

EC516 Contracts and Organisations, LSE

Lecture Notes on Binding Agreements

Debraj Ray

1. SOME COOPERATIVE GAME THEORY

1.1. The Characteristic Function. Cooperative game theory starts with the *characteristic function*, a description of the possibilities open to every possible coalition of players. Formally, let N be a set of players. A *coalition* is any nonempty subset of N . Denote coalitions by S, T , etc. A characteristic function assigns a set of payoff vectors $V(S)$ to every coalition S (efficient for that coalition). These are payoff vectors that takes values in \mathbb{R}^S . An important special case is the transferable utility (TU) characteristic function, in which for each coalition there is a just a *number* $v(S)$, describing the overall worth of that coalition. $V(S)$ is then the set of all divisions of that worth among the players in S . The characteristic function is fundamental to game theory. As we shall see later in this course, it is not a tool that takes us very far in many important game-theoretic situations. But there are many contexts in which it has proved remarkably useful. Here are some examples: 1. *Local Public Goods*. Person i gets utility $u_i(c, g)$, where c is money and g is a local public good. Then I can construct $V(S)$ — or its Pareto frontier — by solving the following problem: For any i and for arbitrary numbers $y_j, j \in S, j \neq i$,

$$\max u_i(c_i, g)$$

subject to $g = g(T)$ (production function), $T = \sum_{k \in S} (w_k - c_k)$ (where w_k is the money endowment of player k), and the restriction that $u_j(c_j, g) \geq y_j$ for all $j \neq i$. If $u_i(c, g) = c + v_i(g)$, we are in the quasi-linear case and this yields a TU characteristic function. 2.

Winning Coalitions. A special subgroup of coalitions can win an election, whereupon they get one unit of surplus. So $v(S) = 1$ for every winning coalition, and $v(S) = 0$ otherwise. 3.

Exchange Economies. Agent i has endowment ω_i . A coalition S can arrange any allocation a such that $\sum_j a_j \leq \sum_j \omega_j$. This generates $V(S)$. 4. *Matching Models*. Each agent has

“ability” α_i . When a group S of agents gets together, they can produce an output $= f_s(\alpha_S)$ (where s is cardinality of S and f_s is a family of functions indexed by s). In general, situations where there are no external effects across coalitions can be represented as characteristic functions. We will return to this business of external effects later in the course. Various assumptions can be made on characteristic functions. One standard one is *superadditivity*: if S and T are disjoint coalitions, with $\mathbf{y}_S \in V(S)$ and $\mathbf{y}_T \in V(T)$, then there is $\mathbf{z} \in V(S \cup T)$ such that $\mathbf{z} \geq \mathbf{y}_{S \cup T}$. For TU games, this just states that

$$v(S \cup T) \geq v(S) + v(T)$$

for all disjoint coalitions S and T . Superadditivity is a good assumption in many cases. In others (like matching models) it may not be.

1.2. The Core.

1.2.1. *Superadditivity and Balancedness.* A central equilibrium notion in cooperative game theory is that of the core. Look at the *grand coalition* N . Say that an allocation $\mathbf{y} \in V(N)$ is *blocked* if there is a coalition S and $\mathbf{z} \in V(S)$ such that $\mathbf{z} \gg \mathbf{y}_S$. *Remarks* We can make this blocking notion weaker; in situations with some transferability it will not matter which definition is used. In the TU case, for instance, we simply say that \mathbf{y} is blocked if there is a coalition S with $v(S) > \sum_{i \in S} y_i$. The *core* of a characteristic function is the set of all unblocked allocations. Superadditivity isn't good enough for a nonempty core: *Example.* Let $N = 123$, $v(i) = 0$, $v(ij) = a$ for all i and j , $v(N) = b$. Then if $a > 0$ and $b > a$, the game is superadditive. On the other hand, suppose that the core is nonempty. Let \mathbf{y} be a core allocation. Then

$$y_i + y_j \geq a$$

for all i and j . Adding this up over the three possible pairs, we see that

$$2(y_1 + y_2 + y_3) \geq 3a,$$

or $b \geq 3a/2$. This is a stronger requirement. There is a classical literature on the existence of the core. Perhaps the most famous theorem here is the characterization theorem of Bondareva and Shapley for TU games. To state this theorem, we need some definitions. For each i , denote by $\mathcal{S}(i)$ the collection of all subcoalitions (i.e., excluding the grand coalition) that contain player i . Let \mathcal{S} be the collection of all subcoalitions. A *weighting scheme* assigns to every conceivable subcoalition S a weight $\delta(S)$ between 0 and 1. A weighting scheme is *balanced* if it has the property that for every player i

$$\sum_{S \in \mathcal{S}(i)} \delta(S) = 1.$$

A TU characteristic function is *balanced* if for every balanced weighting scheme δ ,

$$(1) \quad v(N) \geq \sum_{S \in \mathcal{S}} \delta(S) v(S).$$

Let's pause here to make sure we understand these definitions. Here are some examples of balanced weighting schemes: (a) Weight of 1 on all the singleton coalitions, 0 otherwise. (b) Weight of 1 on all the coalitions in some given partition, 0 otherwise. (c) Weight of 1/2 each on "connected coalitions" of the form $\{i, i+1\}$ (modulo n). Notice that if a characteristic function is balanced the inequality (1) must hold for the weighting system (b). This proves that balancedness implies superadditivity. To show that the converse is false, take the three-person example above, look at the case in which $b < 3a/2$, and apply balancedness for the weighting system (c). Notice that a balanced weighting scheme need not be "symmetric" across players or coalitions. E.g., in (b), any partition will do.

1.2.2. *The Theorem of Bondareva and Shapley.* The following classical theorem characterizes nonempty cores for TU characteristic functions.¹

THEOREM 1. Bondareva (1962), Shapley (1967). *A TU characteristic function has a nonempty core if and only if it is balanced.*

¹For NTU games, an appropriate extension of the balancedness concept is sufficient, but it isn't necessary.

Proof. Suppose that \mathbf{y} belongs to the core. Let δ be some balanced weighting system. Because $\sum_{i \in S} y_i \geq v(S)$ for any coalition S , we know that

$$\delta(S) \sum_{i \in S} y_i \geq \delta(S) v(S).$$

Adding up over all S ,

$$\sum_{S \in \mathcal{S}} \delta(S) v(S) \leq \sum_S \delta(S) \sum_{i \in S} y_i = \sum_i \left\{ \sum_{S \ni i} \delta(S) \right\} y_i = \sum_i y_i \leq v(N).$$

This proves the necessity of balancedness. Sufficiency is a bit harder. To do this, we first think of a characteristic function as a vector in a large-dimensional Euclidean space \mathbb{R}^m , with as many dimensions m as there are coalitions. Suppose, contrary to the assertion, that v is balanced but has an empty core. Pick $\epsilon > 0$. Construct two sets of characteristic functions (vectors in \mathbb{R}^m)

$$A \equiv \{v' \in \mathbb{R}^m \mid v'(S) = \lambda v(S) \text{ for all } S, v'(N) = \lambda[v(N) + \epsilon], \text{ for some } \lambda > 0\},$$

and

$$B \equiv \{v' \in \mathbb{R}^m \mid v' \text{ has nonempty core}\}.$$

The first set contains all the scalings of our old characteristic function, slightly amended to give the grand coalition a bit more than $v(N)$ (by the amount ϵ). The second set is self-explanatory. Now, by our presumption that v has an empty core, the same must be true of the slightly modified characteristic function when $\epsilon > 0$ but small. For such ϵ , then, A and B are nonempty and disjoint sets. It is trivial to see (just take convex combinations) that A and B are also convex sets. So by the well-known separating hyperplane theorem, there are weights $\beta(S)$ for all S (including N), not all zero, and a scalar α such that

$$(2) \quad \sum_{S \in \mathcal{S}} \beta(S) v'(S) + \beta(N) v'(N) \geq \alpha \text{ for all } v' \in A,$$

and

$$(3) \quad \sum_{S \in \mathcal{S}} \beta(S) v'(S) + \beta(N) v'(N) \leq \alpha \text{ for all } v' \in B.$$

First, choosing $v' \equiv 0$ in B (which has a nonempty core) and then by taking λ arbitrarily small in A we easily see from (2) and (3) that α must be zero. Next, notice that $\beta(S) \geq 0$ for every S . For if not for some S , simply find some $v' \in B$ with $v'(s) < 0$ and large, while all other $v'(T) = 0$. This will contradict (3). Third, note that $\beta(N) < 0$. For if not, there are two possibilities. If $\beta(N) > 0$. Then we can violate (3) by choosing $v' \in B$ with $v'(N) > 0$ and large, while all other $v'(S) = 0$. The other possibility is that $\beta(N) = 0$. In this case $\beta(S) > 0$ for some S (all the β 's cannot be zero by the separation theorem). Then take $v' \in B$ with $v'(S) = D > 0$, $v'(N) = 2D$, while all other $v'(T) = 0$. For large D we contradict (3) again. So we can now divide through by $-\beta(N)$ in (3) and transpose terms to get

$$(4) \quad v'(N) \geq \sum_{S \in \mathcal{S}} \delta(S) v'(S) \text{ for all } v' \in B,$$

where we've defined $\delta(s) \equiv -\beta(S)/\beta(N)$ for each S . We claim that δ is a balanced weighting system; i.e., that

$$\sum_{S \ni i} \delta(S) = 1 \text{ for every } i.$$

For any player i and any number D , construct a game v' such that $v(S) = D$ for all $S \ni i$ and $V(S) = 0$ otherwise. Such a game must have a nonempty core: simply consider the allocation $y_i = D$ and $y_j = 0$ for $j \neq i$. So $v' \in B$. But now notice that by taking D to ∞ or $-\infty$ we can contradict (4), unless $\sum_{S \ni i} \delta(S)$ is precisely 1. To complete the proof, apply all this to (2). We have

$$(5) \quad v'(N) \leq \sum_{S \in \mathcal{S}} \delta(S) v'(S) \text{ for all } v' \in A.$$

Take $\lambda = 1$ in A , then — recalling that $\epsilon > 0$ — (5) reduces to

$$v(N) < \sum_{S \in \mathcal{S}} \delta(S) v(S),$$

which contradicts the balancedness of v . □

1.2.3. Remarks on the Theorem, and More on Balancedness. The Bondareva-Scarf theorem as stated here is a bit unsatisfactory. The reason is that to verify balancedness we need an “infinite number” of checks. After all, there are infinitely many balanced weighting schemes and checking all of them is seemingly no less difficult than directly looking for a core allocation. But one can make further progress here. To do so, say that a family of coalitions is *balanced* if there is some balanced weighting scheme under which this family is precisely the set of coalitions that receive positive weight. For instance: any partition is a balanced family; in a three person game, $\{12, 23, 13\}$ is a balanced family but $\{12, 23\}$ can never be a balanced family (why?). Now, observe that the union of two balanced family must also be a balanced family. Simply take *any* convex combination of the balanced weighting systems for the two families, and this will form a balanced weighting system for the union. This also proves, by the way, that there is a continuum of balanced weighting systems. But there is a certain set of balanced families for which the balanced weighting system is unique. These are the so-called *minimal balanced families*, which contain no sub-families that are also balanced. For instance: in a three person game, $\{12, 23, 13\}$ is a minimal balanced family but $\{1, 2, 3, 12, 23, 13\}$ is not (it is balanced, though). CLAIM: a minimal balanced family is associated with a unique balanced weighting system (see Shapley (1967)). Finally, observe that all we have to do to check the balancedness of a characteristic function is to check it for all the minimal balanced families (everything else can be done by taking linear combinations). Thus, because there are only a finite number of minimal balanced families and a unique balanced weighting scheme is associated with each of them, one has only to run a finite number of checks. If the game is already known to be superadditive, we can restrict the analysis further to those minimal balanced families which do not have any pair of disjoint coalitions.

1.2.4. Some Applications of Bondareva-Shapley. Here are some special cases and a couple of economic applications. *Symmetric Games.* A TU characteristic function is *symmetric* if $v(S)$ only depends on S via its cardinality s ; so we can write $v(s)$ instead of $v(S)$. CLAIM: A symmetric game has a nonempty core if and only if $v(n)/n \geq v(s)/s$ for all s . To prove this, assume first that $v(n)/n \geq v(s)/s$. Consider the “constant allocation” $y_i = v(n)/n$. For every coalition S of size s , $\sum_{i \in S} y_i = sv(n)/n \geq v(s)$, so this allocation is in the core. So our condition is sufficient. To prove that it is necessary, fix any $1 \leq s < n$ and consider the balanced weighting scheme $\delta(T) = 0$ if T is not of size s , and $\delta(T) = 1/\binom{n-1}{s-1}$ otherwise.

Since every i belongs to precisely $\binom{n-1}{s-1}$ coalitions of size s , this is a balanced map, and so — invoking Bondareva-Shapley and noting that there are $\binom{n}{s}$ coalitions of size s —

$$v(n) \geq \binom{n-1}{s-1}^{-1} v(s) \binom{n}{s} = v(s)n/s.$$

As an application of this, consider the following simple game: *Majority Voting*. Odd number of voters n . $v(s) = 1$ if $s > n/2$, $v(s) = 0$ if $s < n/2$. This game has an empty core. This observation is connected to the nonexistence of voting equilibria in multidimensional situations. *Convex Games*. A game is *convex* if $v(S \cup T) \geq v(S) + v(T) - v(S \cap T)$. Obviously a strengthening of superadditivity. It is easy to see that a convex game is balanced. The TU version of the public goods game discussed earlier is convex. Suppose that there are n agents, and agent i has endowment w_i . The public goods production function is given by any continuous increasing function $g(r)$, where r is the total resources expended to produce the good. If g is the amount of the good and r_i the resources contributed by agent i , the payoff of that agent is $u(g) + (w_i - r_i)$, where u is some continuous increasing function satisfying the Inada end-point condition at $g = \infty$. To prove that this game is convex, denote by $f(r)$ the composition function $u(g(r))$ and let r_s any maximizer of the function $sf(r) - r$. Then it is easy to see that for any coalition S with cardinality s ,

$$v(S) = sf(r_s) - r_s + \sum_{i \in S} w_i.$$

Pick two coalitions S and T with cardinality s and t . Let $p = |S \cup T|$ and $q = |S \cap T|$. Wlog let $r_s \geq r_t$. Then

$$\begin{aligned} v(S \cup T) + v(S \cap T) &= [pg(r_p) - r_p] + [qg(r_q) - r_q] + \sum_{i \in S \cup T} w_i + \sum_{i \in S \cap T} w_i \\ &= [pg(r_p) - r_p] + [qg(r_q) - r_q] + \sum_{i \in S} w_i + \sum_{i \in T} w_i \\ &\geq [pg(r_s) - r_s] + [qg(r_t) - r_t] + \sum_{i \in S} w_i + \sum_{i \in T} w_i \\ &\geq [sg(r_s) - r_s] + [p - s]g(r_t) + [qg(r_t) - r_t] + \sum_{i \in S} w_i + \sum_{i \in T} w_i \\ &\geq [sg(r_s) - r_s] + [tg(r_t) - r_t] + \sum_{i \in S} w_i + \sum_{i \in T} w_i \\ &= v(S) + v(T). \end{aligned}$$

1.3. The Problem of Farsightedness.

1.3.1. The Credible Core. The core is the set of all allocations which are unblocked by any subcoalition. But this definition does not put subcoalitions to the test in the same way as it does the grand coalition. Specifically, the core definition does not examine the credibility of the blocking allocations. Put another way, agents are not farsighted and do not see forward to the ultimate consequences of their own actions. Recall that an allocation for S is any element of $V(S)$. Think of a solution concept F as one that assigns a subset of allocations

(possibly empty) to every coalition S : thus $F(S) \subseteq V(S)$ for every S . For instance, in the case of the core, $F(S)$ is the set of all unblocked allocations. Formally:

$$C(S) = \{\mathbf{y} \in V(S) \mid \text{there is no subset } T \text{ of } S \text{ and } \mathbf{z} \in V(T) \text{ such that } \mathbf{z} \gg \mathbf{y}_S\}.$$

Credibility would require that we apply the same restriction to the blocking coalition as we do to the original coalition. Here is a core-like definition:

$$(6) \quad C^*(S) = \{\mathbf{y} \in V(S) \mid \text{there is no subset } T \text{ of } S \text{ and } \mathbf{z} \in C^*(T) \text{ such that } \mathbf{z} \gg \mathbf{y}_S\}.$$

Claim. $C^*(S) = C(S)$ for all coalitions S . **Proof.** Clearly $C(S) \subseteq C^*(S)$ for all S . Suppose that for some S , equality does not hold. Then there is $\mathbf{y} \in C^*(S) \setminus C(S)$. Because $\mathbf{y} \notin C(S)$, there is $T \subseteq S$ which blocks \mathbf{y} using $\mathbf{z} \in V(T)$. Because $\mathbf{y} \in C^*(S)$, $\mathbf{z} \notin C^*(T)$. So \mathbf{z} is blocked by some coalition W using an allocation \mathbf{w} in $C^*(W)$. Now $W \subseteq S$. Moreover, it is easy to check that if (W, \mathbf{w}) blocks (T, \mathbf{z}) while (T, \mathbf{z}) blocks (S, \mathbf{y}) , then (W, \mathbf{w}) blocks (S, \mathbf{y}) . But this contradicts our presumption that $\mathbf{y} \in C^*(S)$, and the proof is complete. \square

1.3.2. Refinements of the Core and Credibility. While the core equals the credible core, applications of the credibility idea to *refinements* of the core concept generally take us beyond the core. For instance, consider the *egalitarian solution*, a solution concept in which all coalitions are required to share their worth “as equally as possible” subject to the blocking constraints. More generally, suppose that for every set of “feasible” outcomes $F(S)$ for coalition S , there is a choice rule σ that picks out a *subset* of outcomes $\sigma F(S) \subseteq F(S)$. *Examples.* 1. The usual core. σ is the identity map. 2. Egalitarianism [applies only to TU games]. σ picks out the Lorenz-maximal elements of $F(S)$. 3. Welfare functions. σ is such that for each coalition W , there is a welfare function W_S so that $\sigma F(S)$ is the set of all allocations in $F(S)$ that maximize W_S on $F(S)$. Now define a core-like solution concept using σ ; call it Σ . It has the property that for each coalition S , $\Sigma(S) = \sigma F(S)$, where

$$F(S) = \{\mathbf{y} \in V(S) \mid \text{there is no subset } T \text{ of } S \text{ and } \mathbf{z} \in \Sigma(T) \text{ such that } \mathbf{z} \gg \mathbf{y}_S\}.$$

The *egalitarian solution* is a special case in which σ picks the Lorenz-maximal outcomes. [Warning. Now the blocking notion — “weak” or “strong” — may matter significantly.] *Example.* $N = A \cup B$, where $|B| = |A| - 1 > 0$. For any coalition S , $v(S) = \min\{|S \cap A|, |S \cap B| + 1\}$. This example corresponds to the well-known “right and left gloves” situation. It is well known that the core of this game gives one unit each to members of B , and none to members of A , even if A and B are both very large sets. There exists, however, a unique egalitarian allocation for the grand coalition:

$$\begin{aligned} y_i &= \frac{|B|}{2(|B| + 1)} \text{ if } i \in A \\ &= 1/2 \text{ if } i \in B \end{aligned}$$

1.3.3. What about Non-Subset Coalitions? Once we are studying farsightedness — requiring that coalitions block with allocations that themselves satisfy the same stability property — the question of blocking with non-subsets invariably comes up. The reason is simple: when studying the grand coalition for the usual core, there is no question of blocking with anything but a subset because all coalitions are subsets of the grand coalition. But that changes as soon as we start examining the credibility of the blocking coalition itself. The classic idea of von-Neumann and Morgenstern can be adapted to deal with this situation. They define

what is called the *stable set*. It is a self-referential concept, just as credibility is. Such a set is a subset of the space of all feasible allocations, and must satisfy two properties: (a) *internal stability*; no element of a stable set should block another element of that stable set, and (b) *external stability*; every element outside a stable set should be blocked by *some* allocation inside that stable set. A stable set, then, is a set of allocations that is both internally and externally stable. Notice that this is a general concept, and it can be applied to pretty abstract situations (with or without coalitions) in which there is a set of allocations and a pairwise blocking relation defined on that set. There is a huge literature on this stuff. But I will talk here about the way the Von-Neumann-Morgenstern idea is applied to coalitions and the characteristic function (in fact, it is the route that these authors originally chose). In the sort of notation we've already set up, the stable set is a simple extension of (6), but we have to work on tracking an entire allocation for everyone instead of just for a coalition.

A solution will now consist of a *partition* of the grand coalition — call it π — and an allocation $\mathbf{y} \in \mathbb{R}^n$ such that for every coalition $S \in \pi$, $\mathbf{y}_S \in V(S)$. Let us use the notation x to denote a pair of these things: (π, \mathbf{y}) . The space of all x — call it X — is the set of all allocations. We will say that $x = (\pi, \mathbf{y})$ *blocks* $x' = (\pi', \mathbf{y}')$ if there is a coalition $S \in \pi$ such that $\mathbf{y}_S \gg \mathbf{y}'_S$. The stable set X^* will be a subset of X , and indeed may be described as follows:

$$(7) \quad X^* = \{x \in X \mid \text{there is no } x' \in X^* \text{ such that } x' \text{ blocks } x\}.$$

[Note the self-referential nature of the definition. This is why it can be set up as a fixed point problem. Also note that the definition automatically meets both the internal and external stability requirements.] The conceptual problems with the stable set have been discussed by Harsanyi and even more forcefully by Chwe (1994). Let us unpack the definition in (7). Suppose that x is in the stable set. In what sense is it stable? After all, there may well be x' (not in the stable set, of course, by definition) such that x' blocks x . Suppose that some coalition S is made better off as a result. Now the stable set definition steps in and says, it doesn't matter that S is better off, because x' is going to get blocked in its turn (by something that is stable, as it happens), so *there is no point in looking at the initial block*. On the other hand, if x is *not* in the stable set, it does get blocked by something that is stable. That sort of block must be taken seriously. So out goes x in this case. This second feature seems fine, but what about the first? Just because S 's initial block fails, should it then not initiate the block to start with? It may very well want to do so anyway, because “at the end of the chain of deviations” it may well be better off. For instance, S may be even better off once x' is counterblocked! At this point the blocking notion starts to get problematic and conceptually very shaky. If gains are to be evaluate using chains of deviations (instead of a single one), which chain to look at? Should one use *some* chain or *all chains* to check gains? The problem gets even more complicated when we look at counterblocks. This motivates a noncooperative approach, based on bargaining, to the coalition formation problem. Of course, such an approach has it sown set of problems, but let us see how far it can take us.

1.4. The Problem With Characteristic Functions. These problems are compounded further, and new problems arise, when we realize that the characteristic function itself is inadequate in many situations.

The standard approach to the problem of binding agreements, starting from the normal form, is to convert the normal form game into characteristic function form, and analyze the core of

the cooperative game so induced. There are several options to choose from in making such a conversion. But, in general, the specific conversions used do not enjoy obvious consistency properties. Consider, for instance, the notion of the α -core. This notion presumes that when a coalition deviates, it does not expect to receive more than what it does when members of the complementary coalition act to minimax this coalition's payoffs. There is no reason why the complementary coalition should behave in this bloodthirsty fashion, and there is no reason for the deviating coalition to *necessarily* expect or fear such behavior.²

The easiest way to see the problem is to consider the example of a Cournot duopoly. Here, the α -core is the set of *all* individually rational Pareto optimal allocations. The reason is simple: under weak assumptions, one player can always be pushed to the point where it is not possible for him to earn any profits. But it should be obvious that any agreement that yields a player less than his Cournot-Nash payoff cannot constitute a binding agreement: by breaking off negotiations, this payoff is what he can credibly expect. When von Neumann and Morgenstern introduced the idea of converting games into characteristic functions, they were fully aware of this problem (but unfortunately, their successors have not been equally aware, and the characteristic function gets disproportionate attention in cooperative game theory):

“The desire of the coalition $-S$ to harm its opponent, the coalition S , is by no means obvious. Indeed, the natural wish of the coalition $-S$ should not be so much to decrease the expectation value ... of the coalition S as to increase its own expectation value. These two principles would be identical ... when Γ is a zero-sum game, but it need not be at all so for a general game ...” (p. 540) But they go ahead anyway: “Inflicting losses on the adversary may not be directly profitable in a general game, but it is the way to put pressure on him. He may be induced by such threats to pay a compensation, to adjust his strategy in a desired way, etc. ... It must be admitted, however, that this is not a justification of our procedure — it merely prepares the ground for the real justification which consists of success in examples.” (p. 541)

²A similar conceptual criticism applies to the β -core and the Strong Nash equilibrium.