h^1, \dots, h^n)$ be a game in strategic form. Then $\sigma^N \in \Sigma^N$ is a coalition-proof Nash equilibrium (CNE) if no $S \in 2^N$ has an ICI upon σ^N.*

Remark 2.8. In the original definition of a CNE in Bernheim et al. [5], the concept of an ICI was not used; instead, the equilibrium was defined inductively using *self-enforcing* strategy *n*-tuples, where self-enforcing means that there are no ICI's of proper subcoalitions. The two ways of defining the equilibrium are obviously equivalent; we have opted for the present one since it is more convenient for the applications to follow.

Example 2.9. Let $\Gamma = (\Sigma^1, \Sigma^2; h^1, h^2)$, where $\Sigma^i = \{\sigma_1^i, \sigma_2^i\}$, $i = 1, 2$, and h^1 and h^2 are given by the following matrix:

	σ_1^2	σ_2^2
σ_1^1	$(1,1)$	$(0,0)$
σ_2^1	$(0,0)$	$(0,0)$

Then $\sigma^N = (\sigma_1^1, \sigma_1^2)$ is a CNE. However, the NE $\mu^N = (\sigma_2^1, \sigma_2^2)$ is not a CNE. Indeed, σ^N is an ICI of N upon μ^N. Note also that in the game of Example 2.5, $\sigma^N = (\sigma_1^1, \sigma_1^2)$ is a CNE of Γ. However, σ^N is not a SNE.

Remark 2.10. Let $\Gamma = (\Sigma^1, \dots, \Sigma^n; h^1, \dots, h^n)$ be a game in strategic form. Then every CNE of Γ is a NE of Γ, and every SNE of Γ is a CNE.

There are as far no general existence results available for the CNE's, but in the special case of games with only two players, some results may be obtained; first of all, we notice that if $\Gamma = (\Sigma^1, \Sigma^2; h^1, h^2)$ is a 2-person game, then $\sigma^N \in \Sigma^N$ is a CNE iff (i) σ^N is a NE, and (ii) there is no NE μ^N such that $h^i(\mu^N) > h^i(\sigma^N)$ for $i = 1, 2$.

Theorem 2.11. *Let $\Gamma = (\Sigma^1, \Sigma^2; h^1, h^2)$ be a 2-person game. If*

(i) *Σ^i is a compact metric space, $i = 1, 2$,*
(ii) *Σ^N is a topological space with the product topology,*

(iii) *h^i is continuous, $i = 1, 2$, and*
(iv) *Γ has a NE,*

then Γ has a CNE.

The proof of Theorem 2.11 is left to the reader.

Corollary 2.12. *Let $\Gamma = (\Sigma^1, \Sigma^2; h^1, h^2)$ be a 2-person game. If $|\Sigma^i| < \infty$, $i = 1, 2$, and Γ has a NE, then Γ has a CNE.*

The following example shows that there exist 3-person games having no CNE.

Example 2.13. Let $\Gamma = (\Sigma^1, \Sigma^2, \Sigma^3; h^1, h^2, h^3)$, where $\Sigma^i = \{\sigma_1^i, \sigma_2^i\}$, $i = 1, 2, 3$, and h^i, $i = 1, 2, 3$, are given by the following pair of matrices:

	σ_1^2	σ_2^2		σ_1^2	σ_2^2
σ_1^1	$(0,0,0)$	$(-1,-1,0)$	σ_1^1	$(1,1,-k)$	$(0,0,-k)$
σ_2^1	$(-1,-1,0)$	$(1,1,0)$	σ_2^1	$(0,0,-k)$	$(-1,-1,1)$
	σ_1^3			σ_2^3	

It can be showed that if $0 < k < 1/8$, then $\sigma^N = (\sigma_1^1, \sigma_1^2, \sigma_1^3)$ is the only NE of Γ^*, but $\sigma^N = (\sigma_1^1, \sigma_1^2, \sigma_1^3)$ is not a CNE. Indeed, (σ_2^1, σ_2^2) is an ICI of $\{1, 2\}$ upon σ^N.

We conclude this section by noticing that if improvements of coalitions in the game Γ can be supported by binding agreements, then every improvement may have a destabilizing effect. In that case only SNE's are equilibrium choices. However, we recall that binding agreements, or contracts, are possible only in cooperative games (see, e.g., Sect. 9 of Aumann [4]). In this paper we consider only non-cooperative games, so that binding agreements are not available in our framework.

3 Coalition proof Nash consistency

In this section, we continue our introductory description of the problem, introducing the notions of game forms and consistency.

Let A be a finite set of *m alternatives*, $m \geq 2$. A *linear order* on A is a complete, reflexive, transitive, and antisymmetric binary relation on A. We denote by L the set of all linear orders on A. Let $N = \{1, \ldots, n\}$, $n \geq 2$, be a set of players.

For any set D, the set of all subsets of D is denoted $P(D)$, and we write $P^2(D)$ for $P(P(D))$. The set of all nonempty subsets of D is denoted 2^D.

For $R \in L$ a linear order, we write xRy for the expression "x is better than or equals y in the order R". If $B, C \in 2^A$, we write BRC for xRy, all $x \in B$, $y \in C$. If B is a subset of A, $R|B$ denotes the linear order on B induced by R.

Let $S \in 2^N$. An S-profile is a map $R^S : S \to L$. We write R^i for $R^S(i)$, $i \in S$, so that $R^S = (R^i)_{i \in S}$. An N-profile is called a *profile* and written $R^N =$

$(R^1, \ldots, R^n)$. We identify R^N with $(R^S, R^{N \setminus S})$ for all $S \neq \emptyset$, N and profiles R^N.

Definition 3.1. *A game form is an* $(n+1)$*-tuple* $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$, *where* Σ^i *is a nonempty set for every* $i \in N$, *and* $\pi : \Sigma^N \to A$ *is a function. We assume throughout that* π *is onto.*

If $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$, then Σ^i is the set of *strategies* of player i, $i \in N$, and π is the *outcome function*.

To define a game from a game form, we need the preferences of the players. For convenience in comparison with the game theoretical concepts of Sect. 2, we introduce utility representations of the preferences: Let $R \in L$. A function $u : A \to \mathbf{R}$ is a utility representation of R if for all $x, y \in A$,

$$xRy \quad \text{iff } u(x) \geq u(y).$$

It is trivial that there exist utility representations for any $R \in L$. Moreover, if u and u' are utility representations of R, then each is a monotone transformation of the other one.

Definition 3.2. *Let* $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$ *be a game form and let* $R^N \in L^N$ *be a profile. A game* Γ *is associated with* G *and* R^N *if* $\Gamma = (\Sigma^1, \ldots, \Sigma^n; h^1, \ldots, h^n)$, *where each* h^i *is defined by*

$$h^i(\sigma^N) = u^i(\pi(\sigma^N))$$

for some utility u^i *representing* R^i.

Since we shall be concerned only with coalition proof Nash equilibria, which are obviously invariant under strictly monotone transformations of the players' payoffs, we may choose an arbitrary game associated with each profile R^N and denote it by $\Gamma(G, R^N)$.

Suppose now that $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$ is a given game form. Then each profile $R^N \in L^N$ corresponds to a game $\Gamma(G, R^N)$, and we may consider a certain solution or equilibrium for games, for example Coalition Proof Nash Equilibrium, which then assigns to $\Gamma(G, R^N)$ a set of equilibrium strategy choices, and thereby indirectly gives a set of alternatives to each profile. We shall make this idea of a correspondence between profiles and alternatives precise:

A *social choice correspondence* (SCC) is a function $H : L^N \to 2^A$. An SCC is called a *social choice function* if $H(R^N)$ is a singleton for each profile $R^N \in L^N$. Intuitively, if H is an SCC and $R^N \in L^N$, then $H(R^N)$ is the set of alternatives chosen by the group N according to the rule H.

Definition 3.3. *Let* $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$ *be a game form.* G *is Coalition Proof Nash Consistent (CNC) if for each preference profile, the set of coalition proof Nash equilibria is nonempty, i.e.* $\mathrm{CNE}(\Gamma(G, R^N)) \neq \emptyset$, *all* $R^N \in L^N$.

We note that if G is coalition proof Nash consistent, then the correspondence $R^N \mapsto \pi(\mathrm{CNE}(\Gamma(G, R^N)))$ is a social choice correspondence. Actually, we can say a little more:

Lemma 3.4. *Let $G = (\Sigma^1, \dots, \Sigma^n; \pi)$ be a game form which is coalition proof Nash consistent. Then the social choice correspondence $\pi(\mathrm{CNE}(\Gamma(G, \cdot)))$ is non-imposed in the sense that for each alternative $a \in A$, there is a profile $R^N \in L^N$ such that $\pi(\mathrm{CNE}(\Gamma(G, R^N))) = \{a\}$.*

Proof. Let $R^N \in L^N$ be a profile with $a = \max R^i$ for all $i \in N$. Choose a strategy array σ^N with $\pi(\sigma^N) = a$ (such a strategy array exists since π is onto). Then σ^N is a CNE of $\Gamma(G, R^N)$ since it is even a SNE: There is no coalition S having an improvement τ^S of σ^N because for all i, $\pi(\sigma^N) = a$ is already maximal. Also, if $b \in A \backslash \{a\}$, then $b \notin \mathrm{CNE}(\Gamma(G, R^N)))$. □

4 The representation problem with coalition proof Nash equilibria

We now proceed to introduce one of the main concepts of the paper, namely that of representation. We begin with the concept of an effectivity function:

Definition 4.1. *An effectivity function is a map $E : P(N) \to P^2(A)$ satisfying the following conditions:*

(i) *for all $B \in P(A)$, $B \notin E(\emptyset)$,*
(ii) *for all $B \in 2^A$, $B \in E(N)$,*
(iii) *for all $S \in P(N)$, $\emptyset \notin E(S)$,*
(iv) *for all $S \in 2^N$, $A \in E(S)$.*

An effectivity function assigns to each coalition S a family of sets $E(S)$, with the interpretation that if $B \in E(S)$, then S may force the outcome of society's choice to be an element of B, or equivalently, S may preclude that society chooses something from $A \backslash B$. Thus, the effectivity function may be considered as a description of a system of rights which are to be valid for the society considered, a constitution in the sense of Gärdenfors [8] (see also Peleg [15]).

A game form gives rise to (at least) two different types of effectivity functions:

Definition 4.2. *Let $G = (\Sigma^1, \dots, \Sigma^n; \pi)$ be a game form, $S \in 2^N$ a coalition, and B a nonempty subset of the set A of alternatives in G. We say that S is α-effective for B if there is a strategy choice $\sigma^S = (\sigma^i)_{i \in S}$ such that for all strategy choices $\sigma^{N \backslash S} = (\sigma^i)_{i \in N \backslash S}$, $\pi(\sigma^S, \sigma^{N \backslash S}) \in B$, and β-effective for B if for each strategy choice $\sigma^{N \backslash S}$, there is σ^S such that $\pi(\sigma^S, \sigma^{N \backslash S}) \in B$.*

The α-effectivity function associated with G is the effectivity function $E^G_\alpha : P(N) \to P^2(A)$ given by

$$E^G_\alpha(S) = \{B \in 2^A | S \text{ is } \alpha\text{-effective for } B\},$$

for $S \in 2^N$, and $E^G_\alpha(\emptyset) = \emptyset$.

The β-effectivity function associated with G is the effectivity function E^G_β which to each coalition S assigns the family of subsets $B \in 2^A$ such that S is β-effective for B.

We collect some obvious consequences of these definitions in a lemma:

Lemma 4.3. *The α- and β-effectivity functions of a game form G satisfy:*

(1) *for all* $S \in 2^N$, $E_\alpha^G(S) \subset E_\beta^G(S)$,
(2) $E_\alpha^G(N) = E_\beta^G(N)$,
(3) *for* $B \in 2^A$, $S \in 2^N$, *if* $B \in E_\alpha^G(S)$, *then* $A \backslash B \notin E_\beta^G(N \backslash S)$.
(4) *If* $B \notin E_\alpha^G(S)$, *then* $A \backslash B \in E_\beta^G(N \backslash S)$.

In general, the inclusion in Lemma 4.3.(1) may be proper (see e.g., Peleg [13]). We shall be particularly interested in the situation where equality obtains, and, as it is known from previous work on implementation in strong Nash equilibria (see Moulin and Peleg [12]), such equality does obtain in many interesting situations. Indeed, one of the main results of the present paper is that the equality of α- and β-effectivity functions of a game form G obtains when G is CNC.

Definition 4.4. *Let* $E : P(N) \to P^2(A)$ *be an effectivity function. A representation of E (with respect to coalition proof Nash equilibrium) is a game form* $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$ *such that*

(1) $E = E_\alpha^G$,
(2) *G is coalition proof Nash consistent (CNC).*

The representation problem for a given effectivity function consists in finding a representation. In the interpretation, the existence of a representation means that the system of rights in society described by the effectivity function is *consistent* in the sense that there exists a set of rules (formally, a game form) such that the individuals and groups in society can exercise their rights simultaneously even when acting strategically within the framework provided by their rules.

As noted in the previous section, a CNC game form gives rise to a social choice correspondence $R^N \mapsto \pi(\mathrm{CNE}(G, R^N))$, so that a representation of an effectivity function (with respect to coalition proof Nash equilibrium) induces a particular social choice correspondence.

Definition 4.5. *Let* $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$ *be a game form and let* $H : L^N \to 2^A$ *be a social choice correspondence. Then G CNE-implements H if for every profile* $R^N \in L^N$, $H(R^N) = \pi(\mathrm{CNE}(\Gamma(G, R^N)))$. *H is CNE-implementable if there exists a game form* $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$ *which CNE-implements H.*

Let G be CNC. The social choice correspondence $R^N \mapsto \pi(\mathrm{CNE}(G, R^N))$ clearly is CNE-implemented by G. In the case where G is a representation of an effectivity function E, it might be conjectured that there is some connection between properties of E and properties of the SCC which is implemented by G, and this is indeed the case.

Before we proceed, it will be useful to introduce another way of constructing effectivity functions, namely from social choice correspondences:

Definition 4.6. *Let* $H : L^N \to 2^A$ *be a non-imposed social choice correspondence, let* $S \in 2^N$, *and let* $B \in 2^A$. *We say that S is α-effective for B if there* *exists an S-profile* R^S *such that for all* $(N\backslash S)$*-profiles* $Q^{N\backslash S}$, $H(R^S, Q^{N\backslash S}) \subset B$ *and S is* β*-effective for B if for every* $(N\backslash S)$*-profile* $Q^{N\backslash S}$*, there exists an S-profile* R^S *such that* $H(R^S, Q^{N\backslash S}) \subset B$.

The α*-effectivity function associated with H is the effectivity function* $E_\alpha^H : P(N) \to P^2(A)$ *given by*

$$E_\alpha^H(S) = \{B \in 2^A | S \text{ is } \alpha\text{-effective for } B\},$$

for $S \in 2^N$*, and* $E_\alpha^H(\emptyset) = \emptyset$*, and the* β*-effectivity function associated with H is the effectivity function* E_β^H *which to each coalition S assigns the family of subsets* $B \in 2^A$ *such that S is* β*-effective for B.*

Remark 4.7. It is clear that Lemma 4.3.(1)–(3) hold also for E_α^H, E_β^H. That also 4.3.(4) holds if H is CNE-implementable will emerge as a consequence of the results below.

Now we may combine all the previously introduced notions in our first fundamental result, which is due to Peleg [14]. A proof of the theorem can be found in Abdou and Keiding [2] (Theorem 7.2.2, pp. 141–143).

Theorem 4.8. *Let* $E : P(N) \to P^2(A)$ *be an effectivity function, let* $G = (\Sigma^1, \ldots, \Sigma^n; \pi)$ *be a representation of E, and let* $H : L^N \to 2^A$ *be a social choice correspondence. If*

$$H(R^N) \subset \pi(\mathrm{CNE}(G, R^N))$$

for all $R^N \in L^N$*, then*

$$E = E_\alpha^G = E_\beta^G = E_\alpha^H = E_\beta^H.$$

Remark 4.9. In Abdou and Keiding [2], an SCC H which satisfies the assumptions of the theorem, namely that there is a game form G such that $H(R^N) \subset \pi(\mathrm{CNE}(G, R^N))$ for all profiles $R^N \in L^N$, is said to be *partially* CNE-implemented. However, this is not a standard terminology, cf., e.g., Dasgupta et al. [7].

We conclude this section by noting that as a consequence of Theorem 4.8, an effectivity function which has a representation (with respect to coalition proof Nash equilibrium), must have certain properties.

Definition 4.10. *Let* $E : P(N) \to P^2(A)$ *be an effectivity function.*

(a) *E is superadditive if for all* $B, C \in 2^A$, $S, T \in 2^N$, $B \in E(S)$, $C \in E(T)$*, if* $S \cap T = \emptyset$*, then* $B \cap C \in E(S \cup T)$,

(b) *E is monotonic if for all* $B, C \in 2^A$, $S, T \in 2^N$, $B \in E(S)$*, if* $B \subset C$ *and* $S \subset T$*, then* $C \in E(T)$.

(c) *E is maximal if for all* $B \in 2^A$, $S \in 2^N$*, if* $B \notin E(S)$ *then* $A\backslash B \in E(N\backslash S)$.

The three properties (a)–(c) above are not independent. Indeed, it is rather easy to show that if E is maximal and superadditive, then it is monotonic: If $B \in E(S)$, $B \subset C$, $S \subset T$, and $C \notin E(T)$, then $A\backslash C \in E(N\backslash T)$ by maximality. Now, S and $N\backslash T$ are disjoint coalitions, and by superadditivity, $B \cap (A\backslash C) \in E(S \cup (N\backslash T))$. But $B \cap (A\backslash C) = \emptyset$, and we have a contradiction.

The following is an immediate consequence of Lemma 4.3 and Theorem 4.8, and as such it is essentially a restatement of results in Peleg [14] and Abdou and Keiding [2]:

Theorem 4.11. *Let $E : P(N) \to P^2(A)$ be an effectivity function which has a representation (with respect to coalition proof Nash equilibrium) $G = (\Sigma^1, \dots, \Sigma^n; \pi)$. Then E is superadditive, monotonic, and maximal.*

Proof. By Theorem 4.8, $E = E_\alpha^G = E_\beta^G$. It is easily seen from the definition that E_α^G is superadditive: Indeed, suppose that $S, T \in 2^N$ are disjoint coalitions, that S is α-effective for B and that T is α-effective for C. Then there is a S-strategy σ^S such that for all $(N \backslash S)$-profiles $\tau^{N \backslash S}$, we have $\pi(\sigma^S, \tau^{N \backslash S}) \subset B$. Furthermore, there is a T-strategy μ^T such that for all $(N \backslash T)$-strategies $\nu^{N \backslash T}$, in particular for the strategy $(\sigma^S, \tau^{N \backslash (T \cup S)})$, we have $\pi(\mu^T, \nu^{N \backslash T}) \subset C$. Obviously, the coalition $S \cup T$ is α-effective for $B \cap C$.

To show that E is maximal, we need only combine Lemma 4.3.(4) and Theorem 4.8. Finally, monotonicity follows from maximality and superadditivity. □

Remark 4.12. Let $G = (\Sigma^1, \dots, \sigma^n, \pi)$ be a game form. If G is CNC, then by Theorem 4.8, $E_\alpha^G = E_\beta^G$. If $n = 2$, then G is CNC iff it is Nash consistent. Hence, for $n = 2$, G is CNC iff $E_\alpha^G = E_\beta^G$ (see, e.g., Abdou [1] for a characterization in terms of effectivity functions of two-person game forms which are Nash consistent).

Remark 4.13. Theorem 4.8 supplies a simple necessary condition for CNE-imple-mentability: If an SCC is CNE-implementable, then $E_\alpha^H = E_\beta^H$ is maximal (see Definition 4.10). This condition is easy to check: For example it shows that the Pareto correspondence is not CNE-implementable. Boylan [6] proves (essentially for a generalized lobbying model) that on a restricted domain (he assumes continuity and monotonicity of the utility function of a player in her own transfer), weak (Maskin) monotonicity is a necessary and sufficient condition for CNE-implementability. As Maskin monotonicity does not imply the maximality of the α-effectivity function of a social choice correspondence (see, again, the Pareto correspondence), our results, for unrestricted (finite) domains, are independent of those of Boylan [6].

As we see, if an effectivity function has a representation, then it must be superadditive and maximal. The logical next question is: Do such effectivity functions have further properties?

The answer is no. Given an arbitrary effectivity function E which is superadditive and maximal, it is possible to find a game form which is CNC and represents E. This result is the subject of our discussion in the following sections.

5 Uniform domination and *u*-effectiveness

Before we proceed to state and prove a converse of Theorem 4.11, we insert in this section a short discussion of some solution concepts for games defined in effectivity function form. These solution concepts may have independent applications; the reason for our treatment of them is that they are used later in the definition of the game form which will be used in our solution of the representation problem. For a more detailed discussion of these concepts, the reader is referred to Abdou and Keiding [2].

We start with the notion of uniform domination:

Definition 5.1. *Let $E : P(N) \to P^2(A)$ be an effectivity function and $R^N \in L^N$ a profile. For $S \in 2^N$ a coalition and B a subset of A we say that the alternative x is uniformly dominated (shorthand: u-dominated) by B via S at R^N if $B \in E(S)$, $x \notin B$, and $BR^S A \backslash B$. The alternative x is u-dominated via S at R^N if there is $B \in 2^A$ such that x is u-dominated by B via S at R^N, and x is u-dominated at R^N if there is $S \in 2^N$ such that x is u-dominated via S at R^N.*

Thus for an alternative to be u-dominated via S, we demand that there is some set B of alternatives for which S is effective and which moreover is such that all players in S agree that everything in B is better than everything not in B.

Example 5.2. Let $E : P(\{1,2,3,4\}) \to P^2(\{x,y,z\})$ be defined by the rule

$$B \in E(S) \Leftrightarrow |B| + |S| \geq 4$$

(E is a so-called additive effectivity function, cf. Moulin and Peleg [12]). It is easily seen that E is superadditive and maximal.

Consider the profile

1	2	3	4
x	y	z	y
y	z	x	z
z	x	y	x

Here, x is u-dominated by $\{y,z\}$ via $\{2,4\}$. The alternative z is dominated (in the usual sense of this word) by $\{y\}$ via the coalition $\{1,2,4\}$, but this is not a case of u-domination.

It is clear from the definition that it is rather hard to u-dominate. The following notion is introduced in order to capture the idea that a coalition might not u-dominate an alternative but on the other hand might make it look as if it did.

Definition 5.3. *Let $E : P(N) \to P^2(A)$ be an effectivity function and $R^N \in L^N$ a profile. For $S \in 2^N$ a coalition and $B \in 2^A$ a subset of alternatives we say that S is u-effective for B at R^N if there exists an S-profile T^S such that for each alternative $x \in A \backslash B$, there is $S' \in 2^S$ satisfying the conditions*

(a) *x is u-dominated via S' at $(T^S, R^{N\backslash S})$,*
(b) *$BR^{S'}x$.*

Example 5.4. In the profile of Example 5.2, the coalition $\{1,2,4\}$ is u-effective for $\{y\}$. Indeed, consider the $\{1,2,4\}$-profile $T^{\{1,2,4\}}$

1	2	4
y	y	y
x	z	z
z	x	x

As before x is u-dominated in $(T^{\{1,2,4\}}, R^3)$ by $\{y,z\}$ via $\{2,4\}$, and $yR^{\{2,4\}}x$. Furthermore, z is u-dominated by $\{y\}$ via $\{1,2,4\}$, and $yR^{\{1,2,4\}}z$. We conclude that $\{1,2,4\}$ is u-effective for $\{y\}$.

In the example, the coalition which was u-effective for the subset $\{y\}$ was also effective in the usual sense of the word, that is $\{y\} \in E(\{1,2,4\})$. This is a general fact, as shown from the following lemma (for a proof, see Abdou and Keiding [2], p. 148):

Lemma 5.5. *Let E be a monotonic and superadditive effectivity function, and let $R^N \in L^N$. If $S \in 2^N$ is u-effective for $B \in 2^A$ at R^N, then $B \in E(S)$.*

Having introduced u-effectiveness, the following concept to be introduced is that of *indirect u-domination:*

Definition 5.6. *Let $E : P(N) \to P^2(A)$ be an effectivity function and $R^N \in L^N$ a profile. An alternative $x \in A$ is indirectly u-dominated at R^N by the subset B of A via the coalition $S \in 2^N$ if S is u-effective for B, $x \notin B$, and BR^Sx; x is indirectly u-dominated at R^N if there are $B \in 2^A$, $S \in 2^N$, such that x is indirectly u-dominated by B via S at R^N.*

Example 5.7. At the profile R^N of Example 5.2, the alternative z was dominated but not u-dominated. However, as we saw in Example 5.4, the coalition $\{1,2,4\}$ is u-effective for $\{y\}$ at R^N, so that z is indirectly dominated at R^N. Clearly, the alternative x, which is u-dominated at R^N, is a fortiori indirectly u-dominated at R^N.

The following result linking together the various concepts introduced above will be used in the next section:

Theorem 5.8. *Let $E : P(N) \to P^2(A)$ be an effectivity function which is monotonic and superadditive, and let $R^N \in L^N$ be a profile. If $B \in 2^A$ is a minimal (for inclusion) set such that some coalition S is u-effective for B at R^N, then the elements of B are not indirectly u-dominated at R^N.*

For a proof, the reader is referred to Abdou and Keiding [2], pp. 149–150.

Example 5.9. In Example 5.7, it was shown that $\{1,2,4\}$ is u-effective for $\{y\}$, so the alternative y belongs to a minimal (for inclusion) set of alternatives for which some coalition is u-effective at $R^{\{1,2,3,4\}}$. Therefore, by Theorem 5.8, y is not indirectly dominated at this profile.

To show that there may be alternatives which are not indirectly dominated even at profiles where each alternative is dominated, consider the effectivity function $E : P(\{1,2,3,4\}) \to P^2(\{x,y,z,w\})$ defined by

$$E(S) = \{B \mid |B| \geq 1\} \quad \text{if } |S| \geq 3 \text{ or if } |S| = 2 \text{ and } 1 \in S,$$

$$E(S) = \begin{cases} \emptyset & \text{if } S = \emptyset, \\ \{\{x,y,z,w\}\} & \text{otherwise.} \end{cases}$$

Again E is superadditive and maximal. There is a profile R^N such that each alternative is dominated at R^N, namely

1	2	3	4
x	y	z	w
y	z	w	x
z	w	x	y
w	x	y	z

We claim that x is not indirectly u-dominated at R^N. Indeed, x cannot be indirectly u-dominated via a coalition containing individual 1, so it is enough to show that $\{2,3,4\}$ is not u-effective for any subset of $\{y,z,w\}$. But this follows from the fact that $\{2,3,4\}$ is the only subset of N not containing individual 1 which is effective for any proper subset of $\{x,y,z,w\}$, and only $wR^{\{2,3,4\}}x$; this means $\{2,3,4\}$ could possibly be u-effective for $\{w\}$ but not for other subsets of $\{y,z,w\}$, and since $\{2,3,4\}$ is not u-effective for $\{w\}$, we have shown that $\{2,3,4\}$ cannot indirectly dominate x.

Let $E : P(N) \to P^2(A)$ be a monotonic and superadditive effectivity function. For later use we define a particular social choice function $F : L^N \to A$ as follows: For any profile R^N, there is a set $B \in 2^A$ such that some coalition $S \in 2^N$ is u-effective for B at R^N (N is u-effective for A at every profile). Let $\mathcal{M}(R^N, E)$ be the set of alternatives belonging to a subset B of A which is minimal with the property that there is a coalition $S \in 2^N$ which is u-effective for B; then $\mathcal{M}(R^N, E) \neq \emptyset$. Put

$$F(R^N) = \max(R^1 | \mathcal{M}(R^N, E)),$$

where for any $R \in L$ and $B \in 2^A$, $\max(R|B)$ is the unique alternative $x \in B$ with xRy for all $y \in B$.

The social choice function F will play a crucial role in Sect. 7 when we consider a coalition proof Nash equilibria in the game derived from the game form to be introduced in the next section. The choice of individual 1 as the particular one whose preferences are decisive for the choice of alternative from the set $\mathcal{M}(R^N, E)$ is arbitrary, but once the choice is made, it will matter for the construction of the game form.

6 A representing game form

Throughout this section, $E : P(N) \to P^2(A)$ is a given monotonic and superadditive effectivity function. In the following two sections, we show that such an effectivity function has a representation: We construct a particular game form $G = (\Sigma^1, \dots, \Sigma^n; \pi)$ and show that (1) E is both the α- and β-effectivity function of G, and (2) for any given profile, we can define suitable strategies such that this strategy array is a coalition proof Nash equilibrium. In the present section we introduce the game form and address the problem (1) obtaining a first and partial answer to be completed subsequently.

In the definition of the game form G, we start with the strategy set Σ^1, of player 1, which is defined as

$$\Sigma^1 = L \times \Phi \times A,$$

where Φ is the set of all maps

$$\varphi : 2^{N \setminus \{1\}} \times 2^A \to 2^A$$

taking pairs (S, B) consisting of a coalition S not containing player 1 and a nonempty subset B of A to a subset $\varphi(S, B)$ of A with $\varphi(S, B) \subset B$. Intuitively, the selection function will be used to pick an alternative which is not uniformly dominated via S. Thus, a strategy of player 1 consists of a preference relation, a selection function of the type described above, and an alternative.

For players $i \neq 1$, we define the strategy set Σ^i by

$$\Sigma^i = L \times \Phi \times A \times \mathscr{P}^i \times (\mathbf{N}_0 \times L),$$

where $\mathbf{N}_0$ is the set of nonnegative integers, and where $\mathscr{P}^i$ is the set of maps $b^i : \bigcup_{S \in 2^{N \setminus \{i\}}} L^S \to P(A \times 2^N)$ taking any preference profile Q^S for $S \in 2^N$, $i \notin S$, to a family $b^i(Q^S) = (x, S^x)_{x \in \hat{B}_i}$, where $\hat{B}_i$ is a subset of A and for each $x \in \hat{B}_i$, S^x is a coalition containing i. As it is seen below, b^i contain a list of alternatives x which individual i wants to exempt from u-domination, provided that each individual in S^x has stated the same pair (x, S^x).

To define the outcome function π, we need some notation: For $\sigma^N = (\sigma^1, \dots, \sigma^n)$ a strategy n-tuple in

$$\Sigma^1 \times \cdots \times \Sigma^n,$$

with $\sigma^1 = (Q^1, \varphi^1, x^1)$, $\sigma^i = (Q^i, \varphi^i, x^i, b^i, (t^i, P^i))$, $i \neq 1$, define

$$S[\sigma^N] = \{i \in N \mid \varphi^i \neq \varphi^1 \text{ or } x^i \neq x^1\}$$

(thus, $S[\sigma^N]$ never contains player 1), and let the $S[\sigma^N]$-profile $R[\sigma^N]$ be given by

$$R[\sigma^N] = (Q^i)_{i \in S[\sigma^N]}.$$

Finally, the sets B^i are determined by

$$B^i = \{x \in A \mid \exists S^x \in 2^N, i \in S^x : (x, S^x) \in b^h((Q^j)_{j \in S[\sigma^N] \setminus \{i\}}), \text{ all } h \in S^x\}$$

if $S[\sigma^N] \neq \{i\}$ and $B^i = \emptyset$ if $S[\sigma^N] = \{i\}$.

Now there are two cases:

Case 1: $S[\sigma^N] = \emptyset$. We let

$$\pi(\sigma^N) = x^1,$$

i.e. the outcome is the alternative stated by player 1, which by the definition of $S[\sigma^N]$ is the alternative stated by any player i.

Case 2: $S[\sigma^N] \neq \emptyset$. Define the set $B[\sigma^N]$ as the set of alternatives x for which there is no subcoalition S' of $S[\sigma^N]$ such that x is u-dominated at $R[\sigma^N]$ via S' and $x \notin B^i$ for all $i \in S'$, *i.e.*

$$B[\sigma^N] = A \backslash \Big\{ y \in A \mid \exists S' \in 2^{S[\sigma^N]}, B \in E(S') : \\ BQ^{S'}A\backslash B,\ y \in A\backslash B,\ y \in A \backslash \bigcup_{i \in S'} B^i \Big\};$$

($B[\sigma^N]$ is non-empty since it contains the set of alternatives which are not u-dominated at the profile $R[\sigma^N]$, and the latter set is non-empty by Theorem 5.8). Now let the outcome be given by

$$\pi(\sigma^N) = \max(P^{i^0} | \varphi^1(S[\sigma^N], B[\sigma^N])),$$

where i^0 is the smallest of the integers $i \in S[\sigma^N]$ such that $t^i \geq t^j$, all $j \in S[\sigma^N]$. Thus, the component t^i in the strategy σ^i determines whose preferences (as stated in the strategy) are to be decisive for the final choice. The role of the sets B^i is a little more obscure; they are there to make it possible to exempt certain alternatives from domination.

Essentially, the outcome rule is as follows: At first, the set $S[\sigma^N]$ of players who disagree with player 1 (as shown by the choice of selection function and alternative) is determined. If nobody disagrees, the unanimously stated alternative is chosen. Otherwise, we look at the set of alternatives which are not uniformly dominated in the disagreeing coalition (when it is taken into consideration that the players may exempt alternatives from domination by including them in B^i). From this set, a subset is chosen according to the selection function of player 1. Now the final choice from this subset is made by the member of $S[\sigma^N]$ who has stated the largest t^i, and according to his linear order P^i.

We remark that the notion of *u-domination*, originally introduced in Abdou and Keiding [2] and reviewed in Sect. 5 above, enters into the very definition of the game form.

A first result about G connects it to the given effectivity function E:

Lemma 6.1. *For all coalitions $S \in 2^N$, $E(S) \subset E^G_\beta(S)$. Moreover, if E is maximal, then $E^G_\alpha(S) \subset E(S) \subset E^G_\beta(S)$ for all S.*

Proof. We show that $B \in E(S)$ minimal for inclusion implies $B \in E^G_\beta(S)$. If $1 \notin S$, choose an S-profile Q^S such that $BQ^SA\backslash B$. Then for each $\tau^{N\backslash S}$ with $\tau^1 = (Q^1, \varphi^1, x^1)$, the S-strategy σ^S with

$$\sigma^i = (Q^i, \varphi, x^i, \emptyset, (0, Q^i)),$$

where φ is an arbitrary selection function with $\varphi \neq \varphi^1$, gives $S \subset S[(\sigma^S, \tau^{N\setminus S})]$ and $B[(\sigma^S, \tau^{N\setminus S})] \subset B$, so that $\pi(\sigma^S, \tau^{N\setminus S}) \in B$.

If $1 \in S$, let φ be a selection function such that $\varphi(S', B') \subset B$ whenever $B \cap B' \neq \emptyset$; choose Q^1 such that $BQ^1A\setminus B$, let $x \in B$ and define σ^S by $\sigma^1 = (Q^1, \varphi, x)$,

$$\sigma^i = (Q^1, \varphi, x, \emptyset, (0, Q^1))$$

for $i \in S$, $i \neq 1$. For any $\tau^{N\setminus S}$, if $B[(\sigma^S, \tau^{N\setminus S})] \subset A\setminus B$, then each alternative in B must be u-dominated at the profile $R[(\sigma^S, \tau^{N\setminus S})]$. By the definition of Q^1, such u-domination must be via subcoalitions of $N\setminus S$. It follows by the proof of Lemma 5.5 that we must have $A\setminus B \in E(N\setminus S)$. However, this contradicts superadditivity of E since $B \in E(S)$. We conclude that $B \in E^G_\beta(S)$.

For the second statement of the lemma, let $B \in E^G_\alpha(S)$. If $B \notin E(S)$, then by maximality of E, $A\setminus B \in E(N\setminus S)$, and by the first part of the proof, $A\setminus B \in E^G_\beta(N\setminus S)$. However, from the definition of E^G_α and E^G_β we have that $B \in E^G_\alpha(S)$ implies $A\setminus B \notin E^G_\beta(N\setminus S)$, and we conclude that $B \in E(S)$. □

Remark 6.2. In the case that the game form G is tight in the sense that $E^G_\alpha = E^G_\beta$, we get from Lemma 6.1 that $E^G_\alpha = E^G_\beta = E$. Tightness of G in its turn will be a consequence of the results in the next section, since here it will be shown that G is coalition proof Nash consistent, so that in particular, G is a representation of E^G_α, whence by Theorem 4.8, $E^G_\alpha = E^G_\beta$.

7 Proof of coalition proof Nash consistency

In the present section, we show that the game form defined in Sect. 6 is CNC. We start by defining suitable strategies and then proceed to check that they indeed are equilibrium strategies.

Let R^N be an arbitrary profile; we let $x = F(R^N)$, where F is the particular social choice function defined at the end of Sect. 5.

For player 1, the strategy σ^1 is defined as (R^1, φ, x); the selection function $\varphi \in \Phi$ is defined as follows:

Let $\mathscr{D}^1$ be the set of $(S, B) \in 2^{N\setminus\{1\}} \times 2^A$ such that there is some $y \in B$, $y \neq x$, with xR^iy for some $i \in S$ (some alternative in B is worse than x for some members of S). Choose any such y and let $\varphi(S, B) = \{y\}$.

Let $\mathscr{D}^2$ be the set of (S, B) with BR^Sx for which there is a two-element subset B' of B such that for some $i, j \in S$,

$$\max(R^i|B') \neq \max(R^j|B')$$

(so that the members of S disagree on the best element of B'); such a subset B' is called an admissible pair. An admissible pair $\{z_1, z_2\}$ dominates another admissible pair $\{w_1, w_2\}$ if $\{w_1, w_2\}R^Sz_1$ or $\{w_1, w_2\}R^Sz_2$ (so a pair is dominated by another if both alternatives of the first are considered better than one of the alternatives in the other pair by all members of the coalition under consideration); clearly, if there are admissible pairs, then there are also undominated admissible pairs. Define the partial relation $\succ$ on the undominated admissible pairs by

$$\{w_1, w_2\} \succ \{z_1, z_2\} \quad \text{if } [w_1 = z_1, w_2 R^S z_2],$$

possibly after a renumbering of $\{z_1, z_2\}$ (so that one of the alternatives is the same in the two sets and the other is dominated via S). Then $\succ$ is acyclic and therefore admits minimal elements; Let $\varphi(S, B)$ select an admissible pair which is minimal for $\succ$.

Finally, let $\mathscr{D}^3 = (2^{N\setminus\{1\}} \times 2^A)\setminus(\mathscr{D}^1 \cup \mathscr{D}^2)$; for $(S, B) \in \mathscr{D}^3$, there exists a unique $w \in B$ such that $BR^S w$; put $\varphi(S, B) = \{w\}$.

This concludes the description of the function φ, (which of course depends on the sincere profile R^N). The very elaborate definition of φ, in particular for its restriction to $\mathscr{D}^2$, will be used in the proof of Lemma 7.2 below, where it helps us to establish a recursiveness property of improvements which is crucial for our reasoning.

If $i \neq 1$, then $\sigma^i = (Q^i, \varphi, x, b_\varnothing, (0, R^i))$, where φ and x were defined above, and $b_\varnothing$ is the constant function with value $\varnothing$. To define the preference relation Q^i, we use that according to the definition of $x \in F(R^N)$, there is a pair $(\bar{B}, \bar{S}) \in 2^A \times 2^N$ with $x \in \bar{B}$ such that $\bar{B}$ is minimal for inclusion among sets B belonging to pairs (B, S) with S u-effective for B at the profile R^N. By Definition 5.3, there is an $\bar{S}$-profile $T^{\bar{S}} = (T^i)_{i \in \bar{S}}$ such that each $y \notin \bar{B}$ is u-dominated at $(T^{\bar{S}}, R^{N\setminus\bar{S}})$ and $\bar{B}R^{S'}y$ for some non-empty $S' \subset \bar{S}$. Now define Q^i as T^i if $i \in \bar{S}$ and as R^i otherwise.

This completes the definition of the particular strategy array σ^N. In the remaining part of the section, we show that the strategies defined above are indeed coalition proof Nash equilibria, and that they result in the outcome prescribed by the social choice function F.

We start with the last assertion which is an easy consequence of the definitions:

Lemma 7.1. $\pi(\sigma^N) = x = F(R^N)$.

Proof. We have $S[\sigma^N] = \varnothing$ so that Case 1 in the definition of π applies. □

Now we want to establish that σ^N is a CNE of the game $\Gamma(G, R^N)$. This means that no coalition can have an ICI against σ^N (cf. Definition 2.6). We start with coalitions which do not contain the first player:

Lemma 7.2. *Let $S \in 2^N$ be a coalition with $1 \notin S$. Then S has no internally consistent improvement of σ^N.*

The proof of Lemma 7.2 is rather long, and to facilitate reading we provide an overview of the proof: In order to prove that coalitions containing individual 1 have no ICI we show that if S has an improvement τ^S, then it cannot be an ICI, since some subcoalition has an ICI of $(\tau^S, \sigma^{N\setminus S})$. To show this, we exhibit an improvement with the property that if some subcoalition has an ICI, then this would also be an ICI of $(\tau^S, \sigma^{N\setminus S})$, from which we can then conclude that S has no ICI of σ^N. We indicate the distinct parts (i)–(v) of this reasoning as we proceed in the proof.

Proof. If the coalition S has no improvement of σ^N, we are done. So, let τ^S be an improvement of σ^N, $\tau^i = (\tilde{Q}^i, \tilde{\varphi}^i, \tilde{x}^i, \tilde{b}^i, (\tilde{t}^i, \tilde{P}^i))$, $i \in S$, with $\pi(\tau^S, \sigma^{N \setminus S}) = q$, where $qR^S x$ and $q \neq x$. We must show that τ^S is not internally consistent.

(i) First of all, we choose a particular subcoalition S_y of S having an improvement of $(\tau^S, \sigma^{N \setminus S})$:

If $S[(\tau^S, \sigma^{N \setminus S})] = \emptyset$, then $q = x$ by the definition of the outcome function π, a contradiction. We consider first the case where $S[(\tau^S, \sigma^{N \setminus S})] = S$; the general case is treated afterwards.

Let $R[(\tau^S, \sigma^{N \setminus S})] = T^S$, $B[(\tau^S, \sigma^{N \setminus S})] = B$. If $(S, B) \in \mathscr{D}^1$, then φ would choose an alternative y such that $xR^i y$, contradicting that the outcome is preferred to x by all members of S. Suppose then that $(S, B) \in \mathscr{D}^2$; in this case $\varphi(S, B) = B'$, where B' is such that

$$\max(R^i | B') = y \neq z = \max(R^j | B')$$

for some $i, j \in S$. Suppose w.l.o.g. that $q \neq y$. Then $i \in S$ can improve upon the strategy array $(\tau^S, \sigma^{N \setminus S})$ simply by changing the last component of the strategy τ^i to some pair $(\hat{t}^i, \hat{P}^i)$, where $\hat{t}^i > \max\{\tilde{t}^j \mid j \in S\}$ and $\max \hat{P}^i = y$. This improvement by the coalition $\{i\}$ is internally consistent, so τ^S cannot be an ICI.

It remains to consider the case where $(S, B) \in \mathscr{D}^3$, so that $R^i | B' = R^j | B'$ for all $i, j \in S$ and $B' \subset B$, meaning that $R^i | B = R^j | B$ for all i, j. By the definition of φ, we get that $\pi(\tau^S, \sigma^{N \setminus S}) = q$ satisfies $BR^S q$ (that is, q is the worst alternative in B for all $i \in S$), and $B \subset C$, where $C = \{w \in A \mid wR^S q\}$.

Now $x = F(R^N) = \pi(\sigma^N)$ is not indirectly u-dominated at R^N (Theorem 5.8); using Definition 5.6 we conclude that S is not u-effective for C at R^N; using Definition 5.3 we get the existence of some alternative $y \notin C$ with the property that for each subcoalition S' of S, y is not u-dominated at $(T^S, R^{N \setminus S})$ via S', or $CR^{S'} y$ does not hold. If $qR^S y$, then y is not u-dominated in $(T^S, R^{N \setminus S})$, meaning that $y \in B \subset C$, a contradiction; therefore, $\{i \in S \mid yR^i q\} \neq \emptyset$. Since $CR^{S'} y$ holds for any subcoalition of $S \setminus \{i \in S \mid yR^i q\}$ we get that y is not u-dominated at $(T^S, R^{N \setminus S})$ via any such subcoalition. Thus there is $y \notin C$, i.e. with $qR^i y$ for some $i \in S$, and $S_y \subset S$, $S_y \neq \emptyset$, such that $yR^{S_y} q$ and y is not u-dominated at $(T^S, R^{N \setminus S})$ via any subset of $S \setminus S_y$. We now choose a pair (y, S_y) with the above properties in such a way that S_y is minimal for set inclusion.

(ii) Next, we define suitable strategies v^i for the members of S_y so that v^{S_y} is an improvement of $(\tau^S, \sigma^{N \setminus S})$:

Define strategies $v^i = (\tilde{Q}^i, \tilde{\varphi}^i, \tilde{x}^i, \hat{b}^i, (\hat{t}^i, \hat{P}^i))$ for $i \in S_y$ with $\hat{b}^i$ such that

$$\hat{b}^i(T^{S \setminus \{i\}}) = \tilde{b}^i(T^{S \setminus \{i\}}) \cup \{(y, S_y)\}$$

and $\hat{b}^i$ agrees with $\tilde{b}^i$ on all other profiles, $\hat{t}^i = \max\{\tilde{t}^i \mid i \in S\} + 1$, and $\hat{P}^i \in L$ some linear order with $\max \hat{P}^i = y$, leaving everything else unchanged. Then in the new strategy array $(v^{S_y}, \tau^{S \setminus S_y}, \sigma^{N \setminus S})$, we have

$$B[(v^{S_y}, \tau^{S \setminus S_y}, \sigma^{N \setminus S})] = B \cup \{y\}.$$

By construction, φ assigns to the pair $(S, B \cup \{y\})$ either the set $\{y\}$ (in the case where xR^iy for some $i \in S$, so that $(S, B \cup \{y\}) \in \mathscr{D}^1$, since in that case y is the only alternative which is not preferred to x by all individuals in S), or a subset of $B \cup \{y\}$ containing y (since the preferences R^i for $i \in S$ agree on B but do not agree on $B \cup \{y\}$, so that $(S, B \cup \{y\}) \in \mathscr{D}^2$, and all admissible pairs from $B \cup \{y\}$ must contain y). Consequently

$$\pi(\nu^{S_y}, \tau^{S\setminus S_y}, \sigma^{N\setminus S}) = y$$

and ν^{S_y} is indeed an improvement for S_y of $(\tau^S, \sigma^{N\setminus S})$.

If no subcoalition has an ICI of $(\nu^{S_y}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})$, then the original improvement was not an ICI, and we would be through.

(iii) We now claim that if a coalition has an ICI of $(\nu^{S_y}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})$, then it would also have an ICI of $(\tau^S, \sigma^{N\setminus S})$. Once this claim has been proved, we have shown that S has no ICI of σ^N.

Suppose therefore that some proper subcoalition S'' of S_y has an ICI $\mu^{S''}$ of

$$(\nu^{S_y}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})$$

with

$$\pi(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S}) = z, \quad zR^{S''}y.$$

Then $S[(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})]$ is a subset of S containing $S\setminus S''$ and some (possibly empty) subset S''' of S''; let

$$R[(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})] = (\tilde{T}^{S'''}, T^{S\setminus S''}),$$

where $\tilde{T}^{S'''}$ is some S'''-profile.

(iv) We now show that $\mu^{S''}$ is an ICI of σ^N as well, at least in the case where $\mu^{S''}$ is sufficiently well-behaved:

Suppose that in some of the strategies μ^i, for $i \in S'''$, the first component has been changed to some $\hat{Q}^i \neq \tilde{Q}^i$. Then $R[(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})] \neq T^S$, and

$$B[(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})] = B[(\mu^{S''}, \tau^{S\setminus S''}, \sigma^{N\setminus S})].$$

This means that $\mu^{S''}$ is an improvement of $(\tau^S, \sigma^{N\setminus S})$ as well. Similarly, improvements of $(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})$ are also improvements of $(\mu^{S''}, \tau^{S\setminus S''}, \sigma^{N\setminus S})$ etc., so that if $\mu^{S''}$ is an ICI of $(\nu^{S_y}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})$, then it is also an ICI of $(\tau^S, \sigma^{N\setminus S})$, and we are done.

(v) The final part of the argument consists in eliminating all other cases than the simple one treated under (iv) above. For this, we make use of the the elaborate definition of the selection function φ chosen by individual 1 in σ^1.

Thus, assume that only the components $\hat{b}^i$ and $(\hat{t}^i, \hat{P}^i)$ have been changed in the strategy μ^i, for $i \in S''$, so that $S[(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})] = S$. Let

$$\hat{B} = B[(\mu^{S''}, \nu^{S_y\setminus S''}, \tau^{S\setminus S_y}, \sigma^{N\setminus S})].$$

From the minimality property of S_y we know that no proper subcoalition S_z of S_y is such that $zR^{S_z}q$ and not u-dominated at $(T^S, R^{N\setminus S})$ via $S\setminus S_z$. Since the strategy changes leading to y consisted only of exempting y from domination, we have that no subcoalition of S_y can achieve an alternative $z \notin C$. Thus, $z \in C$.

Suppose that $y \notin \hat{B}$. Then, by our definition of $\tilde{b}^i$ for $i \in S_y$, we have that $\mu^{S''}$ is an improvement upon $(\tau^S, \sigma^{N\backslash S})$ via S''. If $\mu^{S''}$ is an ICI upon $(v^{S_y}, \tau^{S\backslash S_y}, \sigma^{N\backslash S})$ then it is also an ICI upon $(\tau^S, \sigma^{N\backslash S})$. Thus, we may assume that $\{y, z\} \subset \hat{B}$.

It follows from $\{y, z\} \subset \hat{B}$ that $(S, \hat{B}) \notin \mathscr{D}^3$. We cannot have $(S, \hat{B}) \in \mathscr{D}^1$, since in that case $\varphi(S, \hat{B})$ could not choose $z \in C$. Thus, $(S, \hat{B}) \in \mathscr{D}^2$, and φ selects an undominated and $\succ$-minimal admissible pair $\{z, w\}$ containing z. Now zR^Sq, and if also wR^Sq, then $\{z, w\}$ is dominated by $\{y, q\}$, a contradiction. Therefore, qR^iw for some $i \in S$, and $\{q, w\}$ is also an admissible pair. Assume that $\{q, w\}$ is dominated, meaning that there is another admissible pair $\{w', w''\}$ with qR^Sw', wR^Sw'; we cannot have that zR^Sw' since in that case $\{z, w\}$ would also be dominated, consequently $\{z, w'\}$ is admissible; if $\{z, w'\}$ is dominated, then so is $\{z, w\}$, contradiction, so $\{z, w'\}$ is undominated; however, $\{z, w\} \succ \{z, w'\}$, contradicting $\succ$-minimality. We conclude that $\{q, w\}$ is undominated, and then we get another contradiction of the $\succ$-minimality of $\{z, w\}$. We conclude that $\varphi(S, \hat{B})$ cannot be a pair containing z. This establishes the claim that only the case considered in (iv) can occur, and therefore we have established the claim stated in (iii).

Summing up, we have shown that if $S[\tau^S, \sigma^{N\backslash S})] = S$ then τ^S is not an ICI of σ^N. Suppose now that $S[(\tau^S, \sigma^{N\backslash S})] = \tilde{S}$ is a proper subset of S. Then

$$\pi(\tau^{\tilde{S}}, \sigma^{N\backslash \tilde{S}}) = \pi(\tau^S, \sigma^{N\backslash S}) = q,$$

and the restriction $\tau^{\tilde{S}}$ of τ^S is an improvement of $(\sigma^{\tilde{S}}, \tau^{S\backslash \tilde{S}}, \sigma^{N\backslash S})$. By the preceding arguments, we know that there is a subcoalition $\hat{S}$ of $\tau^{\tilde{S}}$ having an ICI of $(\tau^{\tilde{S}}, \tau^{S\backslash \tilde{S}}, \sigma^{N\backslash S})$. But this means that τ^S is not an ICI of σ^N. □

The remaining cases to be treated in order to prove that σ^N is indeed a coalition proof Nash equilibrium, namely those of coalitions containing the individual 1, are much simpler to deal with.

Lemma 7.3. *Let $S \in 2^N$ be a coalition with $1 \in S$. Then S has no internally consistent improvement of σ^N.*

Proof. Suppose that the coalition S with $1 \in S$ has an improvement τ^S of σ^N with $\pi(\tau^S, \sigma^{N\backslash S}) = q$. From the definition of x and the fact that qR^1x, we have that q does not belong to any minimal set B such that some S is u-effective for B. In particular, q does not belong to the set $\bar{B}$ used in the definition of the preference component Q^i of the strategies σ^i, $i = 1, \ldots, n$.

Let $\tau^1 = (\tilde{Q}^1, \tilde{\varphi}^1, \tilde{x}^1)$. If $\tilde{\varphi}^1 = \varphi$ and $\tilde{x}^1 = x$, then $S[(\tau^S, \sigma^{N\backslash S})] \subset S$ and $1 \notin S[(\tau^S, \sigma^{N\backslash S})]$, and $S[(\tau^S, \sigma^{N\backslash S})]$ has an improvement of σ^N. By lemma 6.2 this improvement is not internally consistent, and the same must hold for τ^S.

Suppose now that $\tilde{\varphi}^1 \neq \varphi$ or $\tilde{x}^1 \neq x$. Then $N\backslash S$ is contained in $S[(\tau^S, \sigma^{N\backslash S})] = \hat{S}$. Let $(\bar{B}, \bar{S}) \in 2^A \times 2^N$ be the pair used in the definition of Q^i for $i \neq 1$. Since $q \notin \bar{B}$ we have that there is $S' \subset \bar{S}$ so that q is u-dominated at $Q^{S'}$ and $\bar{B}R^{S'}q$. From $x \in \bar{B}$ we infer that $S' \cap S = \varnothing$. But then $R[(\tau^S, \sigma^{N\backslash S})]$

can be written as $(Q^{S'}, Q^{\hat{S}\setminus S'})$, and since q is u-dominated at this profile, we have a contradiction. □

The results of Lemma 7.1, 7.2, and 7.3 may be summarized as follows:

Theorem 7.4. *Let $R^N \in L^N$, and let $x = F(R^N)$. Then there exist strategies $\sigma^1, \dots, \sigma^n$ such that $\sigma^N = (\sigma^1, \dots, \sigma^n)$ is a CNE in G, and $\pi(\sigma^N) = x$.*

Combining Theorem 7.4 and Lemma 6.1 together with Remark 6.2, we get the desired converse of Theorem 4.11:

Theorem 7.5. *Let $E : P(N) \to P^2(A)$ be an effectivity function which is maximal and superadditive. Then there is a game form $G = (\Sigma^1, \dots, \Sigma^n, \pi)$ which represents E, i.e.*

(i) *E is α- and β-associated with G, $E = E_\alpha^G = E_\beta^G$,*
(ii) *for each $R^N \in L^N$, the game (G, R^N) has a coalition proof Nash equilibrium.*

The result may also be formulated in terms of implementation:

Corollary 7.6. *Let $E : P(N) \to P^2(A)$ be an effectivity function which is maximal and superadditive. Then there is an SCC $H : L^N \to 2^A$ such that*

(i) *H is CNE-implementable,*
(ii) *E is α- and β-associated with H, $E = E_\alpha^H = E_\beta^H$.*

It may be noticed that the maximality property of the effectivity function E is used only to establish the second part of Lemma 6.1, which in its turn is invoked only to obtain the final result of Theorem 7.5.

8 Concluding comments

In this paper we provide a complete solution to the problem of representation of effectivity functions in coalition proof equilibria: For each maximal and superadditive effectivity function E we can find a coalition proof Nash consistent game form G such that $E_\alpha^G = E$. Furthermore, G is tight, that is $E_\alpha^G = E_\beta^G$.

This result has two immediate applications. First, if we model rights-systems by effectivity functions (Gärdenfors [8]), then existence of representations is essential for possible consistent behavior of the members of a society (who obey the rights-system). If an effectivity function E is the constitution of a society, then the members of the society can exercise their rights simultaneously only if E has a consistent representation (Peleg [15]). Clearly, we treated only one solution, namely, coalition proof Nash equilibrium.

Second, our result supplies a simple necessary condition for implementability in coalition proof Nash equilibria (see Remark 4.13). In particular, the Pareto correspondence is not CNE-implementable in our case of unrestricted finite domains. Thus, our results are independent of those of Boylan [6] (see, again, Remark 4.13).

Our result is comparable to the main result of Moulin and Peleg [12]. We recall that Moulin and Peleg [12] prove that an effectivity function is representable in strong Nash equilibrium if and only if it is stable and maximal (see also Peleg [13], Theorem 6.4.4). Finally, our work is linked to the works of Gurvich [9] and Abdou [1] on Nash-consistency of two-person game forms, because a two-person game form is Nash-consistent if and only if it is CNC (see Remark 4.12).

References

[1] Abdou J (1995) Nash and strongly consistent two-player game forms. Int J Game Theory 24: 345–356
[2] Abdou J, Keiding H (1991) Effectivity functions in social choice. Kluwer Academic Press, Dordrecht
[3] Aumann RJ (1959) Acceptable points in general cooperative *n*-person games. In: Tucker AW, Luce RD (eds) Contributions to the Theory of Games IV. Princeton University Press, Princeton, 287–324
[4] Aumann RJ (1974) Subjectivity and correlation in randomized strategies. J Math Econ 1: 67–96
[5] Bernheim BD, Peleg B, Whinston MD (1987) Coalition-proof equilibria. I. Concepts. J Econ Theory 42: 1–12
[6] Boylan RT (1998) Coalition-proof implementation. J Econ Theory 82: 132–143
[7] Dasgupta PS, Hammond PJ, Maskin ES (1979) The implementation of social choice rules: Some general results on incentive compatibility. Rev Econ Stud 46: 185–216
[8] Gärdenfors P (1981) Rights, games, and social choice. Noûs 15: 341–356
[9] Gurvich VA (1975) Solvability in pure strategies, Zh. Vychisl. Mat. i Fiz. 15: 358–371; English translation in USSR Comput Math and Math Phys 15: 1975
[10] Hurwicz L (1972) On informationally decentralized systems. In: Radner R, McGuire CB (eds) Decision and organization, Vol. in honor of J. Marschak. North-Holland, Amsterdam, 297–336
[11] Maskin ES (1977) Nash equilibrium and welfare optimality, mimeo, M.I.T
[12] Moulin H, Peleg B (1982) Cores of effectivity functions and implementation theory. J Math Econ 10: 115–145
[13] Peleg B (1984) Game theoretic analysis of voting in committees. Cambridge University Press, Cambridge
[14] Peleg B (1984) Quasi-coalitional equilibria, part I: Definitions and preliminary results. Research Memorandum No. 59. The Hebrew University, Jerusalem
[15] Peleg B (1998) Effectivity functions, game forms, games, and rights. Soc Choice Welfare 15: 67–80