Soc Choice Welfare (1997) 14: 513–525

Social Choice and Welfare

Non-cooperative implementation of the core

Roberto Serrano, Rajiv Vohra

Department of Economics, Brown University, Providence, RI 02912, USA

Received: 14 September 1994/Accepted: 16 April 1996

Abstract. The aim of this paper is to examine the non-cooperative basis of the core. We provide mechanisms that are motivated closely by the very description of the core, and have the property that their non-cooperative equilibrium outcomes coincide with the core. For general economic environments we construct an extensive form mechanism in which each player proposes a status-quo and then also has an opportunity to recontract with any other coalition. A proposal to recontract is enforced if and only if it meets with the unanimous approval of such a coalition. We show that subgame perfect outcomes of this mechanism coincide with the core allocations of the underlying economy. We also consider situations, such as labor managed firms, in which the mechanism designer does not know the set of feasible allocations but can observe the output (utility).

1. Introduction

As a solution concept in cooperative game theory, the core occupies a central position. In economic environments the core, pioneered by Edgeworth (1881), provides an important resolution to the collective problem of resource allocation. The notion of core allocations as those that are immune to recontracting or to objections from all possible coalitions captures rather naturally the idea of coalitional stability. However, the concept of the core does not make explicit the interaction of coalitions or individual agents that yields the final outcomes, and this is the main focus of the present paper.

If we believe that agents play strategically (non-cooperatively), then the process through which the core allocations emerge is not clear. The issue is

* The authors thank Andreu Mas-Colell, Debraj Ray and an anonymous referee for many helpful comments. Vohra is also grateful for support from National Science Foundation grant SBR-9414142 and a Fulbright Research Award.

one of formulating a non-cooperative framework to support core outcomes. We shall consider private ownership economies as well as general games in characteristic function form. In each case we shall construct a game form that is closely tied to the notion of the core and implements the core in non-cooperative equilibria. In doing so we build upon two different strands of the literature: (1) the general approach of implementation in economic environments in which the feasible sets are known to the designer but utility functions are not, and (2) the literature that considers games in characteristic function form and attempts to provide non-cooperative underpinnings of the core through bargaining in an extensive form.

The general problem of implementing a social choice correspondence through Nash equilibrium has been extensively analyzed. The classic reference is Maskin (1977); see also Maskin (1985) and Moore (1992) for two excellent surveys. Given a social choice correspondence that maps from the space of environments to the space of alternatives (feasible allocations), the problem is one of designing a mechanism or a set of rules such that in every environment the socially desired outcomes coincide with those obtained through strategic behavior on the part of the agents.

The usual approach in this literature is to assume that the set of feasible allocations is known but agents' preferences are unknown to a plannner or mechanism designer. Maskin (1977) showed that any social choice correspondence that can be implemented in Nash equilibrium must satisfy a condition known as monotonicity. In economic environments, such as private ownership economies, monotonicity is also a sufficient condition for Nash implementation. A canonical mechanism was constructed to show that every monotonic social choice correspondence (for example, the Pareto, the constrained Walrasian or the core correspondence) can be implemented.

It is natural to expect that a particular social choice correspondence may be better implemented through a tailor-made mechanism that is more appealing than the general canonical mechanism which involves each agent reporting the entire profile of preferences. Indeed, the (constrained) Walrasian correspondence can be implemented through mechanisms that are based on the Walrasian notions of allocations and prices; see Schmeidler (1980), Hurwicz (1979), and Hurwicz et al (1995). It is intriguing that a similar result on the core is conspicuous by its absence from the literature.[1]

In Sect. 2 we consider general economic environments, including both private goods production economies as well as public goods economies. We construct an extensive form mechanism in which each player proposes a status-quo allocation and then also has an opportunity to recontract with any coalition. The mechanism consist of two stages. In the first stage each player chooses a feasible allocation. If there is unanimous agreement on this allocation (the status-quo) the game moves to the sescond stage.[2] We construct the second stage in such a way that any player who can form an objecting coalition has the ability to make such a proposal. Other members of

[1] One aspect of the problem is illustrated by a result of Dutta et al (1995) which implies that the core cannot be implemented through a mechanism that relies on strategy sets consisting only of allocations (or quantities) and prices.

[2] The mechanism is constructed to ensure that in equilibrium there is unanimous agreement on the status-quo.

such a coalition then have an opportunity to respond sequentially to this proposal.[3] If they all agree, this proposal is enforced and players outside this coalition receive their initial endowments along with any public goods produced by the formed coalition. If the proposal is not unanimously accepted, the outcome is the status-quo. It should be quite clear that such a mechanism has the appeal of being motivated quite closely by the very story underlying the notion of the core. We show that an allocation is a subgame perfect outcome of this mechanism if and only if it belongs to the core of the underlying economy.[4]

It is relevant to compare our mechanism to those proposed by Wilson (1978) and Kalai et al. (1979). Wilson (1978) formulated a two stage, competitive bidding game in which one subgame perfect outcome belongs to the core of the underlying exchange economy. In the bidding game, one player acts as the auctioneer who chooses from proposals of net trades offered by the other players. Unfortunately, Wilson's mechanism does not achieve full implementation – there may exist subgame perfect outcomes that do not belong to the core and there may exist core allocations that do not correspond to a subgame perfect outcome. Kalai et al. (1979) constructed a normal form mechanism that implements the core in strong equilibrium. But the Nash equilibria of their game correspond to the entire set of individually rational allocations.

Quite distinct from the theory of implementation is a growing literature that analyzes the non-cooperative underpinnings of the core. The idea is to construct detailed rules of bargaining in an extensive form game and show that the set of non-cooperative equilibria (either subgame perfect Nash or some refinement thereof) coincide with the core. The planner or mechanism designer is not formally incorporated in this analysis and it is not always apparent what the informational requirements are. In particular, coalitions are always assumed to choose feasible and efficient allocations. The aim of this branch of the literature is to underscore new, non-cooperative aspects of core outcomes.

Building on the analysis of coalitional bargaining in extensive form games in Selten (1981) and Chatterjee et al. (1993), several recent papers have focused more directly on the connection between equilibria of such games and the core. For instance, Perry and Reny (1994), Moldovanu and Winter (1995), Winter (1992) and Winter (1993) obtain the core as the outcome of stationary perfect equilibrium play in different extensive forms, all of which model coalitional bargaining. Alternatively, Serrano (1995) presents an extensive form that resembles a market, where, the core is supported in subgame perfect equilibrium and obtained as those outcomes in which every profit opportunity has vanished from the market.[5] Since the subgame perfect equilibria of

[3] We leave the formal details of this construction to Sect. 2.

[4] It is worth pointing out that the canonical mechanisms for subgame perfect implementation (Moore and Repullo (1988), Abreu and Sen (1990) require players to report profiles of utility functions.

[5] The underlying structure is a TU game in Perry and Reny (1994) and in Serrano (1995); an NTU game in Moldovanu and Winter (1995) and in Winter (1992); and a committee problem in Winter (1993). Bloch (1992) and Ray and Vohra (1995) consider a more general model that allows for externalities across coalitions and analyze an extensive form game of coalition formation. An alternative, but related, approach is followed by Lagunoff (1994) who analyzes allocation specific mechanisms that strategically separate core allocations from non-core allocations.

such games depend crucially on the order in which players move, it is not surprising that all these authors have to rely on some mechanism for making players symmetric in this respect. Moldovanu and Winter (1995) do this by restricting attention only to those equilibria which are "order independent"; Serrano (1995) considers a game in which the order of moves is random; Perry and Reny (1994) consider a dynamic game in continuous time which allows for any players to move at any time.

Our analysis of general NTU games in Sect. 3 is directly related to this literature. We consider a situation in which the mechanism designer does not know the characteristic function but can observe the utilities. We construct a normal form mechanism that implements the core in Nash equilibrium. The normal form provides a relatively simple way of dealing with the issue of anonymity.[6] Players announce a proposed status-quo and possible objections. To allow any player to form an objecting coalition we also require each player to announce a permutation of the player set. The composition of these permutations then determines the order in which coalitioins are actually formed. The idea of using permutations as part of the strategies was used recently by Thomson (1992) to construct simple mechanisms for implementing various solutions to the fair division problem. As will be clear in the following Sections, our results rely very heavily on this construction.

There is a more compelling motivation for our analysis of general NTU games. While the economic environments considered in Sect. 2 are very general, the model (as in the typical formulation of the implementation problem) is one in which the feasible sets are known to the designer but preferences are not. The case in which the feasible sets are unknown but payoffs are observable is also of significant economic interest. For instance, consider the situation in which a labor managed firm produces a single (observable) output and the mechanism designer does not know the productivities of the agents, or the technology. This is one important example of the more general implementation problem in the context of an NTU game in which the designer can observe utilities but does not know the characteristic function.[7] Our approach in Sect. 3 can also be used to extend our result on private ownership economies to cover the case in which the designer has no knowledge of the utility functions, endowments or the technology.

2. Private ownership economies

To consider economic environments to which the notion of the core can be applied we consider economies in which each agent has an initial (private) endowment. We refer to such an economy as a private ownership economy. The model is general enough to include production as well as public goods.

Suppose there are l commodities, one commonly available technology and a set of consumers $N = \{1, \dots, n\}$. A private ownership economy $\mathbf{E}$ is defined

[6] It is interesting to note that a similar normal form mechanism cannot be used for implementing the core in private ownership economies; see Remark 5 in Sect. 3.

[7] Hurwicz, Maskin and Postlewaite (1995) is our closest precursor in this respect. See also Postlewaite and Wettstein (1989)

as $\mathbf{E} = \{(X_i, u_i, \omega_i)_{i \in N}, Y\}$, where $X_i \subseteq \mathbf{R}^l$, $u_i : X_i \mapsto \mathbf{R}$ and $\omega_i \in X_i$ refer to consumer i's consumption set, utility function and endowment respectively. For public goods economies we make the usual assumption that no consumer has an endowment of any public good. The commonly available technology is represented by the production set $Y \subseteq \mathbf{R}^l$. We shall assume that $-\mathbf{R}^l_+ \subseteq Y$ (free disposal) and that for every $i \in N$, $\omega_i > 0$. We will use the convention $\gg$, $>$, $\geq$ to order vectors. Let $\mathcal{N}$ denote the set of all non-empty subsets (coalitions) in N. For $S \in \mathcal{N}$, we use $-S$ to denote the complement of S. Given a collection of vectors or sets, one for each consumer, we will use subscripts to refer to their restrictions to a particular coalition. For example, $X_N = \prod_{i \in N} X_i$, $X_S = \prod_{i \in S} X_i$, and given $(x_i) \in X_N$, $x_S = (x_i)_{i \in S}$ and $x_{-S} = (x_i)_{i \notin S}$. We will denote by $u_S(x_S)$ the profile of utilities $(u_i(x_i))_{i \in S}$. For the grand coalition we will use $u(x)$ to denote $u_N(x_N)$.

We shall assume that the designer knows the production set and the endowments of the consumers but not their utility functions.[8] Thus X_i, Y and ω_i will remain fixed and economies will be distinguished simply by the utility functions of the consumers. Let $\mathscr{E}$ denote the class of economies in which for all $i \in N$, u_i is strictly monotonic in the sense that $u_i(x_i) > u_i(x_i')$ if $x_i > x_i'$.

Each coalition S has a feasible set of consumption plans $A_S \subseteq X_S$ with the property that $\omega_S \in A_S$. Additional properties of A_S, or its precise form, will not be critical in what follows, but we note that if all goods are private goods,

$$A_S = \left\{ x_S \in X_S \,\middle|\, \sum_{i \in S} x_i \in Y + \left\{ \sum_{i \in S} \omega_i \right\} \right\}.$$

From this it should also be clear that our model includes cost sharing problems as well.

The *core* of an economy $\mathbf{E}$ is defined as

$$C(\mathbf{E}) = \{x \in A_N \mid \nexists S \in \mathcal{N} \text{ and } y_S \in A_S \text{ such that } u_S(y_S) > u_S(x_S)\}.$$

We shall construct an extensive form mechanism such that for every economy in $\mathscr{E}$, the subgame perfect outcomes of the game coincide with the core allocations.

An *extensive game form or mechanism* is defined as a game tree with possibly simultaneous moves, i.e., as an array $\Gamma = (N, K, g)$, where N is the set of players, K a game tree and $g : Z \mapsto A_N$ is the outcome function, where Z denotes the set of terminal nodes of the tree K. We will use $g(z)_i$ to denote consumer i's commodity bundle corresponding the allocation $g(z)$. The set of nodes of the tree K is denoted T. The initial node is t_0. Let M_i^t denote the set of choices available to player i at node t and let M_i denote the set of strategies of player i. For a strategy profile $m \in M_N$, let $g(m, t)$ denote the outcome corresponding to m starting at node t. Given an economy $\mathbf{E} = \{(X_i, u_i, \omega_i), Y\}$, the mechanism Γ defines an extensive form game $(\Gamma, \mathbf{E})$, where the payoff to the players corresponding to the strategy profile m is $u(g(m, t_0))$.

A *subgame perfect* equilibrium of a game $(\Gamma, \mathbf{E})$ is a strategy profile $\bar{m} \in M_N$ such that for all $t \in T \backslash Z$ and for all $i \in N$,

$$u_i(g(\bar{m}, t)_i) \geqslant u_i(g(\bar{m}_{-i}, m_i, t)_i) \text{ for all } m_i \in M_i.$$

[8] For the case in which the technology and endowments are also unknown to the designer, see Remark 6 at the end of the next section

Let SPE(Γ, **E**) denote the set of all allocations corresponding to subgame perfect equilibria of the game (Γ, **E**).

An extensive game form Γ is said to *implement in subgame perfect equilibrium* the core in all economies over the class $\mathscr{E}$ if

$\text{SPE}(\Gamma, \mathbf{E}) = C(\mathbf{E})$ for all $\mathbf{E} \in \mathscr{E}$.

We need some additional notation before defining our mechanism. For every i, pick $\varepsilon_i \in \mathbf{R}^l_+$ such that $\omega_i - \varepsilon_i \in \mathbf{R}^l_+$. This is possible since $\omega_i > 0$ for all i. Let Π denote the set of all permutations of N, i.e. one-to-one functions from N to N. Given $\pi = (\pi_i)$, where $\pi_i \in \Pi$ for every $i \in N$, define $p(\pi)$ to be the composition of the permutations (π_i), i.e., $p(\pi) = \pi_1(\pi_2(\ldots (\pi_i(\ldots \pi_n) \ldots)$. The i-th element of $p(\pi)$ will be denoted $p(\pi)_i$. Notice that for every $i \in N$, given π_{-i} and $\pi^* \in \Pi$, there exists $\pi'_i \in \Pi$ such that $p(\pi'_i, \pi_{-i}) = \pi^*$. In particular, any $i \in N$ can make a unilateral change in π_i to make him/herself the first player in the order p. This particular construction was recently used by Thomson (1992) to construct mechanisms for Nash implementation in the context of solutions to the problem of fair division. We shall interpret $p(\pi)$ to be an endogenously determined protocol in our extensive form game. While this construction is reminiscent of integer or modulo game constructions[9] it does capture quite naturally the anonymity that we need to build in the game form to implement the core. As an alternative it is possible to use an integer game construction to determine the order of moves[10] but we believe that in the present context permutations bring out more clearly the role of a protocol that is symmetric (anonymous) with respect to the players.

We shall construct an extensive game form consisting of two stages. In Stage 0, every player i chooses simultaneously from the choice set $M^0_i = A_N \times \Pi$. A typical choice of player i will be denoted $m^0_i = (x^i, \pi_i)$. Note that x^i refers to player i's announcement of an allocation, i.e., $x^i = (x^i_j)_{j \in N}$. We will generally use superscripts to denote an agent's announcement of a profile. Let $m^0 = (m^0_i)$ represent the profile of Stage 0 messages and let $1(m^0) = p(\pi)_1$ and $n(m^0) = p(\pi)_n$ denote the first and the last players according to the order $p(\pi)$. If for any i and j, $x^i \neq x^j$, the outcome is that player $n(m^0)$ receives $\omega_{n(m^0)} - \varepsilon_{n(m^0)}$, and all other players receive their initial endowments. If $x^i = x^j = x^*$ for all i and j in N, proceed to Stage 1. In this case we will refer to x^* as the status-quo.

In Stage 1, player $1(m^0)$ chooses a coalition S containing $1(m^0)$, and $y \in A_S$. Let $S = \{1(m^0), 2(m^0), \ldots, k(m^0)\}$, where $j(m^0)$ refers to the j-th player in S according to the order $p(\pi)$. The other members of S then respond sequentially to this proposal (starting with player $2(m^0)$ and going up to $k(m^0)$) by either accepting it or rejecting it. If all members of S accept y, coalition S is assigned y and all players not in S are assigned their initial endowment along with any public goods produced by coalition S. More precisely, for a public goods economy, let y_p denote the public goods corresponding to $y \in A_S$. Then if m is a strategy profile such that coalition S accepts the proposal y, the final outcome $g(m, t_0) = (y, \omega_{-S})$ in a private goods economy and $(y, (\omega_{-S}, y_p))$ in

[9] See Jackson (1992) for a critique.

[10] It seems that permutations play far more direct role in Thomson's mechanisms where they are used to permute commodity bundles among the players.

a public goods economy. If any player in S rejects the proposal, the final outcome is x^*. This completes the description of the mechanism. Notice that the mechanism is feasible in the sense that $g(m, t_0) \in A_N$ for every $m \in M_N$.

Theorem 1. *The extensive form mechanism* Γ *implements in subgame perfect equilibrium the core in the class of economies* $\mathscr{E}$.

Proof. We begin by showing that if $x^* \in C(\mathbf{E})$, then $x^* \in \mathrm{SPE}(\Gamma, \mathbf{E})$. Consider the strategy profile $\bar{m}$ defined as follows:

(i) $\bar{m}_i^0 = (x^*, \pi^e)$ for all i, where π^e denotes the identity permutation;

(ii) every player i chooses $(S, y) = (N, x^*)$ at every node of Stage 1 where i has to make a proposal;

(iii) at every node of Stage 1 where i has to respond to a status-quo x and a proposal y, i accepts if and only if $u_i(y_i) > u_i(x_i)$.

Clearly, condition (iii) conforms to subgame perfection in all subgames starting at a node where a proposal has to be responded to. Since $x^* \in C(\mathbf{E})$, no player can propose an objection to x^* and this implies that (ii) conforms to subgame perfection in all subgames starting at a node where a player has to make a proposal to the status-quo x^*. Finally, note that since $x^* \in C(\mathbf{E})$, no player i can gain by choosing $x^i \neq x^*$. Thus $\bar{m}$ is a subgame perfect equilibrium. The path corresponding to $\bar{m}$ is one where the status-quo from Stage 0 is x^*, player 1 proposes the grand coalition and x^*, and all other players reject.[11] Thus $g(\bar{m}) = x^*$ and $x^* \in \mathrm{SPE}(\Gamma, \mathbf{E})$.

We now proceed to show that if $\bar{m}$ is a subgame perfect equilibrium of $(\Gamma, \mathbf{E})$, then $g(\bar{m}) \in C(\mathbf{E})$. Consider a subgame perfect equilibrium $\bar{m}$ such that $\bar{m}_i^0 = (x^i, \pi_i)$. Let $\bar{z} = g(\bar{m})$.

Claim 1. $x^i = \bar{x}$ *for all* $i \in N$. Suppose not. Then player $j = n(\bar{m}^0)$ receives $\omega_j - \varepsilon_j$. However, this player can gain by changing π_j to π_j' such that $j \neq p(\pi_{-j}, \pi_j')_n$. A deviation from $\bar{m}_j$ with just such a change in π_j will result in j receiving ω_j instead of $\omega_j - \varepsilon_j$, which contradicts the hypothesis that $\bar{m}$ is a subgame perfect equilibrium.

Claim 2. $u(\bar{z}) \geq u(\bar{x})$. Notice that given π_{-j}, by a suitable choice of π_j', player j can make sure that $j = p(\pi_{-j}, \pi_j')_1$. Suppose that by such a choice j becomes the first player in the order p and then proposes $(N, \bar{x})$ in Stage 1. Irrespective of how the others respond, the outcome is $\bar{x}$. Thus every player can make a unilateral deviation to ensure the status-quo $\bar{x}$. Since $\bar{z} \in \mathrm{SPE}(\Gamma, \mathbf{E})$, the claim follows.

Claim 3. *Let* t *be a node of any subgame in which a player has to respond to the status-quo* $\bar{x}$ *and a proposal* (S, y) *made by* $i \in S$. *Suppose* $u_j(y_j) > u_j(\bar{x}_j)$ *for all* $j \neq i$, $j \in S$. *Then* $g(\bar{m}, t) = (y, \omega_{-S})$ *in a private goods economy and* $g(\bar{m}, t) = (y, (\omega_{-S}, y_p))$ *in a public goods economy.* This is a straightforward consequence of the subgame perfectness of $\bar{m}$ — a responder must accept any proposal that is better than the status-quo.

To complete the proof, suppose $\bar{z} \notin C(\mathbf{E})$. Then there exists an objection (S, y) to $\bar{z}$. Since preferences are monotonic, there exists an objection (S, y')

[11] There exists another subgame perfect equilibrium supporting x^*; one which differs from $\bar{m}$ only in having a weak inequality rather than a strict inequality in condition (iii).

such that $u_S(y') \gg u_S(\bar{z}_S)$. Since, by Claim 2, $u(\bar{z}) \geq u(\bar{x})$, it follows that $u_S(y') \gg u_S(\bar{x}_S)$. From Claim 3 we know that, corresponding to $\bar{m}$, in every subgame following the proposal (S, y'), the outcome is (y', ω_{-S}) in a private goods economy and $(y', (\omega_{-S}, y'_p))$ in a public goods economy. This means that any player $i \in S$ can unilaterally change the outcome from $\bar{z}$ to (y', ω_{-S}), or to $(y', (\omega_{-S}, y'_p))$ in a public goods economy, simply by changing π_i to become the first player and by proposing (S, y'). Since $u_S(y') \gg u_S(\bar{z}_S)$, this contradicts the hypothesis that $\bar{m}$ is a subgame perfect equilibrium, and completes the proof that $g(\bar{m}) \in C(\mathbf{E})$. □

Remark 1. Conditions for the non-emptiness of the core have been investigated by Scarf (1967) and Scarf (1986). It follows from Theorem 1 that in an economy with an empty core, the set of subgame perfect equilibria (in pure strategies) of the mechanism Γ is also empty. In economies that do not have a superadditive structure, for example economies with decreasing returns to scale, it is also of interest to consider the coalition structure core. Our mechanism can be easily modified to implement the coalition structure core. Let $\mathscr{P}$ denote the set of all partitions (coalition structure) of N. The feasible set for a partition $P \in \mathscr{P}$ is defined as

$$A_P = \{x \in X \mid x_S \in A_S \text{ for all } S \in P\}.$$

Let $A_{\mathscr{P}} = \bigcup_{P \in \mathscr{P}} A_P$ and define $M_i^0 = A_{\mathscr{P}} \times \Pi$ for all i. It is easy to see that with this modification the mechanism Γ implements in subgame perfect equilibrium the coalition structure core in $\mathscr{E}$. The interpretation of a status-quo now is a feasible allocation for some coalition structure.

Remark 2. It should be noted that our mechanism leads to wasteful outcomes in case consumers disagree about the status-quo. It is possible to modify our mechanism to ensure non-wastefulness provided $n \geq 3$ (recall that our result imposes no restrictions on n). This can be done as follows. In case $n-1$ consumers announce x^* and player i announces $x^i \neq x^*$, all players are assigned their initial endowment and ε_i is transferred from player i to some arbitrary player. In all other cases in which a unanimous status-quo is not announced, $\varepsilon_{n(m^0)}$ is transferred from the last player to some arbitrary player. The rest of the mechanism remains unchanged.

Remark 3. Strict monotonicity of preferences is used in the proof of Theorem 1 only in the last paragraph of the proof to argue that if there exists an objection (S, y) to $\bar{z}$, then there also exists an objection (S, y') such that $u_S(y') \gg u_S(\bar{z}_S)$. Thus, it is possible to relax the assumption that all utility functions are strictly monotonic by defining objections with strict inequalities. Note however, that in this case footnote 11 may no longer apply.

3. NTU Games

The primary motivation for this Section comes from a situation in which there is a single observable output but the endowments and technology are unknown to the mechanism designer. Think of a simple labor managed firm consisting of n agents. The maximum output of a coalition depends on the technology as well as the endowments and the productivity of its members. If the mechanism designer is not fully informed about any of these variables, she cannot construct the characteristic function for the underlying game, even if the output (utility) is observable. This situation, therefore, is the motivating example for the study of the implementation problem for NTU games in characteristic function form. We will assume that while the mechanism designer does not have full knowledge of the characteristic function, utilities are nevertheless observable.

We shall consider games with a finite set of players $N = \{1, \dots, n\}$. For every $S \in \mathcal{N}$ denote by $\mathbf{R}^S$ the $|S|$ dimensional Euclidean space with coordinates indexed by the elements of S. With a slight abuse of notation we will now use $x \in \mathbf{R}^N$ to denote a profile of utilities rather than an allocation. Our other notational conventions remain unchanged (for example, for $x \in \mathbf{R}^N$, x_S refers to its restriction on $\mathbf{R}^S$). For every coalition $S \in \mathcal{N}$, the set of feasible utilities is $V(S) \subseteq \mathbf{R}^S$. An NTU game in characteristic function form is defined by V such that for every $S \in \mathcal{N}$, $V(S) \subseteq \mathbf{R}^S$. We shall consider the class of games satisfying the following assumption:

(A) For all $S \in \mathcal{N}$, $V(S) \subseteq \mathbf{R}^S$ is non-empty, comprehensive (i.e. $V(S) = V(S) - \mathbf{R}^S_+$) and contains 0.

Let $\mathcal{P}$ denote the set of all partitions (coalition structures) of N. The feasible set of a coalition structure P is defined as

$$V(P) = \{x \in \mathbf{R}^N \mid x_S \in V(S) \text{ for all } S \in P\}.$$

The *core of a coalition structure* P of a game V is defined as

$$C(P, V) = \{x \in V(P) \mid \nexists T \in \mathcal{N} \text{ and } y_T \in V(T) \text{ such that } y_T > x_T\}$$

and the *coalition structure core* of a game V is defined as

$$C(\mathcal{P}, V) = \bigcup_{P \in \mathcal{P}} C(P, V).$$

The *core of a game* V, $C(N, V)$, refers to the core of the grand coalition.

A (*normal form*) *mechanism* consists of strategy sets $(M_i)_{\{i \in N\}}$ and an outcome function $g : M_N \mapsto \mathbf{R}^N$ such that for every game V and $m \in M_N$, $g(m) \in V(\mathcal{P})$.

A mechanism $(g, (M_i))$ is said to *implement the core of a coalition structure*[12] *in Nash equilibrium* if for every game (N, V), $\bar{x} \in C(\mathcal{P}, V)$ if and only if there exists $\bar{m} \in M_N$ such that $g(\bar{m}) = \bar{x}$ and

$$\text{for every } i \in N, \quad \bar{x}_i \geq g(m_i, \bar{m}_{-i})_i \text{ for all } m_i \in M_i.$$

The main difference between our model and that of Hurwicz et al. (1995) is that we are interested in implementation not in an economic environment but in the context of an NTU game. However, the critical difficulty in both cases is similar. It will be instructive to discuss this briefly, before we turn to a description of our mechanism. Hurwicz et al. (1995) showed, when the feasible sets are unknown, it is not possible to carry out successful implementation without making the strategy sets of the agents depend on the feasible set. In our

[12] Our approach can be easily modified to allow for the implementation of the core or the core of any given coalition structure.

context too, an analogous restriction on the strategy sets can be derived. Of course, if the designer can infer the entire characteristic function from the strategy sets, the implementation exercise will become meaningless. We shall, therefore, consider a mechanism in which player i's strategies include elements of $V(S_i)$, where $i \in S_i$. The interpretation is that while players do not reveal the entire set $V(S_i)$ to the designer, they cannot lie about feasibility. When a player claims that $x \in V(S_i)$, the designer can verify whether or not this is true. This interpretation is similar to that in Hurwicz, Maskin and Postlewaite (1995), where it is assumed that players can withhold but cannot exaggerate their endowments. In case of a false claim, the designer may punish a player. "Punishment" will take the form of assigning a player to a singleton coalition and may also include a small penalty. More precisely, let $w_i \geqslant 0$ denote the individual worth of player i. We assume that there exists $\varepsilon > 0$ such that the designer can impose on player i a final payoff $w_i - \varepsilon$ by assigning i to a singleton coalition and imposing a small fine. The use of such punishments allows us to specify strategy sets that are independent of the feasible sets. It needs to be stressed that this interpretation is based on the observability of utilities but not that the designer has full knowledge of the characteristic function.

We now turn to the construction of a mechanism $(g, (M_i))$. The first step is to define the strategy sets. Let $\mathcal{N}_i$ denote all coalitions containing i. The strategy set of player i is defined as

$$M_i = \{(x^i, S_i, y^i, \pi_i) \mid x^i \in \mathbf{R}^N_+, S_i \in \mathcal{N}_i, y^i \in \mathbf{R}^{S_i}_+, \pi_i \in \Pi\}.$$

A typical strategy profile will be denoted $m = (m_i)_{i \in N} = (x^i, S_i, y^i, \pi_i)_{i \in N}$.

To complete the definition of the mechanism we shall now define the outcome function g. To do this we first construct a mapping that assigns to each strategy profile a coalition structure. Given a strategy profile (x^i, S_i, y^i, π_i), player $p(\pi)_1$ is allowed to form coalition $S_{p(\pi)_1}$. Let $T_1 = S_{p(\pi)_1}$. Player $p(\pi)_1$ is called the leader of coalition T_1. If $T_1 = N$, no other coalisions are formed. Otherwise, let i_2 be the first player in the set $N \backslash T_1$. This player, the leader of the next coalition, gets to form the coalition T_2, consisting of all members of S_{i_2} who have not already been claimed by T_1, i.e. $T_2 = S_{i_2} \cap [N \backslash T_1]$. Notice that if i_2 exists, T_2 is non-empty. In a similar way we can define i_3 and T_3 etc. Thus, given a strategy profile m we have constructed a coalition structure

$$P(m) = (T_1(m), \ldots, T_{n_m}(m)).$$

Let $n(m) = p(\pi)_n$ denote the last player according to the order $p(\pi)$. For $T \in \mathcal{N}$, let $e^T = (1, \ldots, 1)$ denote the unit vector in $\mathbf{R}^T$ and let e_i^T denote the vector in $\mathbf{R}^T$ with 1 in the i-th coordinate and 0 elsewhere.

The outcome for a strategy profile $m = (x^i, S_i, y^i, \pi_i)$ is defined according to the following rules:

R.1. If $x^i \neq x^j$ for any $i, j \in N$, then

$$g(m) = w - \varepsilon e^N_{n(m)}.$$

As we have already pointed out, this is achieved by assigning all the players to singleton coalitions and imposing a fine on player $n(m)$.

R.2. If R.1 does not apply, then the payoffs are determined by the worth of each coalition in the coalition structure $P(m)$. The leaders are residual claimants in their respective coalition. They are required to manage their coalition so as to provide every member a payoff at least as high as the proposed x^*. If that is not feasible, the leaders are penalized. The formal construction of the outcome function is as follows. Let t be the leader of coalition $T \in P(m)$ and define

$$g(m, T) = \begin{cases} y^t & \text{if } y^t \in V(T) \text{ and } y^t \geq x^*_T, \\ w_T - \varepsilon e^T_t & \text{otherwise} \end{cases}$$

and let $g(m) = (g(m, T))_{T \in P(m)}$.

This completes the construction of the mechanism. In order to emphasize the informational requirements that have been assumed, we have allowed the individual strategies $m_i = (x^i, S_i, y^i, \pi_i)$ to be such that $x^i \in \mathbf{R}^N_+$ and $y^i \in \mathbf{R}^{S_i}_+$. Since the designer can verify whether or not $y^i \in V(S_i)$, it is possible for her to enforce $g(m)$. Notice also that the designer is only required to verify the feasibility of the leaders' announcements of y^t in the coalition structure $P(m)$. In particular, the designer does not learn the entire characsteristic function of the game.

Finally, observe that the mechanism is feasible in the sense that $g(m) \in V(P(m))$ for all $m \in M_N$.

We can now present the main result of this Section.

Theorem 2. *In the class of all games satisfying assumption* (A), *the mechanism* $(g, (M_i))$ *implements the coalition structure core in Nash equilibrium.*

Proof. It is easy to see that every allocation in the coalition structure core is an outcome of some Nash equilibrium. Suppose $x^* \in C(P, V)$ for some $P \in \mathscr{P}$. For every $S \in P$, let $m_i = (x^*, S, x^*_S, \pi^e)$ for all $i \in S$, where π^e denotes the identity permutation. Since $x^* \in V(P)$, it follows that $g(m) = x^*$. It is easy to see that since $x^* \in C(P, V)$, m is a Nash equilibrium – if any player had a profitable deviation, there would be a coalition with an objection to x^*.

Next, we show that every equilibrium corresponds to a core outcome. Suppose $m = (x^i, S_i, y^i, \pi_i)$ is a Nash equilibrium. We begin by showing that $g(m) \geq w$. Suppose not, i.e., there exists i such that $g(m)_i < w_i$. We claim that i can obtain a payoff of w_i. Consider a strategy $m'_i = (x'^i, S_i, y^i, \pi'_i)$, where $x'^i \neq x^j$ for some $j \neq i$ and π'_i is such that $p(\pi'_i, \pi_{-i})_1 = i$. With this strategy player i becomes the first player in the order $p(m'_i, m_{-i})$ and induces R.1. This yields player i a final payoff of w_i and contradicts the hypothesis that m is a Nash equilibrium. Thus, $g(m) \geq w$. This immediately implies that $x^i = x^*$ for all $i \in N$ and, therefore, R.2. applies. Moreover, since $g(m) \geq w$, it also follows from the construction of R.2 that

$$g(m) \geq x^*. \tag{1}$$

Of course,

$$g(m)_T \in V(T) \text{ for every } T \in P(m). \tag{2}$$

To complete the proof, given (2), it remains to be shown that there is no coalition that has an objection to $g(m)$. Suppose this is false, i.e.,

there exists a coalition S and $y \in V(S)$ such $y > g(m)_S$.

Let $i \in S$ be a player such that $y_i > g(m)_i$. Then player i by becoming the first player and managing coalition S can obtain $y_i > x_i^*$. More precisely, consider $m_i' = (x^*, S, y, \pi_i')$ such that $p(\pi_i', \pi_{-i})_1 = i$. Since $y > g(m)_S$, from (1) we know that $y > x_S^*$. Clearly then, from R.2 it follows that $g(m_i', m_{-i})_S = y$ and $g(m_i', m_{-i})_i > g(m)_i$, which contradicts the hypothesis that m is a Nash equilibrium and proves that $g(m) \in C(P, V)$. □

Remark 4. It is easy to modify the mechanism to implement the core of any given coalition structure P. To do this change R.1 as follows. If m is such that either $P(m) \neq P$ or $x^i \neq x^j$ for any $i, j \in N$, then $g(m) = w - \varepsilon e^N_{n(m)}$. The rest of the proof remains unchanged.

Remark 5. Is it possible to apply the kind of normal form mechanism used in this Section to the private ownership economies of Sect. 2 in which the designer does not know the utility functions? The answer seems to be no. The main difficulty is that when in response to a status-quo a player proposes a coalition and a feasible allocation for its members, the designer cannot determine if all members of this coalition will prefer such a proposal. The second stage of the mechanism in Sect. 2 deals precisely with this issue. If, as in this section, utilities are observable, the designer does not need responses from all members of a coalition to determine whether a proposal is truly an objection. For this reason, it does not seem that a simple normal form mechanism could be used in Theorem 1. But the arguments of this Section applied to an extensive form mechanism can provide a significant generalization of Theorem 1, as the following Remark points out.

Remark 6. Suppose the environment consists of private ownership economies in which the planner does not know preferences, endowments or the technology. By combining the mechanism of this section with that of the previous section it is possible to implement the core even in such environments. As in this section, assume that the planner can enforce a coalition structure, impose small penalties and observe the final allocation. It is possible to use the mechanism of Sect. 2 and follow the penalties used in this section. Specifically, the planner enforces the coalition structure of singletons and punishes the last player according to the ordering p if the status-quo is not unanimously agreed. If the outcome prescribed by the mechanism of Sect. 2 is the formation of coalition S and the allocation (y_S, ω_{-S}), coalition S (possibly the grand coalition) is required to show that it can allocate y_S, while the complement is forced into singletons. If coalition S cannot establish that y_S is feasible, then the coalition structure of singletons is enforced and the proposer is punished.

References

Abreu D, Sen A (1990) Subgame perfect implementation: A necessary and almost sufficient condition. J Econ Theory 50: 285–299

Bloch F (1992) Sequential formation of coalitions with fixed payoff division. Games econ behav (in press)

Chatterjee K, Dutta B, Ray D, Sengupta K (1993) A non-cooperative theory of coalitional bargaining. Rev Econ Studies 60: 463–477

Edgeworth FY (1881) Mathematical Psychics. Kegan Paul Publishers, London

Dutta B, Sen A, Vohra R (1995) Nash implementation through elementary mechanisms in economic environments. Econ Des 1: 173–203
Hurwicz L (1979) Outcome functions yielding Walrasian and Lindahl allocations at Nash equilibrium points. Rev Econ Studies 46: 217–225
Hurwicz L, Maskin E, Postlewaite A (1995) Feasible Nash implementation of social choice rules when the designer does not know endowments or production sets. In: Ledyard JO (ed) The Economics of Informational Decentralization: Complexity, Efficiency and Stability. Kluwer Academic Publishers, Amsterdam, pp 367–344
Jackson M (1992) Implementation in undominated strategies: a look at bounded mechanisms. Rev Econ Studies 59: 757–775
Kalai E, Postlewaite A, Roberts J (1979) A group incentive compatible mechanism yielding core allocations. J Econ Theory 20: 13–22
Lagunoff R (1994) A simple noncooperative core story. Games Econ Behav 7: 54–61
Maskin E (1977) Nash equilibrium and welfare optimality. mimeo, MIT
Maskin E (1985) The theory of implementation in Nash equilibrium: a survey. In: Hurwicz L, Schmeidler D, Sonnenschein H (eds) Social Goods and Social Organization. Cambridge University Press, Cambridge
Moldovanu B, Winter E (1995) Order independent equilibria. Games Econ Behav 9: 21–34
Moore J (1992) Implementation, contracts and renegotiation in environments with complete information. In: Laffont JJ (ed) Advances in Economic theory. VI World Congress of the Econometric Society. Cambridge University Press, Cambridge, (vol. I)
Moore J, Repullo R (1988) Subgame perfect implementation. Econometrica 56: 1191–1220
Perry M, Reny PJ (1994) A non-cooperative view of coalition formation and the core. Econometrica 62: 795–817
Postlewaite A, Wettstein D (1989) Feasible and continuous implementation. Rev Econ Studies 56: 603–611
Ray D, Vohra R (1995) Binding agreements and coalitional bargaining. J Econ Theory (in press)
Scarf H (1967) The core of an *N* person game. Econometrica 38: 50–69
Scarf H (1986) Notes on the core of a productive economy. In: Hildenbrand W, Mas-Colell A (eds) Contributions to Mathematical Economics. Essays in Honor of Gerard Debreu. North-Holland, Amsterdam pp 401–429
Schmeidler D (1980) Walrasian analysis via strategic outcome functions. Econometrica 48: 1585–1594
Selten R (1981) A non-cooperative model of characteristic function bargaining. In: Aumann R et al. (eds) Essays in Game Theory and Mathematical Economics in Honor of Oskar Morgenstern. Bibliographisches Institut Mannheim
Serrano R (1995) A market to implement the core. J Econ Theory 67: 285–294
Thomson W (1992) Divide and permute and the implementation of solutions to the problem of fair division (revised 1993). Rochester Center for Economic Research Working Paper No. 360.
Wilson R (1978) Competitive exchange model. Econometrica 46: 577–585
Winter E (1992) Mechanism robustness in multilateral bargaining. Mimeo, Hebrew University of Jerusalem
Winter E (1993) Bargaining in committees. Mimeo, Hebrew University of Jerusalem