数理解析研究所講究録
861巻 1994年 67-76

Self-Binding Strategies for a Coalition

法政大学経済学部　中山幹夫　(Mikio Nakayama)

Introduction

In the tradition of game theory, it is a common understanding that a cooperative game is a game in which players are able to make a binding agreement; that is, an agreement that is enforceable. For instance, Nash (1953) argued that an external mechanism or "a sort of umpire" is necessary for the enforceability of contracts and commitments. Aumann (1973) also expresses a similar viewpoint.
If a cooperative game is given in strategic form, a coalition of players is then able to discuss and coordinate its choice of strategies in the enforceable way, and, if it wishes to do so, can obtain a certain utility vector independently of the strategies of all others. But, does the game necessarily become noncooperative if the externally enforceable agreements are not available? Is the coalitional behavior described above, in particular, meaningless when the enforceability is assumed away?

The answer will be affirmative if the enforceability is to be replaced with the notion of self-enforcingness as established in the concept of Nash equilibria and coalition-proof Nash equilibria due to Bernheim, Peleg and Whinston (1987). But this will not be the

only answer: the possibility of a coalitional behavior remains if the players are able to make an agreement by themselves without resort to the external mechanism. To be more specific, we can show that a coalition may choose a joint strategy by which it can sustain itself independently of the strategies of all other players. We shall call such a strategy a self-binding strategy for a coalition.

The self-bindingness is an extension of the notion of credibility due to Ray (1989) to games in strategic form. By explicitly considering the strategic interactions, we can forward the analysis a step further. Under the assumption of Scarf (1971) that utility functions are quasiconcave, we will state a sufficient condition for a given coalition to have a self-binding strategy; and then, point out that a market game can be derived from a strategic game in which every coalition has a self-binding strategy. Thus we can conclude that a market game is a cooperative game that in fact does not require the assumption of binding agreements.

The Self-Binding Strategy

Let $G=(N, \{X^i\}_{i\in N}, \{u_i\}_{i\in N})$ be a game in strategic form, where N is a finite set of players, X^i is a nonempty, compact convex set of strategies of player i and u_i is a continuous utility function of player i. For each nonempty $S\subset N$, X^S denotes the Cartesian product of X^i in S, and let $X:=X^N$. Let S be a nonempty proper subset of N. The players in S can communicate with each other

and make an agreement on their choice of strategies, but no authority or an external mechanism is available to make an agreement binding. If coalition S is to form, the members of S must therefore seek to find an agreement that can be made in a self-binding way. By a self-binding strategy, we mean the strategy for S satisfying two requirements. The first is an obvious one that S be not disrupted thereby. Secondly, since there exists no obvious limitation on the strategies taken by players outside S, the first requirement should be met independently of the strategies taken by the complementary coalition. Thus, the self-binding strategy for S is one that can sustain itself for all strategies outside S.

To state the definition formally, we first need a notion of deviation. Let $T \subset S \subset N$. Then, given $z^S \in X^S$ and $x^T \in X^T$, we denote by $z^S \mid x^T$ the strategy $|S|$-tuple in which z^T is replaced with x^T. [1]

Definition 1. For all $T \subset N$, we say that T has a **self-binding α-deviation** at $x \in X$ if and only if there exists $y^T \in X^T$ such that for all $z \in X$,

(i) $u_i(z \mid y^T) > u_i(x)$ for all $i \in T$,

and

(ii) there does not exist $R \subset T$ $(R \neq T)$ which has a **self-binding α-deviation** at $z \mid y^T$.

---

[1] This notation is intended to keep the consistency of the following definitions 1 and 2 when we take T=N and S=N, respectively.

Note the recursion in the definition. Every single player $i \in T$ has a self-binding α-deviation at x iff the maximin value exceeds $u_i(x)$; and then, the definition goes on inductively by the cardinality of the subsets. Note also that every rebellious subset R of T must confront the same strategic environment as that of T, i.e., R must take all the strategies of N-T into consideration.

Definition 2. For all $S \subset N$, we say that $x^S \in X^S$ is a self-binding strategy for S if and only if for all $z \in X$, no $T \subset S$ has a self-binding α-deviation at $z \mid x^S$.

By a self-binding strategy for S, the members of S can assure by themselves a certain level of utilities that is enough to bind themselves in S whatever strategies the complementary coalition may choose. In characteristic function form, Ray (1989) defined a credible coalition to be one that can sustain itself by assuring each of the members a certain level of utility. Thus, the self-bindingness is a formalization of the credibility in strategic form. The following lemma is therefore an exact analog to the Ray's result. We say that a nonempty subset $T \subset N$ has an α-deviation at x if Definition 1 without (ii) is met. Then:

Lemma 1 . Let $x \in X$, and $T \subset N$. (i) If T has a self-binding α-deviation at x, then T has an α-deviation at x. (ii) If T has an α-deviation at x, then some $R \subset T$ has a self-binding α-deviation at x.

Proof. It will be enough to check (ii). Suppose T has an α-deviation at x. Then, there exists $y^T \in X^T$ such that for all $z \in X$,

$u_i(z \mid y^T) > u_i(x)$ for all $i \in T$. If, for any such $z \mid y^T$, there exists no $R \subset T$ $(R \neq T)$ which has a self-binding α-deviation at $z \mid y^T$, then it follows that T has a self-binding α-deviation at x. If, for some $z \mid y^T$, there exists $R \subset T$ $(R \neq T)$ which has a self-binding α-deviation at $z \mid y^T$, then there exists $w^R \in X^R$ such that for all $w^{N-R} \in X^{N-R}$,

$$u_i(w^R, w^{N-R}) > u_i(z \mid y^T) > u_i(x) \quad \text{for all } i \in R,$$

which implies that R has a self-binding α-deviation at x.

The α-core of the game G is the set of those strategies $x \in X$ at which no $S \subset N$ has an α-deviation. Then, Lemma 1 implies that x is in the α-core if and only if there exists no $S \subset N$ which has a self-binding α-deviation at x. The following result shows a general relation between the self-binding strategy and the familiar solution concept, the α-core.

Proposition 2. (i) Let $x \in X$. Then, x is a self-binding strategy for N if and only if x is in the α-core. (ii) If the α-core is empty, then some coalition $S \subset N$ $(S \neq N)$ has a self-binding strategy.

Proof. (i) Immediate from Lemma 1 and the definitions. (ii) Let $x \in X$. Then some $S \subset N$ has a self-binding α-deviation y^S at x by Lemma 1(ii). Since for all $z \in X$, no $T \subset S$ $(T \neq S)$ has a self-binding α-deviation at $z \mid y^S$, y^S will be a self-binding strategy for S if S itself does not have a self-binding α-deviation at $z \mid y^S$. Let w^S be any α-deviation at x satisfying for all $z \in X$ that

$$u_i(z \mid w^S) \geq u_i(z \mid y^S) > u_i(x), \quad \text{for all } i \in S.$$

Then, no $T \subset S$ $(T \neq S)$ must have a self-binding α-deviation at $z \mid w^S$,

since y^S is a self-binding α-deviation. Hence w^S must be a self-binding α-deviation at x. By compactness and continuity, there can be found a maximal w^S satisfying the above inequality, so that we may take one as y^S. Hence S has a self-binding strategy, and $S \neq N$ by (i).

Thus, the self-bindingness for N is equivalent to the concept of α-core; and for any coalition S, either S itself has a self-binding strategy, or its subcoalition has a self-binding strategy.

We now consider when a given coalition has a self-binding strategy. We shall state a sufficient condition for the class of games given by Scarf (1971). Let S be a nonempty proper subset of N. Then, we say that N-S has a damaging strategy $d^{N-S} \in X^{N-S}$ to S if and only if for all $z^S \in X^S$ and $z^{N-S} \in X^{N-S}$,

$$u_i(z^S, z^{N-S}) \geq u_i(z^S, d^{N-S}) \quad \textit{for all } i \in S.$$

Proposition 3. Assume that for all $i \in N$, u_i is quasi-concave in $x \in X$. Then, S has a self-binding strategy if N-S has a damaging strategy to S.

Proof. Let d^{N-S} be the damaging strategy. Then, since $u_i(\cdot, d^{N-S})$ is quasi-concave for all $i \in S$, it follows from Proposition 2(i) and the Scarf's theorem (1971) that there exists a self-binding strategy $x^S \in X^S$ for S in the subgame induced by holding x^{N-S} fixed to d^{N-S}. Then, for any $T \subset S$ and any $y^T \in X^T$, there must exist $z \in X^S$ such that $u_i(z \mid y^T, d^{N-S}) \leq u_i(x^S, d^{N-S})$ for some $i \in T$. Hence, there exists $w \in X$ such that $u_i(w \mid y^T) \leq u_i(x^S, d^{N-S})$ for some $i \in T$. Since d^{N-S} is a damaging

strategy, it follows that for all $x^{N-S} \in X^{N-S}$,

$$u_i(w \mid y^T) \leq u_i(x^S, d^{N-S}) \leq u_i(x^S, x^{N-S}) \quad \text{for some } i \in T,$$

which implies that no $T \subset S$ has an α-deviation at (x^S, x^{N-S}). Hence, for all x^{N-S}, there exists no $T \subset S$ which has a self-binding α-deviation at (x^S, x^{N-S}) by Lemma 1(i), so that x^S is a self-binding strategy for S.

Corollary 4. Under the quasi-concavity of all u_i, every nonempty coalition S has a self-binding strategy if every nonempty N-S has a damaging strategy.

The strategy that hurts uniformly the members of the complementary coalition appeals to intuition, but may not exist in general. If N-S has the damaging strategy, then it is easy to see that what N-S cannot prevent S from getting is precisely those payoff vectors which S can assure by itself. In the language of cooperative game theory, this shows that S is α-effective if and only if S is β-effective, whereas the "if" part is not true in general (see Scarf (1971)). Thus, the existence of a damaging strategy will be limited to special cases.

Nevertheless, there is a typical economic example for Corollary 4. To see this, consider the following pure exchange game G in strategic form (see Scarf (1971), and also Mas-Colell (1987)): For each $i \in N$, let $w^i \in \mathbb{R}^m_{++}$ be an m-vector of initial endowments, and let the strategy be any n vectors describing allocations of player i's endowments among the n players; that is, the strategy set X^i is defined as

$$X^i=\{x^i=(x^{i1},\ldots,x^{in}) \ : \ \sum_{j\in N} x^{ij}\leq w^i \ , \ \text{and } x^{ij}\in \mathbb{R}^m_+ \ \ \forall j\in N\}.$$

The utility function u_i is given by

$$u_i(x) \ = \ f_i\Big(\sum_{j\in N} x^{ji} \Big),$$

where f_i is continuous, quasiconcave in x and monotone nondecreasing in $\Sigma_{j\in N}x^{ji}$. In this game, every N-S may naturally allocate the endowments only among the members of N-S; namely, N-S always has a strategy x^{N-S} satisfying

$$x^{ji} = 0 \in \mathbb{R}^m_+ \ \ \text{for all } j\in N-S \text{ and } i\in S.$$

By the monotonicity of u_i, this strategy x^{N-S} can be easily identified with a damaging strategy to S. Such a damaging strategy makes good sense in the context of pure exchange. Thus, every nonempty coalition should have a self-binding strategy. The public good game mentioned in Mas-Colell (1987) is also an example, in which no contribution by N-S to financing a public good is a natural damaging strategy to S.

Given a pure exchange game G, one can obtain, for each coalition S, a set of utility vectors that S can assure by itself:

$$V(S) \ = \ \{v_S=(v_i)_{i\in S} : \exists x^S\in X^S \ \forall z\in X \ \forall i\in S \ \ u_i(z|x^S)\geq v_i\}, \ \ S\subset N.$$

Then, by the above discussion, we may take as x^S a self-binding strategy for S; and any utility vector in V(S) is attainable by exchanging only within the coalition S. Thus, the pair (N,V) amounts to a market game, which has been usually analyzed without the assumption of binding agreements. By Corollary 4, however, we can now confirm that market games really do not require the binding agreements.

Corollary 4 may also help us understand the fact that every subgame of a market game has a nonempty core. In games with this property, any coalition will not disrupt itself, but this is precisely what the self-binding strategy intends to do. Therefore this property of a market game can be traced back to the fact that every coalition of a pure exchange game has a self-binding strategy.

Concluding Remarks

A market game has been a central economic application of a cooperative game in characteristic function form. No explicit assumption on binding agreements has been made in its traditional analyses, which can now be supported from a general behavioral basis. A market game (N,V) is one that can be derived from a strategic game in which every coalition has a self-binding strategy: V(S) is a set of utility vectors to coalition S obtained from the pure exchange game through the "self-binding α-derivation".

Under a similar environment on communications and agreements among players, Bernheim, Peleg and Whinston (1987) have defined the solution concept, the coalition-proof Nash equilibrium (CPNE). The conceptual difference of CPNE to the self-binding strategy is of course obvious: CPNE is literally coalition-proof, while the self-bindingness is "disruption-proof". Deviating coalitions in CPNE assume no reactions from other players, whereas they must assume every conceivable reaction in the α-deviations.

From a standpoint of a coalition trying to maintain itself in a self-binding way, assuming no reactions from others will not make sense. It is the other extreme that provides a behavioral basis for the coalition that has only insufficient knowledge about how nonmembers will react.

References

R.J.Aumann, Subjectivity and correlation in randomized strategies, *Journal of Mathematical Economics* 1(1973)67-96.

D.Bernheim, B.Peleg and M.Whinston, Coalition-proof Nash equilibria I. Concepts, *Journal of Economic Theory* 42(1987)1-29.

A.Mas-Colell, Cooperative equilibrium, in:J.Eatwell, M.Milgate and P.Newman eds., The New Palgrave: Game Theory (The Macmillan Press,1987).

J.Nash, Two-person cooperative games, *Econometrica* 21(1953)128-140.

D.Ray, Credible coalitions and the core, *International Journal of Game Theory* 18(1989)185-187.

H.Scarf, On the existence of a cooperative solution for a general class of n-person games, *Journal of Economic Theory* 3(1971)169-181.