EC516 Contracts and Organisations, LSE

2. Lecture Notes on Binding Agreements

Debraj Ray

1. An Approach to Coalition formation

1.1. **What We've Learnt So Far.** Here is a summary of the main points from the lecture on cooperative game theory.

1. As the concept of the core teaches us, efficiency need not be an equilibrium outcome, but it is unclear what an equilibrium outcome is.

2. Indeed, even when the core is nonempty (and especially when it *is* empty) the core need not be a good definition of equilibrium, especially of blocking is restricted.

3. Farsightedness is a consideration. It is not enough to block an ongoing agreemen without worrying about how, in turn, the block itself may be further blocked.

4. Most importantly, the characteristic function is itself inadequate to capture most economic situations of interest.

1.2. **Two Examples.** The following situations illustrate some of the main points.

Example. Three Cournot oligopolists produce output at a fixed unit cost, c, in a homogeneous market with a linear demand curve: $p = A - bx$. They are free to form coalitions among themselves, and this includes the option of forming the grand coalition of all three players. Recall that by standard calculations, that the Nash profit accruing to a single firm in an m-player Cournot oligopoly is

$$\frac{(A-c)^2}{b(m+1)^2} = \frac{D}{(m+1)^2},$$

where $D \equiv (A-c)^2/b$. Now suppose that the three firms in our example are deciding whether or not to form a cartel. If they do, they will earn monopoly profits, which from the expression above equals $K/4$. Now it must be the case that in the proposed agreement between the three at least one of the firms is earning no more than $K/12$. What should this firm do?

Von Neumann and Morgenstern's characteristic function tells us that if this firm breaks off, it should anticipate whatever it is that the other firms can hold it down to. But this last number is zero, for it is certainly the case that the other two firms can flood the market and drive prices down to zero. So the characteristic function predicts that our firm should not object to *any* nonnegative return, however small. This is clearly absurd.

On the other hand, suppose that our firm anticipates that in the event of its defection, the other two firms will play a best response to the defector's subsequent actions. This implies that following the deviation, we are in a duopoly, where the deviant's return, using the general expression above is $K/9$. This exceeds $K/12$.

Does this mean a deviation from the three-player coalition is then justifiable? Not really: there are other considerations. Study the situation facing the two remaining firms once our deviant leaves. Their *total* return is $K/9$ as well, which means, of course, that one of them can be earning no more than $K/18$. If this firm were to leave and induce the standard three-person oligopoly, its return would be $K/16$. So faced with the irrevocable departure of one firm from the original agreement, the remaining firms will split up as well. But in this case, the original deviant gets $K/16$ too! So each member of the three-firm coalition should anticipate receiving $K/16$ as a result of such a deviation. It follows that the grand coalition in this example is a stable coalition structure (proposing the joint monopoly outcome with each firm getting at least $K/16$).

Example. Consider the provision of a public good by three symmetric agents. Describe the partition function in the following intuitive way. Assume that if the three players get together, they produce a per-capita utility of 1. If one player leaves, assume that he would get 2 by free-riding on the other players' provisions, *provided* that the other two players stay together. Thus far this is analogous to the Cournot model. What is different is that we consider a case where the remaining two players will indeed wish to stay together. Imagine that by doing so, they can get a per-capita utility of 0.25. If all three players are on their own, assume that no public good is produced and that each player gets 0.

In this case, and in contrast to the first example, a single deviant can credibly expect to get 2, simply because faced with this deviation, the remaining agents will find it in their best interests to cling together. Now we have a problem, because it is clear that in the grand coalition, at least one player must get strictly less than 2. We find it difficult, in this case, to avoid an inefficient outcome.

1.3. **Two Approaches.** These examples illustrate two aspects of a good model of coalition formation: the inadequacy of the characteristic function, and the issue of farsightedness. But it is is still unclear how to write down a satisfactory model.

Approach 1. Stick to the language of cooperative game theory: use notions such as blocking. In such a framework the basic unit of analysis is the *coalition*.

Approach 2. Try to formulate the process of negotiation as a noncooperative game in which *individuals* are the unit of analysis.

In these lectures, we focus on Approach 2. To do this, however, we need a model of proposals and counterproposals for coalitions. To develop such a model we draw on the Rubinstein bargaining model.

2. Rubinstein Bargaining

2.1. **Basics.** Rubinstein bargaining represents possibly one of the simplest examples of an infinite game, and it has many applications. So it is of interest in itself, and will also form the basis for what we do later.

Suppose there are two persons, call them 1 and 2. They are dividing a cake of size 1. They take turns in proposing divisions of the cake; at each round, person i proposes a division and person j must accept or reject. If there is an acceptance, the game ends and the proposed division is implemented. If there is a rejection, we move on to the next round, and proposer and responder switch roles.

If a period passes, the next period is discounted. The discount factor of player i is $\delta_i \in (0,1)$. Thus, if a division $(x, 1-x)$ is settled on at date t, the two payoffs are $\delta_1^t x$ and $\delta_2^t(1-x)$, and if no division is ever settled on at all, then the payoffs are zero.

A (proposal or response) *history* is a listing of all that has transpired up to any proposal or response round. A strategy for player i prescribes an initial proposal (if she is the very first proposer) and then conditional on every history, a fresh proposal (if she is the proposer at that history) or an accept-reject decision (if she is the responder at that history).

It is well known that *every* conceivable division of the cake can be sustainable as a Nash equilibrium. To see this, fix some division $(x_1, x_2) = (y, 1-y)$ and have each proposer stubbornly propose this division following any proposal history, and have each responder i use the response rule following any response history: yes if and only if $x \geq x_i$, where x is the latest proposal made under the response history.

However, such an equilibrium is not subgame perfect. Under the strategies described above, nodes in which player i is offered a bit less than x_i will never be visited. But subgame perfection requires that the optimality of the strategy at those nodes (visited or not) should be checked. If we do so, we see that the above strategies are not optimal. If person i feels that she will get x_i tomorrow she should certainly be willing to accept $\delta_i x_i + \epsilon$ today. But for small ϵ this number is in fact smaller than x, and the going strategy is telling her to refuse such an offer. So the above strategies are not "credible".

2.2. **Perfect Equilibrium in the Rubinstein Model.** The following is true:

THEOREM **1.** *There is a unique subgame perfect equilibrium payoff vectorin the two-person Rubinstein bargaining model.*

Proof. Let M_i be the *maximum* equilibrium payoff and m_i be the *minimum* equilibrium payoff to person i.

Note that when it is i's turn to propose, she can be sure that a proposed payment of $\delta_j M_j + \epsilon$ to j will be accepted by j for every $\epsilon > 0$. This means that i can always get at least $1 - \delta_j M_j$ in equilibrium. This proves that

$$m_i \geq 1 - \delta_j M_j. \tag{1}$$

Now examine M_j. Suppose that j wants to try and clinch an agreement today. She cannot get more than $1 - \delta_i m_i$. On the other hand, if j makes an unacceptable offer, the max she can get from tomorrow (discounted to today) is $\delta_j M_j$. It follows that

$$M_j \leq \max\{1 - \delta_i m_i, \delta_j M_j\}.$$

It is easy to see that $M_j > 0$ (why?). Therefore, the second term on the RHS above cannot be the one that attains the max. Consequently,

$$M_j \leq 1 - \delta_i m_i. \tag{2}$$

Combining (1) and (2), it is easy to see that

$$m_i \geq 1 - \delta_j M_j \geq 1 - \delta_j(1 - \delta_i m_i),$$

or

$$m_i \geq \frac{1-\delta_j}{1-\delta_i\delta_j}. \tag{3}$$

Now combining (1) and (2) in a slightly different way and using the same logic,

$$M_j \leq 1 - \delta_i m_i \leq 1 - \delta_i(1 - \delta_j M_j),$$

so that

$$M_j \leq \frac{1-\delta_i}{1-\delta_i\delta_j}.$$

Flipping the indices i and j,

$$M_i \leq \frac{1-\delta_j}{1-\delta_i\delta_j}. \tag{4}$$

Combining (3) and (4), we conclude that

$$m_i = M_i = \frac{1-\delta_j}{1-\delta_i\delta_j}. \tag{5}$$

□

We can now unpack this to figure out supporting strategies. Player 1 as proposer will always propose the division $(x^*, 1-x^*)$, where

$$x^* = \frac{1-\delta_2}{1-\delta_1\delta_2},$$

and accept any proposal that gives her at least $\delta_1 x^*$. Likewise, Player 1 as proposer will always propose the division $(1-y^*, y^*)$, where

$$y^* = \frac{1-\delta_1}{1-\delta_1\delta_2},$$

and accept any proposal that gives her at least $\delta_2 y^*$.

If the two are equally patient with common discount factor δ, then the proposer picks up $1/(1+\delta)$ and the responder picks up $\delta/(1+\delta)$. The difference arises from the first-mover advantage that the proposer has but in any case will wash out as the discount factor converges to one.

2.3. **More Than Two Players.** Additional issues arise in Rubinstein bargaining when there are three players or more. Of these, the most basic concerns issues of *protocol*: who proposes, who responds, etc. Two standard examples:

1. First rejector of going proposal proposes next.

2. New proposer drawn at random

In what follows we abstract from questions of heterogeneous patience; assume that everyone has the same discount factor δ. As before, there is a cake of unit size to divide.

2.3.1. *Equilibrium in Stationary Strategies.* A *stationary strategy* (sometimes called a Markovian strategy) calls upon the agent to *take the same action* always when it is her turn to propose or respond, provided that the ambient situation is the same.[1] neither history nor calendar time matters. [Of course, when we require such strategy profiles to be equilibria, they must be full-blown equilibria in the class of *all* strategies, stationary or not.

A proposal now is a vector $\mathbf{x} \in \mathbb{R}^n$ such that $\sum_i x_i \leq 1$.

Focus on responses. With stationary strategy profiles they must look like this: say yes if the amount given to you, person i, is above some threshold m_i, otherwise say no. Actually it is a bit more complicated than that because if the proposal gives you more (than your threshold) but gives some responder who responds after you less (than her threshold), you don't want to accept because you want to grab the initiative, at least in the rejector-proposes protocol. So the technically correct description of a response vector is a collection $(m_1^*, \ldots m_n^*)$ such that each responder i says yes if the proposal gives $x_j^* \geq m_j$ to every responder j who responds after i and says no if the proposal gives $x_j^* < m_j$ to some responder j who comes after i.[2]

Now go back to proposals made by i. He will simply get try to

$$z_i \equiv 1 - \sum_{j \neq i} m_j,$$

assuming this is nonnegative, and zero otherwise.

Variant 1. First rejector proposes. By rejecting i takes the initiative and gets z_i tomorrow, valued today at δz_i. Therefore

$$m_i = \delta z_i = \delta(1 - \sum_{j \neq i} m_j) \tag{6}$$

It is easy to check (do so!) that there is a unique solution to (6) with $m_i = m_j \equiv m^*$ for all i and j, so

$$m^* = \frac{\delta}{1 + (n-1)\delta},$$

and indeed, this is what each responder gets. The proposer picks up the remainder, which is easily seen to be

$$\frac{1}{1 + (n-1)\delta}.$$

All of this goes to equal division as $\delta \to 1$.

Variant 2. Random choice of proposers. If i rejects a proposal, two things can happen. One (with probability $1/n$), i gets the initiative and therefore $(1 - \sum_{j \neq i} m_j)$ tomorrow. Two

---

[1] In the present context, "the ambient situation is the same" simply means that she is responding to the same proposal.

[2] Strictly speaking, i is perfectly allowed to randomize if $x_i = m_i$ but we simply presume that he will accept then. It is easy to see that if he rejects with some probability in the indifference case then this cannot form part of any stationary equilibrium, because the corresponding proposer will not have a well-defined best response.

(with probability $(n-1)/n$), i remains a responder and then gets m_i tomorrow. Discounting and then taking expected values, we must conclude that Therefore

(7)
$$m_i = \delta \left[\frac{1}{n}(1 - \sum_{j \neq i} m_j) + \frac{n-1}{n} m_i \right].$$

Again, it is easy to see that (please verify) that there is a unique solution to (7) with $m_i = m_j \equiv \hat{m}$ for all i and j, so

$$\hat{m} = \frac{\delta}{n},$$

and this is what each responder gets. [The proposer picks up the remainder as in Variant 1.]

Notice that $\hat{m}$ in Variant 2 is smaller than m^* in Variant 1 (check this). This is as it should be, because in Variant 1 the rejector has "more power". Notice, however, that even in this case the outcome goes to equal division as the discount factor converges to one.

2.3.2. *Other Equilibria.* Unfortunately, the beautiful uniqueness result for two-person Rubinstein bargaining no longer survives with three or more players. This is why we shall be invoking the additional assumption of stationary strategies.

3. Games of Coalition Formation

3.1. **Partition Functions.** $N = \{1, \ldots, n\}$ is the set of *players.* A *coalition structure of* N is a partition π of N. A *partition function* $\mathbf{v}$ assigns to each coalition S in a coalition structure π a worth $v(S, \pi)$.

Notice the connection with TU characteristic functions. There, a coalition's worth depends *only* on that coalition itself. Here, it depends on the entire coalition structure.

Example. Cournot oligopoly. Here, for any coalition S, and partition π,

$$v(S, \pi) = \frac{D}{(m+1)^2},$$

where $D \equiv (A - c)^2/b$, and m is the *number* of coalitions in the partition π.

Example. Public goods with externalities. n agents (regions). Each contributes towards (global) pollution control. z units of control requires cost $c(z)$ (increasing, strictly convex). Payoff to agent when global control is Z and region contributes z is

$$Z - c(z).$$

Now consider a *coalition* S of regions of size s. Each region i contributes z_i. Then payoff is

$$[\sum_{i \in S} z_i + Z_{-i}]s - \sum_{i \in S} c(z_i).$$

So the coalition maximizes $s[sz - c(z)]$ by choice of z. Let $z(s)$ be the solution. So for every partition π with coalitions of size $\{s_1, s_2, \ldots, s_m\}$, and for $S_i \in \pi$,

$$v(S_i, \pi) \equiv \left(s_i z(s_i) - c(z(s_i)) + \sum_{j \neq i} s_j z(s_j) \right) s_i.$$

3.2. **Proposals and Responses.** Agents make proposals to coalitions and respond to proposals made to coalitions to which they belong. For each coalition S (to be interpreted in this sentence as the "remaining set" of players in some subgame) is assigned a probability distribution over initial proposers; take it to be uniform. Likewise, to each coalition S *to* which a proposal has been made, there is a given order of respondents (excluding the proposer of course).

Interpret the bargaining game as follows. Some initial proposer starts the game with player set N. She chooses a coalition S and then makes a proposal to this coalition.

Loosely speaking, a proposal is the division of the worth of a coalition among its members. But given a partition function, a worth is not well-defined until a coalition structure has formed in its entirety. Therefore a proposal must consist of a set of *conditional statements* that describe a proposed division of coalitional worth for every contingency; i.e., for every conceivable coalition structure that finally forms. More precisely, a proposal is a pair $(S, \mathbf{y})$, where $\mathbf{y}$ is a collection of allocations $\{y(\pi)\}_\pi$, one for each partition π that contains S, feasible in the sense that

$$\sum_{i \in S} y_i(\pi) = v(S, \pi).$$

Once a proposal (S, y) is made by a proposer i, attention shifts to the respondents in S, the order of which is obtained from $\rho^r(S)$ (with i, the proposer, eliminated from the list if needed). By a response we mean simply an acceptance or rejection of the going proposal. If all respondents accept, the players in S retire from bargaining, and the game shifts to the set of players remaining in the game.

If there is a rejection, the response process is immediately halted, and one unit of time passes which is discounted by everybody by a discount factor δ. It is assumed that the rejector gets to make the new proposal.[3] The proposer makes a fresh proposal, and the game continues exactly as before.

If and when all agreements are concluded, a coalition structure forms. Each coalition in this structure is now required to allocate its worth among its members as dictated by the proposals to which they were signatories. If bargaining continues forever, it is assumed that all players receive a payoff of zero.

We will assume that each coalition does strictly better by forming than by not forming at all; i.e., $v(S, \pi) > 0$ for all S and π with $S \in \pi$.

[3]Broader protocols can be accommodated. See comments in class.

3.3. **Strategies and Equilibrium.** A *(stationary) strategy* for a player requires her to make a proposal whenever it is her turn to propose, where the (possibly probabilistic) proposal is conditioned only on the current state of the game — the current player set and the coalitions that have already formed.

It also requires her to accept or reject proposals at every node where she is supposed to respond. Again, we impose the restriction that this (possibly probabilistic) decision not depend on anything else but the current set of players, the coalitions that have already left, as well as the identity of the proposer and the nature of the proposal that she is responding to.[4]

A *stationary (perfect) equilibrium* is defined to be a collection of stationary strategies such that there is no history at which a player benefits from a deviation from her prescribed strategy.

Note that our notion of equilibrium allows for mixed (behavior) strategies in three ways: (a) the proposer may randomly choose a coalition, (b) given the choice of a coalition, the proposer may randomly choose offers, and (c) respondents may mix over accepting and rejecting a proposal.

But it turns out that an equilibrium exists with a minimal need to randomize, as described in the theorem below.

Theorem **2.** *There exists a stationary equilibrium where the only source of mixing is in the (possibly) probabilistic choice of a coalition by each proposer.*

Proof omitted (see Ray and Vohra (1999) if you are interested).

4. Symmetric Partition Functions

A partition function is *symmetric* if the worth of a particular coalition in a given partition depends *only* on the number of individuals in each coalition in that partition.

With a little abuse of notation, the worth of a coalition $S_i \in \pi$, $v(S_i, \pi)$, can be written as $v(s_i, \mathbf{n})$, where $\mathbf{n}(\pi) \equiv (s_1, \ldots, s_k)$ is the collection of coalition *sizes* in π. We might call this a *numerical coalition structure.*

The examples given for partition functions above are both symmetric.

We begin the analysis by constructing a particular numerical coalition structure.

4.1. An Algorithm.

Our results make essential use of a simple recursive algorithm which we now describe.

For *any* vector $\mathbf{n} = (n_i)$ of positive integers (not necessarily a numerical coalition structure), define $K(\mathbf{n}) \equiv \sum n_i$. Use the notation ϕ to refer to the "zero-dimensional" or null vector containing no entries, and set $K(\phi) = 0$.

[4] Of course, it is only fair to also let her condition her yes-no decision on the identity and order of the other respondents, but this is already accounted for, because once the proposer and proposal is given, the protocol fixes the order of respondents.

Let $\mathcal{F}$ be the family of all such vectors (including ϕ) satisfying the additional condition that $K(\mathbf{n}) < n$. You can think of these as numerical coalition *substructures*, and the interpretation in what follows is that this is some structure that has "already formed" in a subgame.

We are going to construct a rule $t(\mathbf{n})$ that assigns to each member of this family a positive integer. [Interpretation: given that the substructure $\mathbf{n}$ has already formed, $t(\mathbf{n})$ is the size of the coalition that forms *next*.]

By applying this rule repeatedly starting from ϕ, we will generate a particular numerical coalition structure, to be called $\mathbf{n}^*$. There is no game theory in what follows and you can look at it as a pure mathematical construction. What will give it meaning are the theorems that will link the equilibria of our game to the structures generated by this algorithm

Step 1. For all $\mathbf{n}$ such that $K(\mathbf{n}) = n - 1$, define $t(\mathbf{n}) \equiv 1$.

Step 2. Recursively, suppose that we have defined $t(\mathbf{n})$ for all $\mathbf{n}$ such that $K(\mathbf{n}) = m + 1, \ldots, n-1$, for some $m \geq 0$. For any such $\mathbf{n}$, define

$$c(\mathbf{n}) \equiv (\mathbf{n}.t(\mathbf{n}).t(\mathbf{n}.t(\mathbf{n}))\ldots),$$

where the notation $\mathbf{n}.t_1 \ldots . t_k$ simply refers to the numerical coalition structure obtained by concatenating $\mathbf{n}$ with the integers $t_1, \ldots t_k$.

Step 3. For any $\mathbf{n}$ such that $K(\mathbf{n}) = m$, define $t(\mathbf{n})$ to be the *largest* integer in $\{1, \ldots, n-m\}$ that maximizes the expression

$$\frac{v(t, c(\mathbf{n}.t))}{t}. \tag{8}$$

Step 4. Complete this recursive definition so that t is now defined on all of $\mathcal{F}$. Define a numerical coalition structure of the entire set of players N by

$$\mathbf{n}^* \equiv c(\phi).$$

This completes the description of the algorithm.

4.2. **Connecting the Algorithm to the Equilibria of Our Game.**

4.2.1. *A Class of Equilibria that Yield n^*.* An equilibrium is *standard* if at the proposer at every stage makes — with positive probability — an acceptable proposal that includes herself in coalition. Presently we shall concern us with conditions that guarantee when a standard equilibrium exists, but for now let us see what the concept gives us.

Theorem **3.** *There exists $\delta^* \in (0, 1)$ such that for all $\delta \in (\delta^*, 1)$, any standard equilibrium must be of the following form. At a stage in which any substructure π has left the game (with associated numerical substructure $\mathbf{n}$), the next coalition that forms is of size $t(\mathbf{n})$ and the payoff to a proposer is*

$$a(\mathbf{n}, \delta) \equiv \frac{v(t(\mathbf{n}), c(\mathbf{n}))}{\delta t(\mathbf{n}) + (1-\delta)]}. \tag{9}$$

In particular, the numerical coalition structure corresponding to any such equilibrium is $\mathbf{n}^*$.

Theorem 3 shows that if acceptable offers are made (with some positive probability) at every stage, the equilibrium coalition structure of the bargaining game *must* yield the same numerical coalition structure as our algorithm.

The proof of this theorem will be provided in the next set of notes.