Bargaining Games

Andreu Mas Colell

Department of Economics, Harvard University, Cambridge, MA 02138, USA

This is a close rendition of the delivered lecture. The manuscript has been written and prepared by Shira Lewin, currently a doctoral candidate at Harvard University. The lecturer wishes to acknowledge and thank her for a very well done job.

1 Introduction

The topic to be reviewed in this lecture is included in what Bob Aumann described in his lecture as the bridges between cooperative and noncooperative theory. If I had all the time in the world, I would begin by presenting the basics of noncooperative game theory, but I cannot possibly do this. I will therefore remain very elementary, and I will be somewhat loose about the noncooperative concepts.[1] The flavor of what I will be doing today consists in writing down or describing game procedures, understood as noncooperative mechanisms for interaction, discussion, and the formulation of agreements about how to split things. These bargaining procedures will be set in a context which will stay very close to the frameworks presented by earlier lecturers. We will then see how the noncooperative solutions of the bargaining procedures relate to the axiomatic procedures presented earlier by others.

In interpreting these procedures, there are two positions that we can take — not quite two points of view, but two sources of light with which we can look at this sort of theory. The first is the descriptive source of light, and the second is the prescriptive.

The descriptive view In the descriptive view, the noncooperative procedure comes first, not only logically but also conceptually and theoretically. We are discussing bargaining procedures, and when we analyze these procedures, we may discover that the equilibria exhibit some relationship with an axiomatically based solution. Then, if we wish, we may call the bargaining procedure under discussion the noncooperative foundation of the axiomatic

[1] For general references on game theory, see Myerson (1991) or Osborne & Rubinstein (1994).

solution. But we certainly view the noncooperative approach as the conceptual starting-point.

The prescriptive approach The prescriptive approach relates more to implementation theory. (The elements of this theory will be presented in a forthcoming lecture.) Here, the point of view is to think of the axiomatic solutions as well founded on the axiomatic grounds on which they are presented. However, one recognizes that to reach them, it may be necessary to design devices — call them bargaining procedures — that will yield the axiomatic solutions as noncooperative equilibria. So, logically, the cooperative part comes first, and we really think of the noncooperative part of the theory as an instrument with which to obtain the cooperative result.

As I said, the distinction is very much there, but I would not want to trace a rigid boundary for the purposes of this lecture.

In this first part of the lecture, I will discuss two player games, and in the second part, I will say something about N-player games.

2 Two-player games

In this section, we have two players, those in $N = \{1, 2\}$. I will start with transferable utility (TU) games, but I will move to the non-transferable utility (NTU) case very soon.

2.1 TU case

2.1.1 Cooperative approach

Think of two players that have to split a pie. If they cooperate, then they will get a total amount of utils, or dollars or whatever, equal to $v(N)$. If they do not cooperate, then they will get a certain point $(c_1, c_2) = c \in \mathbf{R}$.

It would be tempting to adhere to the exact framework of a characteristic function, and write, say, $c_1 = v(1)$ and $c_2 = v(2)$. We could do this, but we will not. It is not clear that we should really think of c_1 and c_2 as if they were exactly what $v(1)$ and $v(2)$ would be in a cooperative framework. There is no need to regard c_1 as what 1 could get by himself, and similarly for c_2. We only require that the combination $c = (c_1, c_2)$ is what would happen if there were no cooperation.

We do assume that

$$c_1 + c_2 < v(N)$$

so that there is some reason to cooperate. Graphically, we present this in figure 1. Here, the segment AB is the utility possibility frontier $\{u_1, u_2 \,|u_1 + u_2 = v(N)\}$. Now, let us make a slight conceptual jump, and let us associate the vector c with the threat point of a bargaining problem (like those presented in W. Thomson's lecture). Then we see that each of

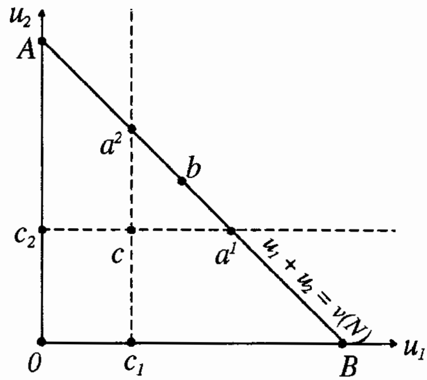

Figure 1: Standard solution in the TU case

the solutions that we have discussed from an axiomatic perspective has the property that it splits the surplus above c, rather than, say, giving each player $\frac{1}{2}v(N)$. Thus, we end up at the midpoint between a^1 and a^2 shown above in figure 1. We will call this split the **standard solution**:

$$b = \left(c_1 + \frac{v(N) - c_1 - c_2}{2}, c_2 + \frac{v(N) - c_1 - c_2}{2}\right) = \frac{1}{2}\left(a^1 + a^2\right)$$

where $v(N) - c_1 - c_2$ is the surplus that is split.

2.1.2 Noncooperative approach

There is a very simple way to obtain the standard solution noncooperatively: the all-or-nothing (or take-it-or-leave-it) mechanism. Consider the following bargaining procedure:

- Choose one player by tossing a coin. Call this player the proposer.
- The proposer proposes a split of $v(N) : (u_1, u_2)$.
- The respondent accepts or rejects.
  - acceptance $\Rightarrow (u_1, u_2)$.
  - rejection $\Rightarrow (c_1, c_2)$.

To solve this (extensive form) game, the natural noncooperative solution concept is backward induction (or, if you prefer, you can say that I am choosing the perfect equilibrium of this game). If 1 is the proposer, then how will 1 reason? He will say, "If I don't offer 2 at least c_2, then 2 will reject, because by rejecting she can get c_2." Therefore, 1 will propose the split that gives him the maximum amount compatible with 2 getting c_2, and this is the point a^1 in figure 1. (We always assume that player 2 is cooperative enough with player 1 that she breaks ties in his favor, so I won't have to worry about little ϵ's.) Thus, 1 will propose a^1, and similarly, when 2 is the proposer, she will propose a^2 and get all the surplus herself. If we accept the von Neumann-Morgenstern expected utility theory, then, in expectation, the outcome is $\frac{1}{2}\left(a^1 + a^2\right)$, which is exactly the standard solution defined above.

So, are we done? Have we implemented the standard solution noncooperatively? The answer is, not quite. Why not? Well, the procedure just described has some drawbacks. One is very apparent, and the other will become clear momentarily when we move to the NTU case. The apparent drawback is that we get the correct outcome only in expectation. It is true that, ex ante, each player i gets b_i, but the proposals that actually take place in the game are not the standard solution. They are either a^1 or a^2. (Actually, in the current TU case, there is an easy fix for this problem — just perform the procedure twice instead of once — but as we will now see, the matter is not always so simple.)

2.2 NTU case

2.2.1 Failure of the take-it-or-leave-it procedure

We draw the problem graphically as before (figure 2). Note that, with a^1

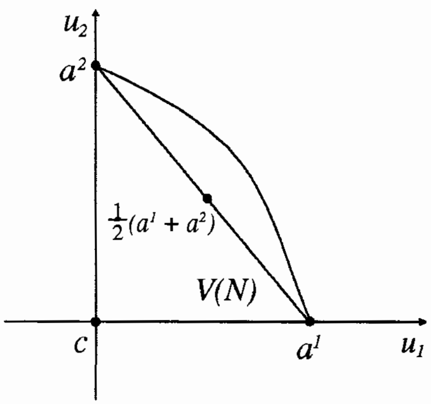

Figure 2: Failure of the one-stage mechanism in the NTU case

and a^2 defined as before, the 50/50 expectation of the two outcomes is not even efficient. Of course, we could get efficiency by not randomizing, and by simply imposing that 1 begin, or 2 begin, but then we would not implement the standard solution. While this is a very simple example, it illustrates an issue that tends to be a general problem. If you are in the TU case, then you can implement (in expectation) by relatively short bargaining procedures; but these procedures are not likely to be efficient if you have an NTU problem.

2.2.2 A multi-stage procedure

At this juncture, it seems logical to argue that if we want to get better outcomes, perhaps we should work with more elaborate bargaining procedures — in particular, bargaining procedures that keep repeating themselves, so that if somebody rejects, then this is not the end of the world; another round of negotiation may yet take place. I will now present a particular, but typical, multi-stage procedure. It goes as follows:

- As before, a player is chosen by tossing a coin, and she makes a (feasible) proposal $(u_1, u_2) \in V(N)$.
- The other player can accept or reject.
  - Acceptance $\Rightarrow (u_1, u_2)$.
  - Rejection $\Rightarrow$
    * With probability $\rho < 1$, the game repeats.
    * With probability $1 - \rho$, the players get c.

Note that in the case of rejection, the probability of breakdown is not 1, but only $1 - \rho < 1$. This number (which is a parameter of the problem) could be large or small, but I want you to think of it as small, so that if there is persistent rejection, then with high probability, the procedure will not terminate with breakdown immediately, but will do so only quite far in the future. A typical interpretation (but not the one that I want to emphasize here), is to think of $1-\rho$ as a rate of time-discounting. (In this case, we should interpret c as the utility that will be obtained if there never is agreement.) The point is that there is a cost of delaying agreement one round. This cost may be that of time passing, or it may be something else. For example, in the case of implementation, the designer can set a device which incorporates the possibility of breakdown.

The procedure just described is not the only possible one. Note, in particular, that it is time-stationary. We could, for example, also have a non-time-stationary rule. Fix a horizon $T < \infty$, such if there is no agreement by time T, everything stops and players get c. However, assume that up to time T there is no cost of delaying agreement. This is somewhat discontinuous (and nonstationary), and it doesn't lend itself to a simple analysis, whereas the device proposed above does.

2.2.3 Stationary Perfect Equilibrium

Which solution concept will we adopt? We have a game that will terminate with probability 1, but which is, in principle, infinite. There is no end of time, and therefore I cannot use backward induction. Instead, I will adopt a very simple solution concept called *stationary perfect equilibrium. Perfect* implies that we are still within the framework of backward induction. *Stationary* means that we are focusing on equilibria where chosen strategies do not depend on history or on calendar time. Proposals will be independent of whatever has happened in the past. Similarly, the responses will depend only on the proposal received, and not on past proposals.

Finally, note that we are still talking about an *equilibrium.* That is, I am referring to an equilibrium, in which the strategies happen to be stationary ones, but it is a true equilibrium. In particular, there is no restriction on the strategy set, and players do contemplate the possibility of every sort of complicated nonstationary deviation. In the universe of perfect equilibria — where every equilibrium is as good as any other — the equilibria that are most descriptively simple are the stationary ones. So just as, when one looks at a dynamical system, one first looks at the rest points, it makes some sense to look at the stationary equilibria first.

2.2.4 Graphical solution by equilibrium equations

I'm going to try to solve the equilibrium problem graphically. The treatment is not meant to be rigorous. Focus on a particular stationary perfect equilibrium. Call $b = (b_1, b_2)$ the expected payoffs at $t = 0$ when this equilibrium is played. Since the utility possibility set is convex, b must be a feasible point, but it does not need to be at the boundary of $V(N)$, and in figure 3, it is shown in the interior. Remember that we do not know a priori that efficiency is guaranteed (and in fact, as we will see, it is not).

Now suppose that player 1 is chosen to be the proposer. What will 1 propose? He will try to evaluate how much it costs 2 to reject. Well, if 2 rejects 1's proposal, then with probability ρ, everything is repeated, and because of stationarity, we come back to b. With probability $1-\rho$, we go to c. So the expected payoff vector is $\rho b+(1-\rho)\,c$. Hence, by rejecting 1's proposal, 2 can guarantee herself $\rho b_2 + (1-\rho)\,c_2$. Player 1 will therefore propose the point a^1 (shown in figure 3) that maximizes his own payoff subject to 2 getting at least $\rho b_2 + (1-\rho)\,c_2$, the minimum payoff that guarantees 2's acceptance. Similarly, 2 will propose the point a^2 that maximizes her payoff, subject to 1 getting at least $\rho b_1 + (1-\rho)\,c_1$, which guarantees that 1 accepts as well. Note that along the equilibrium path, there will be no rejection.

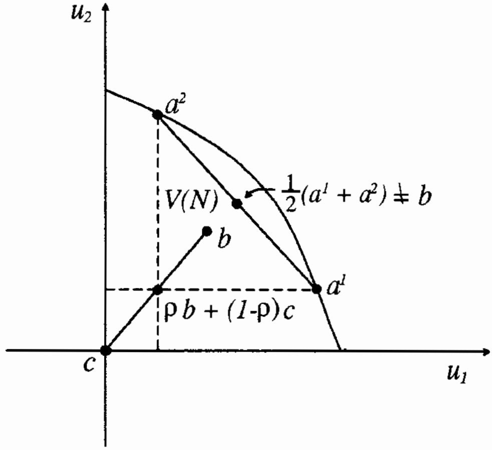

Figure 3: Graphical analysis of the mult-stage procedure

We conclude that, if 1 is the proposer, the outcome is a^1, and if 2 is the proposer, the outcome is a^2. So far, we have not brought any consistency conditions into the analysis. To close the system, we notice that the expected outcome of the equilibrium is b. Therefore, we must have

$$b = \frac{1}{2}\left(a^1 + a^2\right)$$

That is, the vector b must lie at the midpoint of the line segment between a^1 and a^2. This is a real condition. If I start with an arbitrary b, and then I construct a^1 and a^2 (as indicated above) and take their midpoint, I need not come back to b, and in figure 3, I do not. If I do happen to come back to b, then I have found an equilibrium. This is the case in figure 4.

The stationary equilibrium payoff vector b depends on ρ, but it can be verified that, given ρ, b is unique. Note that it is not efficient.

At this point, we can observe something very interesting. Normalize c to $(0,0)$ (this is just for convenience). Take the straight line through a^1 and a^2, and extend it until it hits the axes (figure 5). The two triangles BOA and a^1Da^2 are similar, and the line Ob splits them in half. Now, since b is the midpoint of the hypotenuse of a^1Da^2, it follows that b is also the midpoint of the hypotenuse of BOA. Now imagine that $1-\rho$ is very small, so that the triangle a^1Da^2 is also very small. Then the slope of AB is almost equal to the slope of the boundary of $V(N)$ near a^1 and a^2 (assuming that $V(N)$ has a smooth boundary).

So, for $1-\rho$ very small, we have, almost, the following property: The equilibrium payoffs b are efficient and are such that when we take the tangent

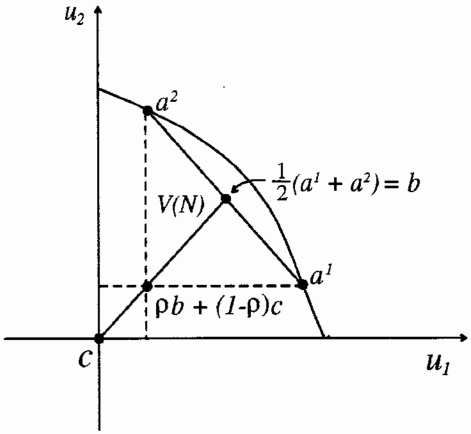

Figure 4: Equilibrium condition for the multi-stage procedure

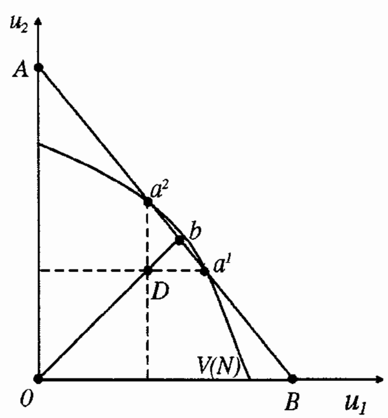

Figure 5: Approximate efficiency for large ρ

to the boundary of $V(N)$ at the equilibrium payoffs, b falls at the midpoint of this tangent (more precisely, the midpoint of the segment of this tangent that lies between its intersections with the axes). We should recognize this as the defining property of the Nash bargaining solution. Recall that for a utility possibility set as illustrated below (figure 6), if the vector b has the property that the segments Ab and bB have the same length, then it follows that b is the Nash solution. This comes right out of the axioms of the Nash

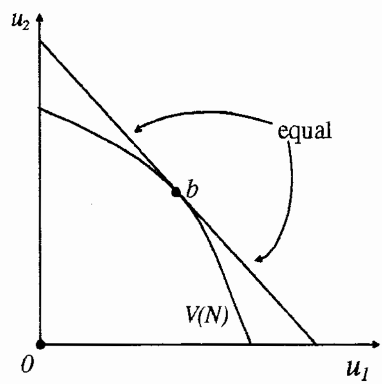

Figure 6: The Nash solution

solution. Consider first an economic budget set — a straight line (See figure 6). Then if we maximize the product of the coordinates on this line, we come to the midpoint. So b is the Nash solution for the budget set. But then the contraction independence axiom implies that, since when we go from the economic budget set to $V(N)$, we only make the utility possibility set smaller while b remains feasible, the solution remains b after the change.

We conclude that if the cost of renegotiation is very low, then at the first stage of the stationary perfect equilibrium of the bargaining procedure, the proposer will propose a payoff which is close to the Nash bargaining solution, and the respondent will accept. We emphasize that:

1. We obtain this result because, in principle, negotiation can go on for a very long time. But in fact, it will not go on for long. It will end in the first round.

2. The proposer does not really matter. Both agents will propose almost the same outcome.

2.3 Some Remarks

2.3.1 On the stationarity restriction

I have been talking about stationary perfect equilibria. In fact, there is a notable result due to Rubinstein (1982), which asserts that, for this model, to get the equilibrium payoffs we derived above, the stationarity restriction is actually not required. Every perfect equilibrium of this procedure has exactly the characteristic that we just described: On the equilibrium path, one player (whoever is the proposer) makes some proposal (the same for all equilibria), and this proposal is accepted. This is a very singular result, but I will not emphasize it. It is quite remarkable, but something of a jewel — admirable and beautiful, but hard to replicate. In particular, it does not generalize to more than two players.

2.3.2 A dynamic analysis

Associated with the previous discussion there is a nice dynamic analysis. I can not be precise here, but M. Maschler and collaborators have done much research on this topic. Suppose that you take exactly the bargaining procedure I have presented, except that you truncate it at period $T \gg 0$, when the world ends. Thus, we are contemplating a nonstationary problem such that the disagreement point is, say, 0, and such that at $t = 0$, we have a utility possibility set $V_0 = V(N)$. Then, there is a contracting sequence of utility possibility sets $V_1, V_2, \ldots V_T$ (figure 7) such that each $V_t = \rho^t V(N)$ is the set of feasible expected payoffs if there is no agreement before time t. Assume also that T is so large that ρ^T is nearly equal to 0.

This problem can be solved by backward induction. You just think about what would happen at the end of the world, and then given that, you look at stage $T - 1$, etc. Figure 7 illustrates the construction of the equilibrium expected payoffs $b(T-1)$ at $t = T-1$ from the equilibrium expected payoffs $b(T) = \frac{1}{2}\left[a^1(T) + a^2(T)\right]$ in the last period. You can proceed in this manner until, finally, you derive $b(0)$. The process will begin to look like a differential equation. We can then make the jump to real differential equations, so that the backward induction yields a system of differential equations which, as it turns out, converges to the Nash solution.

2.3.3 Variation in the breakdown point

I have assumed that the breakdown point c is given independently of the history that leads to breakdown. I could consider (why not?) a more complicated model in which the breakdown point depends, for example, on who has been responsible for the breakdown, perhaps the refuser or perhaps the last proposer. One can think of many variations. But let me focus on one.

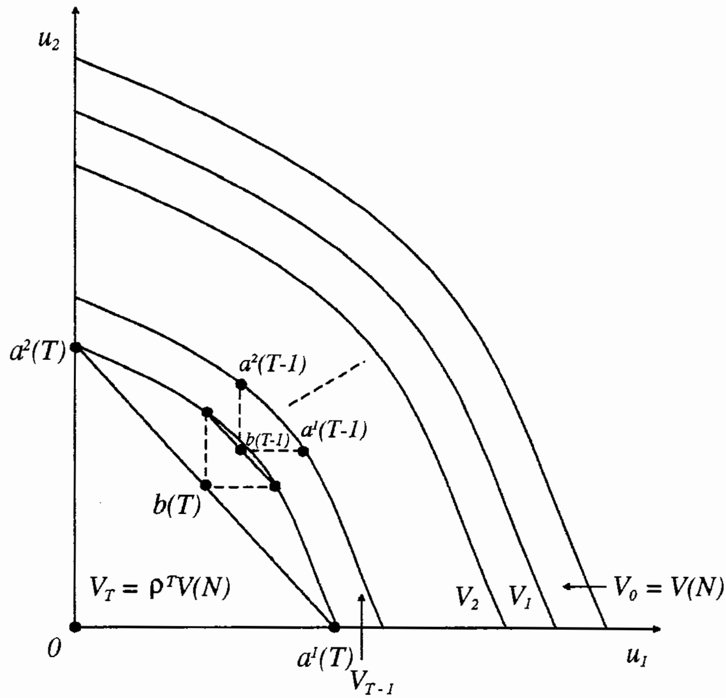

Figure 7: Truncated multi-stage problem

Suppose that $V(N)$ is as before, but that the breakdown point depends on who the last proposer was before breakdown. Otherwise, the bargaining procedure is as before. Again we have a ρ. The only change is that as we go through time, if player 1 makes a proposal, 2 rejects it, and there is breakdown, then we go to some point c^1. If the breakdown happens after 2 proposes and 1 rejects, then we go to c^2 instead (figure 8). Note that if you think of ρ in terms of time discounting then this doesn't make sense, but for other interpretations, it does make sense. When 1 evaluates the utility 2 gets from rejecting, he should consider c^1. If 2 rejects, then c^1 occurs with probability $1-\rho$, and play continues with probability ρ. It turns out (this is very easy to check) that to solve this model, we can proceed by constructing a kind of fake disagreement point (shown in figure 8): $c = \left(c_1^2, c_2^1\right)$. Then we can continue exactly as we did before. Taking c as the disagreement point, if ρ is large, the outcome will be nearly the Nash solution calculated from this fake disagreement point. Note that this disagreement point has no reality. Outcome c will never occur. What can occur is c^1 or c^2. But the theory can still make use of the fake disagreement point.

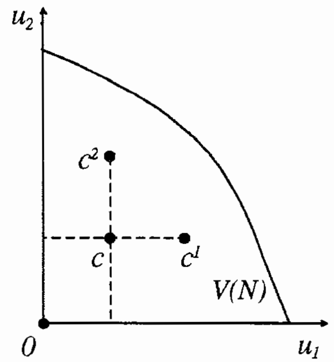

Figure 8: The breakdown point depends on the last proposer

One interesting point that I would like to mention is that there is no reason why things have to be as I drew them in figure 8. In fact, it could be as in figure 9. The only requirement is that c^1 and c^2 be in the feasible set $V(N)$. But, as constructed in figure 9, c need not be feasible itself. But we can still look at the bargaining procedure and derive its stationary equilibria. What will we get? The equilibrium will have to satisfy exactly the same equations as before. Given an equilibrium payoff vector b, we construct a^1 and a^2 as we did earlier, and it must be the case that b is at their midpoint. (See figure 9.) If $1-\rho$ is small, then to find something that is almost the solution, you look at c as a disagreement point, and you look at the point b on the boundary of $V(N)$ such that when you take the tangent at b, b is at the midpoint of the line segment AB shown in figure 10. Two things are worth noting about this construction. First, it amounts to guaranteeing the first order conditions (but not the second!) of the Nash product "maximization problem." Second, the solution need not now be unique.

2.3.4 What about Kalai-Smorodinsky?

My entire discussion has led us to the Nash bargaining solution. You heard in W. Thomson's lecture that the Kalai-Smorodinsky solution is as important as Nash's, so you may ask whether I can get Kalai-Smorodinsky's solution by a bargaining procedure similar to the one I have described above. I cannot give you an affirmative answer to this question. I could offer you some bargaining procedures, but these would be of a very different character. However, I can offer you some insight by obtaining a solution which is in the spirit of Kalai-Smorodinsky.

We do this as follows: Put $\rho = 1$, so that there is no cost of delay. To avoid being degenerate, also assume a fixed time horizon $T < \infty$; i.e. we repeat only T times. We now apply backward induction. Consider the last period. If 1 is the proposer, then 1 will offer $a^1(T)$; similarly, 2 will offer $a^2(T)$, as shown

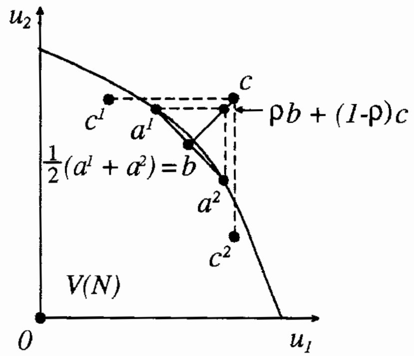

Figure 9: Disagreement point outside feasible set

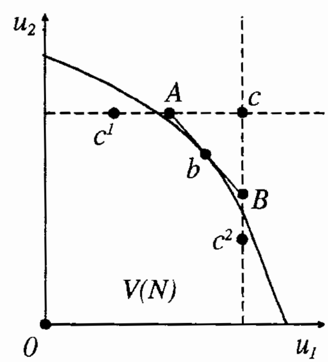

Figure 10: Near efficiency with a fake disagreement point

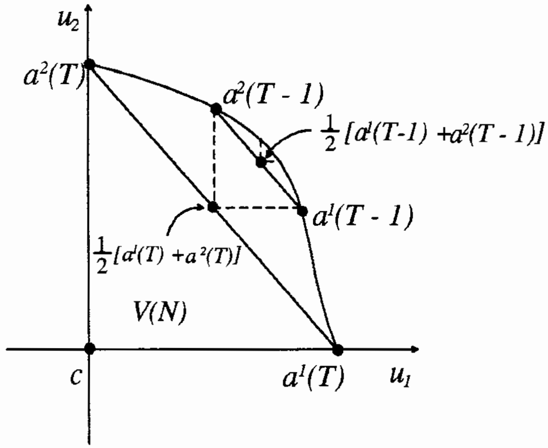

Figure 11: The Raiffa solution

in figure 11. Therefore, if there is rejection at $T-1$, the expected payoff is the midpoint $b(T) = \frac{1}{2}\left[a^1(T) + a^2(T)\right]$. What will happen in period $T-1$? Both players know that if the other rejects, then they get $b(T)$, so player 1 proposes $a^1(T-1)$ shown above, and 2 proposes $a^2(T-2)$. We continue this construction to get $a^1(T-2)$, $a^2(T-2)$, etc. As T grows large, we approach the boundary. This is not the Kalai-Smorodinsky solution. It is called the Raiffa solution, but it clearly seems to be in the same general category as Kalai-Smorodinsky's.

3 N-Player Games

I will now move to N players. The theory is less settled here, and so I will be much more particular and merely illustrative. There is much work on this topic, and I cannot possibly cover all the available results. Fortunately, the lecture by P. Reny will also touch on this general area, and he will complement our discussion quite well. The presentation from now on is in the spirit of implementation theory. Thus, I will just present instances of how, under certain restrictions, this or that solution concept can be supported by a noncooperative procedure. But I will not discuss whether such a procedure is sufficiently descriptive of "real bargaining."

3.1 Background

Since I want to relate my discussion to cooperative game theory, let me begin by describing the physical situation that is being contemplated in terms of the formalism of the characteristic function. I assume that there are N players and that for all $S \subset N$, $V(S)$ is the attainable set for S. I want to be a bit more precise here. Thus, in the spirit of an economic approach, let us think of this V as describing an economic situation in which people are endowed with resources and where there are no externalities of any kind, so that $V(S)$ is the set of utility possibilities that the group S can reach by using the resources of its members (this set is independent of whatever the players not in S eventually get).

I am focusing on this particular scenario because it is a very simple one in which there is no complication whatsoever in interpreting what the characteristic function is. In general I think that for the type of discussion that we are carrying out, it is indispensable to have an interpretation, a particular story about this characteristic function. Otherwise it is very hard to know what one is talking about. For example, in a model with externalities and other interactions, the characteristic function may have been constructed taking into account a certain number of strategic considerations that concern how the members of a coalition and its complement will behave. In this case, it would be artificial to analyze the problem using other strategic considerations. If we are focusing on the core, then the kind of strategic considerations that go into its definition may also be strategic considerations that lead us to a certain kind of characteristic function; while for the Shapley Value, say, we may be led to another characteristic function. I think, therefore, that you cannot separate the construction of the characteristic function from the solution concept that you are going to use, except in completely natural cases like pure resource problems. Since matters are already complicated enough here, we will stick to this case.

3.2 Two approaches to the N-player case

In the spirit of what I have done so far, I want to generalize the bargaining procedures presented earlier, and in a manner that fits in with the lessons of cooperative game theory. This already points one in certain directions. For example, in the $N = 2$ case, I focused on the Nash solution. That means I have already focused on a single-valued solution. It would not be natural now to move in the direction of the core, which is multi-valued even in the case of two players, and which certainly does not equal the Nash solution. In fact, the approach I am taking is directing me towards value-type solutions, and I will not resist this direction.

Let me remark that, in my view, it is not yet well understood what distinguishes the types of bargaining solutions that take us to the core from those that take us to value. Very vaguely, my impression is that the distinction has something to do with the meeting technology of the players — that is, with how different players meet. If the choice of players that meet is very strategic — namely, a player chooses and looks around for partners — then I think that we are pushed towards core-like notions; while if the meeting technology is that people meet at random somehow, so that people have very little choice about how they find their companions, then I think we are pushed more towards value solutions.

From now on, we have N players, and they need to meet to bargain over how to split some sort of pie. To analyze this situation further, it is critical to establish how players meet. In the value-oriented literature, we find two varieties of what we could call "meeting technologies." Because I am more familiar with one of them I will follow that one, but I must mention both. They both generalize the formulation we adopted in the two-player case.

The first is the technology of pair-wise meetings. The meeting of pairs (the "buyer" and the "seller") is pervasive throughout economics (cf. Gale 1986, Rubinstein & Wolinsky 1985). For the current problem, the technology of pair-wise meetings has been used by Gul (1989). In his important paper, there is a collection of people who have resources and who meet at random in pairs. When they meet, one proposer is chosen at random. The proposer makes a proposal to buy the resources of the respondent. The respondent may accept, in which case he disappears from the game with the payment, or he may not accept, in which case both members of the pair go back to the pool of players and negotiation continues in this manner.

The second technology, which is the one I will adopt, is that of multilateral meetings. More precisely, I mean by this that at any point in time, there is an assembly of all the bargainers, and the proposer (chosen in some manner) addresses the entire assembly. I will follow this approach, but I also recommend that you look at the Gul paper.

3.3 An illustrative example

3.3.1 Setup

I am going to present, as an example, a bargaining procedure which is a generalization of the previous (2-player) bargaining procedure. It is taken from Hart & Mas-Colell (1992). The key feature of this procedure is that players may drop out throughout the negotiation process.

The procedure is as follows: Assume that $S \subset N$ is the set of players still involved in negotiation. Initially, we will just have $S = N$.

- Choose a player i at random from S using a uniform distribution.

- Player i proposes a payoff vector $u \in V(S)$.

- Other players are asked (sequentially) if they agree or dissent.
  - All agree $\Rightarrow u$ is implemented.
  - Any player dissents $\Rightarrow$
    - With probability ρ, the game repeats.
    - With probability $1-\rho$, breakdown occurs.

What does breakdown mean? I don't want it to mean, as before, that everything is ended and that we go to some disagreement point $c \in V(S)$. I avoid this meaning because I want subcoalitions to matter. There are a multitude of other meanings of "breakdown" that one could consider. I encourage you to consider some. For example, "breakdown" could mean that some player is at risk of disappearing. It could be that there is some $\delta < 1$ such that each player disappears with his own resources with probability δ. If δ is small, then the probability of two players disappearing simultaneously is negligible, and so breakdown means that one of the players, chosen at random, will disappear, and we will go on with a smaller game.

To be specific, I will focus on another particular meaning of breakdown. I choose this particular example purely because it fits with the analysis of the Nash solution I presented before, and because I want to tie this analysis to the Shapley Value. Thus, I will present a breakdown technology which has the feature that if I look at the equilibria, then in the pure bargaining case I get the Nash solution, and in the case of transferable utility — another leading case for analysis — I get the Shapley Value.

The breakdown technology which I will use, and which, I could argue, is the only technology which works for this purpose, is the following: With probability $1-\rho$, the proposer disappears (taking with him and consuming his own resources). The game then repeats with only the players in $S \setminus \{i\}$. So proposers that are frivolous enough to invite rejection run the risk of being out of the game. At the same time, of course, they are not always thrown out of the game because they have resources that the other players value.

As before, I will look at the stationary perfect equilibrium. If I could do with perfect equilibrium, I would be happier, but unfortunately, in the games that we are analyzing, the set of perfect equilibria is large (if $N > 2$).

3.3.2 The equilibrium conditions

How do we analyze problems like this? We already know how to determine the stationary perfect equilibrium equations. For the case $N = 2$, I drew a picture (figure 4). Now, since there are many more than two equations, I cannot draw a picture, but I can still write down the equilibrium equations without any difficulty. The logic is the same.

The equilibrium objects are the proposals. For every foreseeable coalition S of players still in the game and for every $i, j \in S$, let $a_j^{S,i}$ represent the proposal that i makes to j if i is the proposer. Let's see now what sort of consistency conditions these numbers need to satisfy. Define the expected payoff of the players in S to be

$$a^S = \frac{1}{|S|} \sum_{i \in S} a^{S,i} \in V(S)$$

This is the average of all the payoffs that result if all players' offers are accepted. There will be two equilibrium conditions:

1. For all S and $i \in S$, $a^{S,i} \in \partial V(S)$; i.e. $a^{S,i}$ is efficient, and is therefore on the boundary of $V(S)$ and not in the interior.

2. If $i \in S$ is the proposer, then he will offer j the minimum possible payoff, which is what j would get if she rejected. He must make sure that j will not reject. If j rejects, then with probability ρ, everything will be repeated, and she will get a_j^S, and with probability $(1-\rho)$. i will be thrown out of the game, and j will expect to get $a_j^{S\backslash\{i\}}$ instead. Thus,

$$a_j^{S,i} = \rho a_j^S + (1-\rho) a_j^{S\backslash\{i\}} \tag{1}$$

We can then solve the system using conditions 1 and 2. Notice that the result is a stationary perfect equilibrium.

3.3.3 A Remark

If ρ is very close to 1, then the term $(1-\rho) a_j^{S\backslash\{i\}}$ is very small. So, the proposals to j will depend on who the proposer i is, but in fact, no matter who i is, this proposal will be very close to the average a_j^S, which does not depend on i. Therefore, the proposals of all the players in S will lie close together on the boundary of $V(S)$, so that the average of these proposals, a^S, will be almost efficient, as illustrated in figure 12. If we examine the equilibrium equations, then as long as ρ is close to 1, we see that, first, it won't matter very much who the first proposer is and, second, the average proposal will be approximately efficient.

3.3.4 The TU case

Do we know anything else about the case in which ρ is large? Well, in the pure bargaining case for two players, once we defined the disagreement point, the stationary equilibrium of this model was the Nash solution. The same argument generalizes to N players if the strict subcoalitions cannot generate gains from trade (the pure bargaining case). But we are interested in the general situation in which the worth of a subcoalition matters.

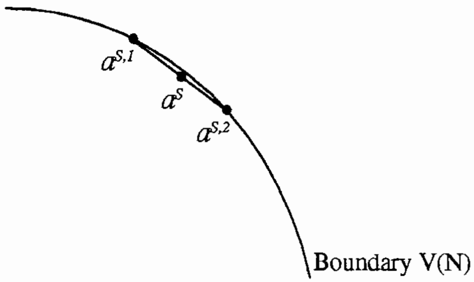

Figure 12: Approximate efficiency for large ρ

Consider the well understood TU case. Here, it turns out that, for any ρ, the stationary equilibrium payoffs are the Shapley values. This is, incidentally, why I chose this particular breakdown technology. While this result holds for all ρ, remember that in order to make sure that the proposals themselves (not just their average) are the Shapley values , we do need ρ to be close to 1. I will try to provide some intuition for this result. If you are familiar enough with the Shapley value, then you look at the system of equations and say, "Of course." But let me argue directly from the axioms. The Shapley value is characterized by four axioms.

1. Efficiency This is guaranteed for equilibrium payoffs because all the equilibrium proposals are on the boundary and the boundary is flat, so that the expectation is also on the boundary.

2. Equal Treatment I have never distinguished any particular player from any other. Thus, clearly, the "stationary perfect equilibrium payoffs" solution must be symmetric.

3. Additivity (Linearity) Write down the system of equations. In the linear (TU) case we simply have, for all S and $i \in S$,

$$\sum_{j \in S} a_j^{S,i} = v(S) \tag{2}$$

and therefore

$$a_i^{S,i} = v(S) - \sum_{\substack{j \in S \\ i \neq j}} a_j^{S,i} \tag{3}$$

A quick examination of (1) and (3) reveals that we can solve these for all the $a_j^{S,i}$'s recursively. We will thus obtain some complicated expression, but this expression will be linear in the $v(S)$'s.

4. Dummy Axiom This is where our particular breakdown technology comes into play. Intuitively, this technology implies that when some $i \in S$

makes a proposal to the players in S, if she has nothing to contribute to the other players, then these players do not pay any cost when they force a delay by rejecting her offer. Either the game is repeated, or i is kicked out, which makes no difference since the pie remains the same. Thus, the procedure gives no power to a dummy player. In a bargaining procedure, a player can have two types of power: one that derives from her resources, and another that derives from her ability to prevent agreement. A dummy has no resources, and this bargaining procedure makes the delays caused by her harmless to the other players. She therefore has no power of either kind.

Here is a formal proof that a dummy player gets nothing in a stationary perfect equilibrium. The proof is by induction. Suppose that the claim holds for all games up to size $N-1$. I will prove that it also holds for games of size N. Suppose that i is a dummy. I need to show, first, that when i proposes, he proposes 0 for himself, and, second, that when another player proposes, the proposal to i is 0.

Suppose that i is the proposer. How much will i propose to the other players? Adding the equilibrium equations (1) above, we get:

$$\sum_{j\neq i} a_j^{N,i} = \sum_{j\neq i}\left[\rho a_j^N + (1-\rho)\, a_j^{N\setminus\{i\}}\right]$$

Then substituting using (3), we get

$$v(N) - a_i^{N,i} = \rho\left[v(N) - a_i^N\right] + (1-\rho)\underbrace{v(N\setminus\{i\})}_{=v(N)} = v(N) - \rho a_i^N$$

since i is a dummy. Therefore, i proposes for himself

$$a_i^{N,i} = \rho a_i^N$$

Now suppose that $j \neq i$ is the proposer. Then, again using (1), we have

$$a_i^{N,j} = \rho a_i^N + (1-\rho)\, a_i^{N\setminus\{j\}}$$

Because i is a dummy, and $N\setminus\{j\}$ has only $N-1$ players, the induction hypothesis tells us that $a_i^{N\setminus\{j\}} = 0$. Therefore, in parallel to what we derived above, we have

$$a_i^{N,j} = \rho a_i^N$$

and this is true for all $j \neq i$.

Hence, whether i is the proposer or not, he gets ρ times his expected value. So by definition,

$$a_i^N = \frac{1}{N}\sum_{j\in N}\underbrace{a_i^{N,j}}_{=\rho a_i^N} = \rho a_i^N$$

which implies that

$$a_i^N = 0 \text{ because } \rho < 1$$

Therefore $a_i^{N,j} = \rho \cdot 0 = 0$ for all $j \in N$. ■

We have now shown that all the axioms of the Shapley value are satisfied. Therefore, the outcome must be the Shapley value.

3.3.5 Closing Remarks: The NTU Case

I repeat that what I have shown you her is only an example. However, it prompts a question: We have considered a bargaining procedure which in familiar cases gives very familiar solutions. Can we now do some conceptual boot-strapping, so that, after using cooperative theory to motivate a non-cooperative procedure, we can then go back to cooperative theory and discover what the particular non-cooperative procedure yields in the general NTU case? Interestingly enough, for $\rho \approx 1$, this procedure yields the "consistent solution" introduced by Maschler & Owen (1992). It does not yield either of the two more familiar solutions: the Shapley NTU value solution or the Harsanyi solution. Note that our particular bargaining procedure was not designed to yield the consistent solution, and that Maschler and Owen derived it from very different consistency-like requirements. Now you can ask, "What is the consistent solution?". I could spell it out for you, and I could also add:

Theorem: *When ρ is close to one, the stationary perfect solution of the bargaining procedure is close to the consistent solution.*

But since I have only −2 minutes left, I am going to simplify matters by transforming a theorem into a definition, and I will answer your question by saying that the consistent solution is the limit as $\rho \to 1$ of the stationary perfect equilibrium of the particular bargaining procedure that I have described!

References

Gale, D. (1986) Bargaining and computation, Part I: Characterization. *Econometrica* 54, 785-806.

Gul, F. (1989) Bargaining foundations of Shapley value. *Econometrica* 57, 81-95.

Hart, S., Mas Colell, A. (1992) Bargaining and value. *Econometrica,* forthcoming.

Maschler, M., Owen, G. (1992) The consistent Shapley value for games without side payments. In: R. Selten (ed.) *Rational Interaction.* Berlin: Springer-Verlag.

Myerson, R. B. (1991) *Game Theory.* Cambridge: Harvard University Press.

Osborne, M., Rubinstein, A. (1994) *A Course in Game Theory.* Cambridge: MIT Press.

Rubinstein, A. (1982) Perfect equilibrium in a bargaining model. *Econometrica* 50, 97-109.

Rubinstein, A., Wolinsky, A. (1985) Equilibrium in a market with sequential bargaining. *Econometrica* 53, 1133-1150.