Econ. Theory 7, 531–539 (1996)

Economic Theory

The production and cost-sharing of an excludable public good*

Benny Moldovanu

Department of Economics, University of Bonn, Adenauerallee 24–42, D-53113 Bonn, GERMANY

Received: December 7, 1993; revised version May 31, 1995

Summary. We study a model of negotiation and coalition formation concerning a public expenditure and its financing. The agents must determine which coalition will jointly produce a public good, how much will be produced, and how the cost is to be shared. Agents that do not belong to the final coalition are excluded from consumption of the public good. Subgame-perfect Nash equilibria in stationary strategies lead to the formation of the grand coalition with an agreed alternative in the core of the economy. Conversely, for each alternative in the core, there exists a subgame-perfect Nash equilibrium in (pure) stationary strategies that leads to the formation of the grand coalition with that alternative.

1. Introduction

We study the coalition formation process in a collective decision problem concerning a public expenditure and its financing. The public expenditure relates to the joint production of a good that has two main characteristics:

1. Non-rivalry in consumption.
2. Exclusion is possible.

For example, cable TV and many other club goods (without congestion) possess these properties. The decision of the community must answer the following questions:

1. Who will contribute and be allowed to consume the public good?
2. What is the suitable level of production?
3. How the costs are to be allocated between the consumers?

One cannot address these questions separately. The size of the coalition that will produce the good and its members' identities determine the available resources. The

* I wish to thank Dieter Balkenborg, Georg Noldeke, Shlomo Weber and Eyal Winter for helpful discussions. The comments of an anonymous referee greatly improved the quality of the exposition. Financial support from Deutsche Forschungsgemeinschaft, SFB 303, is gratefully acknowledged.

resources determine the production possibilities. The production cost must be covered by allocating the cost between the members. Finally, the level of supplied goods and the cost-sharing rule form the basis on which agents decide to cooperate in production and consume, or not. Insisting on Pareto-optimality is not enough: there are usually many ways in which the three above mentioned issues can be simultaneously resolved. The selection of one alternative between the many possible ones, if undertaken by the agents themselves, has the character of a complex negotiation process.

The core expresses ideas of group rationality: A chosen alternative must be stable in the sense that no coalition can produce by its own (and cover the costs) while increasing the utility of its members. The core approach (see, among others, Pauly, 1967 and Foley, 1970) adds valuable insight to the theory of public goods and clubs, and its fundamental importance will be illustrated here as well. However, this approach neglects the institutional aspect and the strategic behavior of agents that wish to influence the choice of the final outcome.

It is not at all clear that strategic behavior based on individual utility maximization is consistent with the group-rational outcomes described by the core. For example, in Moldovanu (1992) we show that the subgame-perfect Nash equilibria of related negotiation games do not necessarily lead to core outcomes in exchange economies with private goods. Core equivalence is obtained only by using coalition proof Nash equilibria.

Here, we study the following situation: A public good can be produced with one input, say "money". The technology is represented by a given cost function. An agent is characterized by an initial endowment in money and an utility function that depends on the amounts of public good and money consumed. Informally, agents make proposals that consists of a coalition, a level of public good and a cost allocation to the members of that coalition. Proposed coalitions must belong to an a-priori determined set of "winning" coalitions. Thus, we assume that the "legislative" power is a-priori determined, independently of the result of the present negotiations. A feasible proposal must ensure that the cost of production can be covered. A coalition forms if the proposal is unanimously accepted by its members. A new proposal can be brought to debate by an agent that refuses a standing proposal. The game ends after a winning coalition has formed. An agent that is not a member of the final coalition is excluded from the consumption of the public good.

Subgame-perfect Nash equilibria (SPNE) in stationary strategies lead to the formation of the grand coalition with an agreed alternative in the core of the economy. Conversely, we show that for each alternative in the core there exists a SPNE in (pure) stationary strategies that leads to the formation of the grand coalition with that alternative. Thus, the decentralized process of negotiation leads to efficient, group rational outcomes, and exclusion does not occur.

Bagnoli and Lipman (1989) show that some refinements of the Nash equilibria lead to core allocations in a model of public good provision (without exclusion). Some of their main assumptions are that the utility functions are linear and that the public good comes in discrete units. Admati and Perry (1991) show how efficiency can be achieved in a two-person sequential game of public good provision without commitments. Bergstrom et al. (1986) study the effect of wealth redistribution in a model of private provision of public goods.

Harsanyi (1974), Selten (1981), Chatterjee et al. (1993), Perry and Reny (1994), Winter (1994) and Hart and Mas-Colell (1992) study negotiation procedures based on underlying games in coalitional form with transferable utility. In all these papers utility functions are (implicitly) assumed to be separable, additive and linear in money. Income effects cannot be taken into account. We allow here rather general utility functions so that income effects can be considered. In Moldovanu and Winter (1995) we show that, for non-transferable utility (NTU) pure exchange markets, the set of core outcomes coincides only with the set of payoffs attainable in *order independent equilibria* of a game where several coalitions may form.

The paper is organized as follows. In Section 2 we describe the public good economy. In Section 3 we describe the negotiation game. In Section 4 we prove the main result that relates outcomes in strategic equilibria of the negotiation game to outcomes in the core of the public good economy.

2. The economic model

Consider a society of n agents, $N = \{1, 2, \ldots, n\}$. A coalition S is a non-empty subset of N. The set $\mathscr{W}$ is an a-priori determined, non-empty family of *winning* coalitions. We require that $\mathscr{W}$ is *monotonic*, i.e. $-S \in \mathscr{W}$ and $T \supset S$ imply $T \in \mathscr{W}$. Because we study a model where only one coalition can form it makes sense to require also that $S \in \mathscr{W}$ implies that $N \backslash S \notin \mathscr{W}$.

Let $g(\alpha)$ denote the cost, in terms of money, of producing the amount (or level) α of a public good. We assume that g is continuous, strictly increasing, with $g(0) = 0$, Each agent i has an initial endowment in money $w^i > 0$. If agent i consumes amount α of public good and if i's share of the cost is c^i, then this agent has utility level $U_i(\alpha, w^i - c^i)$. We assume that U_i is continuous and strictly increasing in the amount of public good and money. We normalize the utility functions such that $U_i(0, w^i) = 0$. The results are not dependent on this normalization. If a winning coalition forms, it controls the technology g, hence it can produce any level of public such that it can afford the cost. For a vector of costs c^S denote by $c(S)$ the sum$\sum_{i \in S} c^i$. An alternative (α, c^S) is *feasible* for a winning coalition S if $g(\alpha) \leq c(S)$, and if $0 \leq c^i \leq w^i$ for all $i \in S$. The requirement $c^i \leq w^i$ means that the agents are not allowed to borrow. Agents not included in the final coalition can be excluded, without costs, from the consumption of the public good, and hence their utility is zero (see the normalization above).

Let (α, c^N) be a feasible alternative for the grand coalition N (of course, N is winning). This alternative can be *improved upon* if there exist a winning coalition S, and an alternative (β, d^S) feasible for S such that $U_i(\beta, d^i) \geq U_i(\alpha, c^i)$ for all $i \in S$, with at least one strict inequality. The *core* is the set of all alternatives that cannot be improved upon. It is well known that the core is not empty in this situation (see Champsaur, 1975, Demange, 1987). Heuristically, the intuition is that, because of increasing marginal contributions to coalition scale, cooperation becomes more and more profitable as additional agents join in.

3. A negotiation scenario

We study the following simple model of negotiation: A fixed agent, say agent 1, has the first initiative. He may pass the initiative to another player, or he may make a proposal. A proposal consists of a winning coalition S and an alternative (α, c^S) that is feasible for S. The initiator designates also a responder to the proposal, a player of S. (Alternatively one can assume that a proposal is considered by responders in a fixed, predetermined order.) A responder can accept or refuse the proposal. If he refuses, then this responder has the initiative, and he may make another proposal. This may be a different feasible alternative for coalition S, or an alternative for another coalition. If the responder accepts, then there are two possibilities: If the responder was the last player in S to propose or accept the proposal, then the coalition S forms and the game ends. Otherwise, the responder selects the next responder to the existing proposal, and the game continues in the same fashion. If a coalition S forms at an end node of the game and its members have all agreed on an alternative (α, c^S), then the payoff to $i \in S$ is given by $U_i(\alpha, w^i - c^i)$ and the payoff of the excluded agents (i.e. not in S) is given by $U_i(0, w^i) = 0$. If there is perpetual disagreement, then no amount of public good can be provided and the payoff to all players is zero.

We concentrate here on simple, stationary strategies. A stationary behavior strategy for a player in this game assigns to each decision node of that player a probability distribution with finite support over the set of actions available at that node. A player uses the same mixture of possible actions whenever he acts as proposer. The action of a responder depends only on the existing proposal and on the set of players that have proposed or accepted this proposal. Stationary strategies constitute here the simplest patterns of behavior consistent with fully rational behavior. Without this assumption one obtains results of the Folk Theorem type. (see also Chatterjee et al., 1990).

Subgame perfect Nash equilibria (SPNE) are defined in the usual way – optimization is required at *all* decision nodes. To remove some boundary phenomena in the proof of our main Theorem we need a tie-breaking rule:

Assumption: Let (α, c^S) and (β, d^T) be feasible alternatives for winning coalitions S and T, respectively. Assume that $S \subset T$ and that for all $i \in S$ it holds that $U_i(\alpha, w^i - c^i) = U_i(\beta. w^i - d^i)$. Then:

1) If $j \in S$ is indifferent between proposing coalition S with (α, c^S) or coalition T with (β, d^T) then she proposes coalition T with (β, d^T).
2) If $j \in S$ is indifferent between accepting or rejecting the proposal T with alternative (β, d^T) then she accepts.

4. Strategic equilibria and the core

We presented a model where agents decide in a decentralized way which coalition will produce, finance, and consume the public good. Forming a coalition that excludes some agents is an explicit strategic possibility, and nothing insures that the agreed alternative will be in the core, even if the grand coalition forms. Strategic equilibria are immune against individual deviations, but not necessarily against the coalitional "deviations" that are behind the core idea.

Given non-rivalry and the possibility of exclusion, we now show that the public good model is remarkable, in the sense that it offers strong incentives to cooperate, even if the agents behave strategically.

Theorem: In any SPNE in stationary strategies the grand coalition forms with an agreed alternative in the core. Conversely, for every alternative in the core there exists a SPNE in pure, stationary strategies that yields the formation of the grand coalition with that alternative.

The proof of the Theorem is based on the following two Lemmata:

Lemma 1: Let σ be a SPNE in stationary strategies. There exists a vector of utilities $q = q^N(\sigma)$ such that if, at an end node possible under σ, coalition S forms with the agreed alternative (β, d^S), then it must hold that $U_i(\beta, w^i - d^i) = q^i$ for all $i \in S$. The vector q has the following properties:

Property 1: For any player $i \in N$ there exist a winning coalition S and a feasible alternative for S, (α, c^S), such that $i \in S$ and $U_j(\alpha, w^j - c^j) = q^j$ for all $j \in S$.

Property 2: For any winning coalition S, and for any feasible alternative (γ, e^S), it can not be the case that $U_i(\gamma, w^i - e^i) \geq q^i$ for all $i \in S$, with at least one strict inequality.

Proof: Define q^i to be the expected utility level of player i given that σ is played and i has the initiative. This is well defined because the stationarity of the profile σ. Let q be the vector of the above defined expectations for all players in N.

We start with Property 2: Let S be a winning coalition, and let (γ, e^S) be a alternative feasible for S such that $U_i(\gamma, w^i - e^i) \geq q^i$ for all $i \in S$, with at least one strict inequality. Let $j \in S$ such that $U_j(\gamma, w_j - e^j) > q^j$. By continuity and monotonicity of the utility functions we can find $\varepsilon > 0$ such that it also holds $U_j(\gamma, w^j - (e^j + \varepsilon)) > q^j$. Let (γ, f^S) be the alternative where $f^j = e^j + \varepsilon$ and $f^i = e^i - \varepsilon/(|S| - 1)$ for $i \in S \backslash \{j\}$. Then it is clear that (γ, f^S) is feasible for the coalition S, and that, for all $i \in S$, it holds $U_i(\gamma, w^i - f^i) > q^i$.

Assume now that j proposes coalition S with alternative (γ, f^S). Assume that $k \in S$ is the last player in S that has not yet responded to this proposal (i.e. – all other players in S have already accepted). If k refuses, then he has the initiative, and by definition, his expected payoff is q^k. Then it is clearly optimal for k to accept the proposal. By backward induction all players in S will accept as well. The proposal S with alternative (γ, f^S) could not have been part of j's action as a proposer in σ, because, by definition, his action there would yield only q^j. Hence the deviation of j is beneficial, a contradiction to the assumption that σ is a subgame perfect equilibrium.

Assume now that at an end node possible under σ a coalition S forms with alternative (β, d^S). Note that in equilibrium player i will never accept a proposal which, if accepted by all other responders, will yield him less than q^i. This is so because i can refuse, become initiator, hence securing q^i. Also, it is not optimal for a proposer i to propose an alternative that, if accepted by all responders, yields less than q^i. These remarks yield $U_i(\beta, w^j - d^i) \geq q^i$ for all $i \in S$. By Property 2 we obtain $U_i(\beta, w^j - d^i) = q^i$ for all $i \in S$. This is the wished result in the first part of the Lemma.

We conclude with Property 1. If $q = 0^N$ then Property 1 holds trivially because the alternative $(0, 0^N)$ is clearly feasible for N. Assume than that q has some positive coordinates, and let $i \in N$. If $q^i > 0$, then, by the definition of the expectations, agent i must be a member of any coalition that forms under σ in any subgame that starts with player i as initiator, and the result follows from the first part of the Lemma. Now let i be such that $q^i = 0$, and assume that the condition in Property 1 does not hold for this player. Consider a winning coalition S and a feasible alternative for S, (α, c^S), such that $U_j(\alpha, w^j - c^j) = q^j$ for all $j \in S$, and such that α is positive. Such a coalition exists because q has some positive coordinates. By assumption $i \notin S$. By mononicity of the family of winning coalitions we obtain that $S \cup \{i\}$ is also winning. The alternative $(\alpha, (c^S, 0^i))$ is feasible for $S \cup \{i\}$, and we obtain that $U_i(\alpha, w^i) > U_i(0, w^i) = 0 = q^i$. This is a contradiction to Property 2. **Q.E.D.**

It is not al all clear that there exists a feasible alternative that generates levels of utility as in q. Different players may form expectations based on different coalitions, level of productions, etc... The next key Lemma shows that the expectations can be, in a sense, well harmonized. We strongly employ here the special structure of the public good economy.

Lemma 2: Let q be any vector with Properties 1 and 2. Then there exists an alternative (α, c^N) in the core such that $U_i(\alpha, w^i - c^i) = q^i$ for all $i \in N$.

Proof: Consider the family $\mathscr{B}$ of all winning coalitions S for which we can find a feasible alternative (α, c^S) such that $U_j(\alpha, w^j - c^j) = q^j$ for all $j \in S$. By Property 1 this family is not empty. Let T be inclusion maximal in $\mathscr{B}$, i.e. – if $R \supsetneqq T$ then $R \notin \mathscr{B}$. Let (α, c^T) be the feasible alternative for T with the property that $U_j(\alpha, w^j - c^j) = q^j$ for all $j \in T$. Because of feasibility and Property 2 it must hold that $g(\alpha) = c(T)$. If $T = N$ then the result is clear. Otherwise, let $i \in N \backslash T$. By Property 1 there exists a coalition S such the $i \in S$ and $S \in \mathscr{B}$. Let (β, d^S) be the feasible alternative for S with the property that $U_j(\beta, w^j - d^j) = q^j$ for all $j \in S$. As before, it holds that $g(\beta) = d(S)$. Assume without loss of generality that $\alpha \leq \beta$. We obtain:

(1) $g(\alpha) + g(\beta) = c(T) + d(S) = c(S \cap T) + c(T \backslash S) + d(S)$.

If $g(\beta) \leq c(T \backslash S) + d(S)$ then the alternative $(\beta, (c^{T/S}, d^S))$ is feasible for $S \cup T$. Because $\alpha \leq \beta$ we have

(2) $U_j(\beta, w^j - c^j) \geq U_j(\alpha, w^j - c^j) = q^j$ for $j \in T \backslash S$.

By definition we have also

(3) $U_j(\beta, w^j - d^j) = q^j$ for $j \in S$.

If $\alpha = \beta$ then there are only equalities in (2). The coalition $S \cup T$ is winning and (2), (3) imply together that $S \cup T \in \mathscr{B}$. This is a contradiction to the maximality of T. If $\alpha < \beta$ then there are strict inequalities in (2), and this is a contradiction to Property 2. Hence it must hold that

(4) $g(\beta) > c(T \backslash S) + d(S)$.

By (1) we obtain that

(5) $g(\alpha) < c(S \cap T)$.

This is a contradiction to $g(\alpha) = c(T)$. The assumption that N does not belong to $\mathscr{B}$ leads always to contradiction, and the result is proved. **Q.E.D.**

Proof of the Theorem: Let q be the associated expectations vector of the equilibrium σ. Denote by $\mathscr{U}$ the set of payoff vectors generated by alternatives in the core. By Lemma 1 and 2 we obtain that $q \in \mathscr{U}$.

If q is in the relative interior of this set then there is no coalition S, $S \neq N$, and feasible alternative (α, c^S) such that $\forall i \in S$, $U_i(\alpha, c^i - w^i) = q^i$. Hence at any possible end node under σ only the grand coalition can form. If the grand coalition forms with an agreed alternative (β, d^N) then, by the first part of Lemma 1, it must hold that $U_i(\beta, w^i - d^i) = q^i$ for all $i \in N$. By Property 2 in Lemma 1, the alternative (β, d^N) cannot be improved upon and therefore it is in the core.

If q is on the boundary of $\mathscr{U}$, then let (β, d^N) be an alternative in the core such that $U_i(\beta, w^i - d^i) = q^i$ for all $i \in N$. By Lemma 1, if a coalition S forms with agreed alternative (α, c^S), it must hold that $U_i(\alpha, w^i - c^i) = q^i = U_i(\beta, w^i - d^i)$ for all $i \in S$. Assume that it is optimal for $j \in S$ (given the strategies of other players) to propose (α, c^S). By backward induction and by the definition of the expectations vector q, it is also optimal for j to propose (β, d^N). (This proposal will be accepted by the tie-breaking rule). Hence, player j is indifferent between proposing (α, c^S) or (β, d^N). By the tie-breaking rule the alternative (β, d^N) will be proposed in equilibrium. This concludes the proof of the first part.

For the converse part, let (α, c^N) be in the core, and consider the following strategy profile τ: Each player proposes the coalition N with alternative (α, c^N) and a responder. A responder accepts a proposal only if all responder j who have not yet accepted, including himself, obtain at least the utility level $U_j(\alpha, w^j - c^j)$. It is easy to see that τ is an SPNE: If a player i proposes a coalition S with an alternative such that his utility level is higher than $U_i(\alpha, w^i - c^i)$ then, because (α, c^N) is in the core, there exists a player $k \in S$ that obtains less than $U_k(\alpha, w^k - c^k)$. This player will reject the proposal. Obviously it is optimal for a responder j to accept a proposal that yields him at least $U_j(\alpha, w^j - c^j)$. Also, it is optimal to reject a proposal that yields less than $U_k(\alpha, w^k - c^k)$ for a responder k that has not yet accepted, because this player will reject this proposal anyway. **Q.E.D.**

Note that without the tie-breaking rule, a sub-coalition may form in equilibrium. This may happen if and only if the vector of expectations associated to that equilibrium happens to be on the boundary of the set of core payoffs. However, the chosen alternative will be in the core of the sub-economy and the members of the forming coalition are indifferent between the equilibrium outcome and the formation of the grand coalition with an alternative in the core.

5. Concluding remarks

As in any strategic model, the negotiation scheme is special because it must exactly detail the institutional setting. Nevertheless, it seems that models that present increasing returns to coalition scale are robust in the sense that several bargaining procedure may lead to group-rational outcomes. This is a major message to be learnt from Harsanyi (1974), Chatterjee et al. (1993) and Winter (1994) where negotiation games based on *convex* TU games are studied. Note that a version of our model where utility functions are linear, additive and separable in money yields a strictly convex TU game. Moreover, if we construct a NTU game based on our public good economy with general utility functions, we obtain an *ordinal convex* NTU game.

We conclude with some comments on the robustness of our results with respect to the introduction of time-discounting.

For general TU games, there are two sources of inefficiencies in the model of Chatterjee et al. The first source – delayed agreement – is connected to the fact that the exact order of proposers (or *protocol*) is very important in models where several coalitions can form (see also Perry and Reny, 1994, Moldovanu and Winter, 1995). Inefficiencies occur for some protocols because agent i, by delaying an agreement that includes her, may cause the early formation of a coalition S, $i \notin S$; in the new situation that ensues, the given protocol may be favorable to i, yielding a higher payoff overall. In our context, this kind of behavior is potentially harmful, and never advantageous (when restricting attention to stationary strategies): if a coalition S, $i \notin S$, forms, the game ends, and i is excluded from consumption. Note that, since our model allows for the formation of a unique coalition, a protocol is here just a designation of the first proposer. Moreover, the exact order of the responders is irrelevant.

The second source of inefficiency is the fact that coalitions other than the grand one may form in equilibrium even in superadditive games with non-empty cores. This kind of inefficiency does not occur in our model (for discount factors close enough to 1) because of the convexity properties, as discussed above. We can conclude that for all discount factors δ close enough to 1, all stationary equilibria are efficient and there is no delay. Moreover, the limit outcomes, as $\delta \to 1$, are core allocations. This result is in line (and can be similarly proven) with the results of Chatterjee et al. (1993) for strictly convex TU games (see their Propositions 7, 8).

The converse part of our Theorem, namely that all core allocations can be attained as SPNE payoffs, is not true for δ strictly less than 1 (the correspondence is not lower hemi-continuous). This is a common phenomenon that can be also observed in simpler models (e.g., Rubimstein's two-person bargaining model).

References

Admati, A. R., Perry, M.: Joint projects without commitments. The Review of Economic Studies **58**, 259–276 (1991)

Bergstrom, T., Blume, L., Varian, H.: On the private provision of public goods. Journal of Public Economics **29**, 25–49 (1986)

Bagnoli, M., Lipman, B.: Provision of public goods: Fully implementing the core through private contributions. Review Of Economic Studies **56**, 583–601 (1989)

Champsaur, P.: How to share the cost of a public good? International Journal of Game Theory **4**, 113–129 (1975)

Chatterjee, K., Dutta, B., Ray, D., Sengupta, K.: A noncooperative theory of coalitional bargaining. Review of Economic Studies **60**, 463–477 (1993)

Demange, G.: Nonmanipulable cores. Econometrica **55**, 1075–1108 (1987)

Foley, D.: Lindahl's solution and the core of an economy with public goods. Econometrica **38**, 66–72 (1970)

Harsanyi, J. C.: An equilibrium point interpretation of stable sets and an alternative definition. Management Science **20**, 1422–1495

Hart, S., Mas-MColell, A.: N-person noncooperative bargaining. Discussion paper no. 10, Center for Rationality, Hebrew University, Jerusalem, 1992

Moldovanu, B.: Coalition-proof Nash equilibria and the core in three-player games. Games and Economic Behavior **4**, 565–581 (1992)

Moldovanu, B., Winter, E.: Order independent equilibria. Games and Economic Behavior **9**, 21–34 (1995)

Pauly, M. V.: Clubs, commonality, and the core: an integration of game theory and the theory of public goods. Economica **34**, 314–324 (1967)

Perry, M., Reny, P. J.: A noncooperative view of coalition formation and the core. Econometrica **62**, 795–819 (1994)

Selten, R.: A non-cooperative model of characteristic function bargaining. In: V. Böhm, H. Nachtkampf (eds.) Essays in game theory and mathematical economics in honor of O. Morgenstern. Mannheim: Bibliographisches Institut 1981

Winter, E.: Demand commitment bargaining and snowballing cooperation. Economic Theory **4**, 255–275 (1994)