慶應義塾大学学術情報リポジトリ
Keio Associated Repository of Academic resouces

| Title | SELF-BINDING COALITIONS |
|---|---|
| Sub Title | |
| Author | 中山, 幹夫(NAKAYAMA, Mikio) |
| Publisher | Keio Economic Society, Keio University |
| Publication year | 1998 |
| Jtitle | Keio economic studies Vol.35, No.1 (1998. ) ,p.1- 8 |
| JaLC DOI | |
| Abstract | Assuming away the possibility of binding agreements, we define a strategy for a coalition, called a self-binding strategy, that sustains the coalition independently of the strategies of all other players. It is shown that a market game in characteristic function form is one that can be derived from a strategic game in which every coalition has a self-binding strategy. |
| Notes | |
| Genre | Journal Article |
| URL | https://koara.lib.keio.ac.jp/xoonips/modules/xoonips/detail.php?koara_id=AA00260492-19980001-0001 |

慶應義塾大学学術情報リポジトリ(KOARA)に掲載されているコンテンツの著作権は、それぞれの著作者、学会または出版社/発行者に帰属し、その権利は著作権法によって保護されています。引用にあたっては、著作権法を遵守してご利用ください。

The copyrights of content available on the KeiO Associated Repository of Academic resources (KOARA) belong to the respective authors, academic societies, or publishers/issuers, and these rights are protected by the Japanese Copyright Act. When quoting the content, please follow the Japanese copyright act.

KEIO ECONOMIC STUDIES 35(1), 1–8 (1998)

# SELF-BINDING COALITIONS

Mikio NAKAYAMA

*Faculty of Economics, Keio University, Tokyo Japan*

*First version received September 1997; final version accepted May 1998*

*Abstract*: Assuming away the possibility of binding agreements, we define a strategy for a coalition, called a *self-binding* strategy, that sustains the coalition independently of the strategies of all other players. It is shown that a market game in characteristic function form is one that can be derived from a strategic game in which *every* coalition has a self-binding strategy.

## 1. INTRODUCTION

In cooperative games of coalitional form, every coalition is assigned, through the so-called characteristic function, what it can achieve by itself. For example, in market games, every coalition is able to rearrange endowments among the members of the coalition and thereby achieve maximal utility vectors within the coalition. The set of utility vectors that a coalition can obtain is usually thought of as well-defined if, as in the market game, the utilities of the members of a coalition are not affected by the actions of players outside the coalition. But if we are to be more faithful to the description of actions of players, even in the pure exchange situation (on which the market game is constructed), no player can be independent of actions of other players. The 'pure exchange game' considered by Scarf [1971] clearly shows this, since a consumption vector of each player is the aggregate vector of goods transferred from all players. In what sense, then, the attainable payoff vectors for a coalition in a market game is well-defined? Put it more generally, what can a coalition achieve by itself when strategic interactions among players are explicitly taken into consideration?

The founders of game theory, John von Neumann and Oskar Morgenstern, treated this problem by the maximin argument between a coalition and its complement; and later, Aumann and Peleg [1960] gave general definitions of attainable payoff vectors of a coalition via $\alpha$ and $\beta$ notions which are extensions of the maximin and minimax arguments, respectively. These approaches rely, at least conceptually, on the implicit assumption that players are able to make binding

*Acknowledgement.* The author is grateful to Mamoru Kaneko and Jun Wako, and an anonymous referee for helpful comments.

agreements when a game is in a non zero-sum situation. Von Neumann and Morgenstern wrote: "The player who lives up to his agreement must possess the conviction that the partner too will do likewise", and assumed the *auxiliary concepts such as "agreements", "understandings"* (von Neumann and Morgenstern [1947, pp. 223–224]). Nash (1953) argued that an external mechanism or "a sort of umpire" is necessary for the enforceability of contracts and commitments; and also Aumann [1973] stated in a similar way that in cooperative games, the agreement must be externally enforced.

While this assumption of externally enforceable agreements may potentially enable players to form any coalition, it is also true that such an assumption is not always satisfied in economic situations, in which case coalition formation should not be a matter of assumption but of theoretical feasibility. And also, from a methodological point of view, if an assumption can be dispensed with, then it should be dispensed with at all. Thus, if it is possible to construct a theory of coalitions without the assumption of externally enforceable binding agreements, it would replace the existing one.

In this paper, we will make a modest step along this direction. Namely, the problem is this: given a coalition, is it possible for this coalition to attain by itself payoff vectors which are enough to maintain itself in a self-binding way? In other words, under what conditions can the members of a given coalition make an agreement on the choice of joint strategies by themselves without resort to external authorities? Such an agreement would have to meet several conditions; of which we shall be concerned with the following two. One is that the coalition must not break up thereby. This will be taken care of in a consistent way that the coalition admits no credible deviations by subcoalitions, where credible deviations are deviations at which no further credible deviations by subcoalitions occur. The word *credible* is used here in a similar sense to that given by Ray [1989]; and also in an analogous spirit to the coalition-proofness given by Bernheim, Peleg and Winston [1987]. The other requirement is that the coalition must prepare for the worst state that might be inflicted upon by the complementary coalition. This conservative requirement would be justified because the agreement must be convincing among the members of a coalition; that is, it must deliver safe and sure prospects for all conceivable actions of players outside the coalition, if it is to be self-binding at all. If the players outside the coalition did not react in an antagonistic way, it only makes the coalition better off. The classical notion of $\alpha$- or $\beta$-effectiveness would therefore be appropriate in taking care of this second requirement.

We will show that a coalition may, independently of the strategies of all other players, choose a joint strategy by which it can sustain itself in the sense that no further sub-coalitional credible deviations occur. We shall call such a strategy a *self-binding* strategy, and a coalition with this strategy a *self-binding coalition.*

Taking explicitly into consideration the strategic interactions, we may ask a legitimate question that under what conditions a given coalition has a self-binding strategy. Referring to the theorem of Scarf (1971), we will state a sufficient condition. Apparently this is a very stringent condition, requiring that the complementary coalition of a given coalition have a strategy that uniformly hurts the members of a given coalition. But it will turn out that the 'pure exchange game' mentioned above is a natural example of a game satisfying this condition. We can show that a strategic game describing the pure exchange has the property that every coalition is self-binding. Since a *market game* is a game in characteristic function form that can be derived from a strategic game in this class, this result tells us that a market game is a typical example of a cooperative game for which there is a theoretical base to dispense with the assumption of externally enforceable binding agreements.

## 2. 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 $X:=X^N$. 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 need an auxiliary notion of deviation that embodies the self-bindingness. Let $T\subset S\subset N$. Then, given $z^S\in X^S$ and $x^T\in X^T$, we denote by $z^S|x^T$ the strategy $|S|$-tuple in which $z^T$ is replaced with $x^T$.[1]

DEFINITION 1. For all nonempty $T\subset N$, we say that $T$ has an *$\alpha$-deviation* at $x\in X$ if and only if there exists $y^T\in X^T$ such that for all $z\in X$, $u_i(z|y^T)>u_i(x)$ for all $i\in T$.

DEFINITION 2. For all nonempty $T\subset N$, we say that $T$ has a *self-binding $\alpha$-deviation* at $x\in X$ if and only if $T$ has an $\alpha$-deviation $y^T\in X^T$ at $x$ such that for all $z\in X$ there exists no $R\subset T\,(R\neq T)$ which has a *self-binding $\alpha$-deviation* at $z|y^T$.

Note the recursion in the definition. Every single player $i\in T$ has a self-binding $\alpha$-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

[1] This notation is intended to keep the consistency of the following definitions 1, 2 and 3 when we take $T=N$ or $S=N$.

subset $R$ of $T$ must confront the similar strategic environment as that of $T$, i.e., $R$ must take all the reactions of $N$–$R$ into consideration.

DEFINITION 3. 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 $\alpha$-deviation at $z \mid x^S$.

By a self-binding strategy for $S$, the members of $S$ can assure themselves a certain level of utilities that is enough to bind themselves in $S$ whatever strategies the complementary coalition may choose. A *self-binding coalition* is one that has a self-binding strategy. 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 strategic formalization of the credibility concept. The following lemma is therefore a restatement of the credibility in the strategic form.

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

*Proof.* It will be enough to check (ii). Suppose $T$ has an $\alpha$-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 $\alpha$-deviation at $z \mid y^T$, then it follows that $T$ has a self-binding $\alpha$-deviation at $x$. If, for some $z \mid y^T$, there exists $R \subset T\,(R \neq T)$ which has a self-binding $\alpha$-deviation at $z \mid y^T$, then there exists $w^R \in X^R$ such that for all $w^{N-R} \in X^{N-R}$,

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

which implies that $R$ has a self-binding $\alpha$-deviation at $x$.

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

PROPOSITION 2. (i) *Let $x \in X$. Then, $x$ is a self-binding strategy for $N$ if and only if $x$ is in the $\alpha$-core.* (ii) *If the $\alpha$-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 $\alpha$-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 $\alpha$-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 $\alpha$-deviation at $z \mid y^S$. Let $w$ be any $\alpha$-deviation at $x$ satisfying for all $z \in X$ that

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

Then, no $T \subset S$ $(T \neq S)$ must have a self-binding $\alpha$-deviation at $z \mid w^S$, since $y^S$ is a self-binding $\alpha$-deviation. Hence $w^S$ must be a self-binding $\alpha$-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 of the grand coalition $N$ is equivalent to the existence of an $\alpha$-core; and for any coalition $S$, either $S$ itself is self-binding, or its subcoalition is self-binding.

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 an *opposing strategy* $d^{N-S} \in X^{N-S}$ to $S$ if and only if for all $z^S \in X^S$ and $z^{N-S} \in X^{N-S}$,

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

PROPOSITION 3. *Assume that for all $i \in N$, $u_i$ is quasi-concave in $x \in X$; and that $S \subset N$ is nonempty and proper. Then, $S$ has a self-binding strategy if $N$–$S$ has an opposing strategy to $S$.*

*Proof.* Let $d^{N-S}$ be the opposing 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 an opposing strategy, it follows that for all $x^{N-S} \in X^{N-S}$,

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

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

## 3. DISCUSSION

The opposing strategy, a strategy that hurts uniformly the members of the complementary coalition, appeals to intuition, but will be hard to exist in general. The existence of such a strategy entails a simple structure of payoffs to $S$. Recall that coalition $S$ is *$\alpha$-effective* for a payoff vector $v_S = (v_i)_{i \in S}$ if $S$ can assure itself the payoff vector $v_S$; that is, if there exists a strategy $x^S \in X^S$ such that for all $x^{N-S} \in X^{N-S}$, $u_i(x^S, x^{N-S}) \geq v_i$ for all $i \in S$. And coalition $S$ is $\beta$-effective for a payoff vector $v_S$ if $N$–$S$ cannot prevent $S$ from getting $v_S$; namely, if for all $x^{N-S} \in X^{N-S}$, there is a strategy $x^S \in X^S$ such that $u_i(x^S, x^{N-S}) \geq v_i$ for all $i \in S$. Note that if $S$ is $\alpha$-effective for $v_S$ then it is $\beta$-effective for $v_S$, but not conversely in general (see Aumann and Peleg (1960)). But an opposing strategy makes these two notions equivalent as shown below.

PROPOSITION 4. *Assume that N–S has an opposing strategy, where $S \subset N$ is nonempty and proper. Then, given a payoff vector $v_S$, S is α-effective for $v_S$* iff *S is β-effective for $v_S$.*

*Proof.* Let $S$ be $\beta$-effective for $v_S$. Then, for all $x^{N-S} \in X^{N-S}$ there is a strategy $x^S \in X^S$ such that $u_i(x^S, x^{N-S}) \geq v_i$ for all $i \in S$. Thus, for an opposing strategy $d^{N-S} \in X^{N-S}$ there is $x(d^{N-S}) \in X^S$ such that $u_i(x(d^{N-S}), d^{N-S}) \geq v_i$ for all $i \in S$. But, since $d^{N-S}$ is an opposing strategy, it follows that for any $z^{N-S} \in X^{N-S}$,

$$u_i(x(d^{N-S}), z^{N-S}) \geq u_i(x(d^{N-S}), d^{N-S}) \geq v_i \qquad \text{for all} \quad i \in S\,,$$

which implies that by choosing $x(d^{N-S}) \in X^S$, $S$ can assure itself the payoff vector $v_S$.

Thus, what $S$ can assure itself is precisely those payoff vectors which $N$–$S$ cannot prevent $S$ from getting. Moreover, the set of payoff vectors that $S$ can assure itself is determined by the opposing strategy of $N$–$S$. Jentzsch (1964) called a strategy with this property "optimal" and the payoffs structure with the optimal strategy "classical" paying attention to the zero-sum-like situation between $S$ and $N$–$S$. Thus, the existence of an opposing strategy will be limited only to a narrow class of games—the "classical" games, especially in economic contexts.

Nevertheless, there is a natural important economic example for Propositions 3 and 4; namely, the pure exchange game (see Scarf (1971), and also Mas-Colell (1987)). Consider the following game $G$ in strategic form: For each $i \in N$, let $w^i \in \Re^m_{++}$ 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

$$X^i = \{x^i = (x^{il}, \cdots, x^{in}) : \textstyle\sum_{j \in N} x^{ij} \leq w^i\,, \text{ and } x^{ij} \in \Re^m_+ \text{ for all } j \in N\}\,.$$

The utility function $u_i$ is given by

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

where $f_i$ is continuous, quasiconcave in $x$ and monotone nondecreasing in $\sum_{j \in N} x^{ji}$.

Scarf (1971) proved that this game has an $\alpha$-core. The grand coalition $N$ is therefore self-binding by Proposition 2 (i). Moreover, every coalition is self-binding in this game. For, every nonempty proper coalition $N$–$S$ may naturally allocate the endowments only among the members of N–$S$; namely, N–$S$ always has a strategy $x^{N-S}$ satisfying

$$x^{ji} = 0 \in \Re^m_+ \quad \text{for all} \quad j \in N\text{–}S \quad \text{and} \quad i \in S\,.$$

By the monotonicity of $u_i$, this strategy $x^{N-S}$ can be easily identified with an opposing strategy to $S$. Such an opposing strategy makes good sense in the context of pure exchange. Thus by Proposition 3, we may state:

COROLLARY 5. *In the pure exchange game G, every nonempty coalition S is self-binding.*

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 opposing strategy of N–$S$.

Given a pure exchange game $G$, one can derive, through the self-binding strategy, the set $V(S)$ of payoff vectors that $S$ can assure itself:

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

$V(S)$ is the set of payoff vectors for which $S$ is $\alpha$-effective; and hence, is also $\beta$-effective by Proposition 4. Now, $y^S=\{y^i\}_{i\in S}$ is called an $S$-allocation if $y^i\in R^m_+$ for all $i\in S$, and $\sum_{i\in S}y^i=\sum_{i\in S}w^i$. Then, defining $x^S$ for any given $S$-allocation $y^S$ by

$$x^{ij,k}=\frac{w^{i,k}y^{j,k}}{\sum\limits_{i\in S}w^{i,k}}\,; \qquad i,j\in S\,, \quad k=1,\cdots,m\,,$$

the set $V(S)$ can be shown to coincide with the following set $v(S)$:

$$v(S)=\{v_S=(v_i)_{i\in S}: \exists S\text{-allocation } y^S\ \forall i\in S\ f_i(y^i)\geq v_i\}\,, \qquad \text{for all} \quad S\subset N\,,$$

which is the characteristic function of a market game (see Scarf (1967)). Thus a market game $(N, v)$ can be derived from the pure exchange game, which together with Corollary 5 provides a theoretical base to justify the tradition that a market game, a typical cooperative game, has been usually analyzed without referring to the binding agreements.

Corollary 5 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 is self-binding.

## 4. 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 is made in its traditional analyses, which may now be supported from a general behavioral basis. A market game $(N, v)$ is intuitively plausible and well defined, but it is also a game that can be derived from a more basic strategic game in which every coalition has a self-binding strategy.

The coalition-proof Nash equilibrium (CPNE) by Bernheim, Peleg and Whinston (1987) is a solution concept defined under a similar environment on communications and agreements among players. The conceptual difference of CPNE to the self-binding strategy lies in that CPNE is literally coalition-proof, while the self-bindingness is "disruption-proof". Deviating coalitions in CPNE assume no reactions from other players at all, whereas they must assume every conceivable reaction in the $\alpha$-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


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

Aumann, R. J. and Peleg B. (1960) Von-Neumann-Morgenstern solutions to cooperative games without side payments, *Bulletin of the American Mathematical Society* **66**: 173–179.

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

Jentzsch G. (1964) Some thoughts on the theory of cooperative games, in M. Dresher, L. S. Shapley and A. W. Tucker eds., *Advances in Game Theory*, Ann. of Math. Study 52, Princeton University Press.

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

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

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

Scarf H.(1967) The core of an n person game, *Econometrica* **35**: 50–69.

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

von Neumann J. and Morgenstern O. (1947) *Theory of Games and Economic Behavior*, Princeton University Press.