Binding Agreements II: Coalition Formation

Debraj Ray, October 2006

1. The Coase Theorem

The bargaining approach can be usefully applied to address some classical questions in cooperative game theory. One of these concerns the formation of coalitions and the writing of binding agreements among groups of players. Many years ago, Ronald Coase argued that such negotiations would invariably end in an efficient outcome, and that bargaining power — whatever that is — would be reflected in the *allocation* of the surplus from that efficient outcome across the different players.

With transferable payoffs, in particular, the "Coase Theorem" would suggest that an "equilibrium outcome" of the negotiation process must maximize aggregate surplus. If there is a unique surplus-maximizing outcome, then this outcome must arise irrespective of the allocation of power across agents.

In these notes, I first do a little cooperative game theory. I then explain why traditional concepts in game theory are inadequate to fully address the Coase Theorem. Then I apply notions of perfect equilibrium and a natural variant of the bargaining model to go a little further in examining the Coase Theorem.

2. Some Cooperative Game Theory

2.1. **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. 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 the traditional development of cooperative game theory. 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.

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.

2.2. **The Core.** 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 $\mathbf{z} \gg \mathbf{y}_S$.[1]

In the TU case, 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.

A TU game is *symmetric* if the worth of each coalition is expressible as a function of the *number* of players in that coalition. Thus, with some mminor abuse of notation, we may write $v(S)$ as $v(s)$, where s is the cardinality of S. For symmetric TU games it is very easy to follow the lines of the example above to find a necessary and sufficient condition for the core to be nonempty.

[1] We can make this blocking notion weaker; in situations with some transferability it will not matter which definition is used.

Observation **1.** *The core of a symmetric TU game is nonempty if and only if*

$$\frac{v(n)}{n} \geq \frac{v(s)}{s} \tag{1}$$

for every coalition of size s.

Proof. To see that (1) is sufficient simply use equal division and show that it is a core allocation. For the necessity of (1), use the obvious extension of the argument in the example above. □

Of course, this elementary proposition can (and has) been further generalized. For TU games there is the famous Bondareva-Shapley theorem, while for NTU games we have the weaker but equally celebrated theorem of Scarf. Here is a brief discussion.

2.3. **More On Core Existence.** First, 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,

$$v(N) \geq \sum_{S \in \mathcal{S}} \delta(s) v(S). \tag{2}$$

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 (2) 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).

The following classical theorem characterizes nonempty cores for TU characteristic functions. [For NTU games, an appropriate extension of the balancedness concept is sufficient, but it isn't necessary. This is the theorem of Scarf, which we omit in these notes.]

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

Here is an entertaining (to some, anyway!) proof of this theorem, which I include for completeness. Separating hyperplanes make an appearance again.

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,

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

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 $\mathbb{R}^m$)

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

and

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

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

$$\sum_{S \in \mathcal{S}} \beta(S) v'(S) + \beta(N) v'(N) \geq \alpha \text{ for all } v' \in A, \tag{3}$$

and

$$\sum_{S \in \mathcal{S}} \beta(S) v'(S) + \beta(N) v'(N) \leq \alpha \text{ for all } v' \in B. \tag{4}$$

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 (3) and (4) 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 (4).

Third, note that $\beta(N) < 0$. For if not, there are two possibilities. If $\beta(N) > 0$. Then we can violate (4) 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 (4) again.

So we can now divide through by $-\beta(N)$ in (4) and transpose terms to get

$$v'(N) \geq \sum_{S \in \mathcal{S}} \delta(S) v'(S) \text{ for all } v' \in B, \tag{5}$$

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 (5), unless $\sum_{S \ni i} \delta(S)$ is precisely 1.

To complete the proof, apply all this to (3). We have

$$v'(N) \leq \sum_{S \in \mathcal{S}} \delta(S) v'(S) \text{ for all } v' \in A. \tag{6}$$

Take $\lambda = 1$ in A, then — recalling that $\epsilon > 0$ — (6) reduces to

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

which contradicts the balancedness of v. □

2.4. **So What For Agreements?** The core is a useful solution concept that takes a first step towards identifying outcomes that should or should not survive a process of negotiations. But it has several shortcomings.

First, if the core is empty, *something* still has to happen, presumably. Perhaps a structure of subcoalitions forms, or perhaps an allocation for the grand coalition still comes about, on the grounds that blocking will be "further" blocked. Cooperative game theory does attempt to travel along these lines by considering blocks that exhibit various levels of credibility.

However, once these routes open up, there is no guarantee that a core allocation will be implemented even if the core is nonempty. The broader possibilities described in the previous paragraph may still be pertinent even when the core is nonempty.

Put another way, the core is too black-boxed a concept. It makes no prediction when it is empty, and when it is nonempty, it does not examine the blocking allocations very carefully. Besides, focussing on the core as a definition of what's implementable *presumes* that the outcome of all negotiations should indeed be efficient.

Now, there are several ways to proceed. One might actually try to build a theory to deal with these issues that proceed using the methodology of blocking. Or one might try a more overtly noncooperative approach to the process of achieving cooperation. This is what we do here.

3. Coalitional Bargaining

3.1. **A Particular Game.** Take as given a TU characteristic function v. Note that the standard bargaining setup is a special case, in which $v(N) = 1$ and $v(S) = 0$ for all other coalitions. We are going to study a model in which everybody can make proposals, but an agreement once implemented cannot be reversed. Loosely speaking, proposals will be made to coalitions — possibly the grand coalition — and once a proposal is unanimously accepted by the relevant coalition that coalition walks away from the game and enjoys the payoff from the proposal. This process continues until a full coalition structure — possibly including some standalone singletons — has formed, otherwise the process goes on forever.

I now describe proposals and responses. Suppose that some set of players has already exited the game. To each "remaining" set of "active" players is assigned a uniform probability distribution over initial proposers. Likewise, to each coalition of active agents *to* which a proposal has been made, there is a given order of respondents (excluding the proposer of course).

A chosen proposer makes a proposal $(S, \mathbf{y})$, where S is a coalition to which he belongs — a subset of the set of currently active players — and $\mathbf{y}$ is a division of the worth $v(S)$ of S.

Once a proposal is made to a coalition, attention shifts to the respondents in that coalition. A response is simply acceptance or rejection of the going proposal. If all respondents accept, the newly-formed coalition exits, and the process shifts to the set of still-active players remaining in the game.

The rejection of a proposal creates a bargaining friction, exactly as in the Rubinstein-Ståhl setup. Payoffs are delayed by the passage of some time, which is discounted by everybody using a discount factor δ. We suppose that the first rejector of the proposal gets to be the next proposer.

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 forever-active players receive a payoff of zero, and that already-formed coalitions received their agreed-upon allocation, discounted by the date at which they mannage to agree.

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

This completes the description of the basic model.

3.2. **Strategies and Equilibrium.** A *strategy* for a player requires her to make a proposal whenever it is her turn to do so, where the choice of proposal could depend on events that have already unfolded. It also requires her to accept or reject proposals at every stage in which she is required to respond. A *perfect equilibrium* is a profile of strategies such that there is no history at which a player benefits from a deviation from her prescribed strategy.

We already know that when $n \geq 3$, there is a plethora of perfect equilibria. This can be dealt with in more than one way. For the pruposes of these notes, we restrict ourselves to the use of stationary, Markovian strategies. These depend on a small set of payoff-relevant state variables. The current proposal or response is allowed to depend on the current set of active players and — in the case of a response — on the going proposal.

A *stationary* (perfect) *equilibrium* is then a collection of stationary strategies which forms a perfect equilibrium.

3.3. **Equilibrium Response Vectors.** let $x_i(S, \delta)$ be the equilibrium payoff to player i when i is the proposer and S is the set of active players. For each i, define

$$y_i(S, \delta) = \delta x_i(S, \delta).$$

Then, if j were to receive a proposal $(T, \mathbf{z})$ at player set S such that $z_k \geq y_k(S, \delta)$ for all $k \in T$ yet to respond, including j, then j should accept. On the other hand, if this inequality held for all such k *except for* j, then j should reject.[2]

We therefore use the terminology *equilibrium response vector* to describe the vector $\mathbf{m}(S, \delta)$. The following lemma is basic:

Lemma **1.** *For every* (S, δ) *and each* $i \in S$,

$$y_i(S, \delta) \geq \delta \max_{T : i \in T \subseteq S} \left[v(T) - \sum_{j \in T - i} y_j(S, \delta) \right]. \tag{7}$$

Proof. By making an offer $\mathbf{z}$ to any coalition T (with $i \in T$) such that $z_j > y_j(S, \delta)$ for every $j \in T$, $j \neq i$, i can guarantee acceptance of the proposal. It follows that

$$x_i(S, \delta) \geq \max_{T : i \in T \subseteq S} \left[v(T) - \sum_{j \in T - i} y_j(S, \delta) \right],$$

and using $y_i(S, \delta) = \delta x_i(S, \delta)$, we are done. □

Notice that if (7) indeed holds with strict inequality for some i, then i *must* be making an equilibrium proposal $(T, \mathbf{z}0$ such that $z_j < y_j(S, \delta)$ for some $j \in T$. Look at the last player in T's response order for which this inequality holds. That player must reject — assuming the proposal makes it that far — so the proposal cannot be acceptable.

Can this sort of "delay" happen in equilibrium? It can. Consider this example:

Delay Example. $N = \{1, 2, 3, 4\}$, $v(1, j) = 50$ for $j = 2, 3, 4$, $v(ij) = 100$ for $i, j = 2, 3, 4$, and $v(S) = 0$ for all other S.

We will provide a formal argument later but for now, note that if player 1 is called upon to propose he will always make an unacceptable offer

The intuition is simply this: player 1 is a "weak partner" and therefore his proposals will be turned down by the other players unless he makes one of them an offer equal to his outside

[2] The other possibilities have unclear implications at this stage. For instance, if $z_j < y_j(S, \delta)$ but it is also the case that $z_k < y_k(S, \delta)$ for a later respondent k, should j reject? Unclear. Indeed, it is unclear what j should do even if $z_j > y_j(S, \delta)$ in this case.

option in dealing with one of the other players. It may be better for player 1 to simply let one of these other partnerships form and *then* proceed to deal with the remaining player on more equal terms.

The point of this exercise is not that you should be taking the delay seriously. The delay is in part an artifact from the assumption that player 1 *must* make a proposal and is not allowed to simply pass the initiative to another player, or to make a proposal for some other coalition. In both these cases the "delay" will go away. But the point is that an inequality like (7) will still hold with strict inequality in such a case.

Say that an equilibrium is *no delay* if after *every* history, every proposer makes an acceptable proposal. Then, of course, (7) holds with equality for such equilibria. Let $\mathbf{m}(S, \delta)$ be the solution to the equality version of (7).

Theorem **2.** *For every* (S, δ), $\mathbf{m}(S, \delta)$ *exists and is unique.*

Proof. Existence is just an application of Brouwer's fixed point theorem; standard. [But don't neglect it, make sure you can do it.]

Uniqueness crucially depends on the following lemma. Roughly speaking, it states that if player i is making an acceptable proposal in equilibrium, then *anyone he includes in his best acceptable offer must enjoy a higher equilibrium response payoff.*

Lemma **2.** *Let* $\mathbf{y}(S, \delta)$ *be any equilibrium response vector, and suppose that* (7) *holds with equality for some* $i \in S$. *Then for any* T *that attains the maximum in* (7) *and for all* $j \in T$, $y_j(S, \delta) \geq y_i(S, \delta)$.

Proof. For i and T as described in the statement of the lemma, we have that

$$y_i(S, \delta) = \delta \left[v(T) - \sum_{k \in T - i} y_k(S, \delta) \right]. \tag{8}$$

while for $j \in T - i$,

$$y_j(S, \delta) \geq \delta \left[v(T) - \sum_{k \in T - j} y_k(S, \delta) \right]. \tag{9}$$

Adding $-\delta y_j(S, \delta)$ to both sides of (9) and using (8), we see that

$$(1 - \delta) y_j(S, \delta) \geq \delta \left[v(T) - \sum_{k \in T - j} y_k(S, \delta) \right] - \delta y_j(S, \delta) = (1 - \delta) y_i(S, \delta).$$

. □

Now return to the proof of the theorem. Suppose, on the contrary, that there are two solutions $\mathbf{m}$ and $\mathbf{m}'$ to the full equality version of (7). Define K to be the set of all indices in S in which the two solutions differ; i.e., $K \equiv \{i \in S | m_i \neq m'_i\}$ and pick an index k such that one of these m-values is maximal; wlog:

$$m_k = \max\{z | z = m_i \text{ or } m'_i \text{ for } i \in K\}.$$

By definition, $m_k > m'_k$. Choose $T \subseteq S$ such that

$$m_k = \delta \left[v(T) - \sum_{j \in T-k} m_j \right].$$

Of course,

$$m'_k \geq \delta \left[v(T) - \sum_{j \in T-k} m'_j \right].$$

By Lemma 2, $m_j \geq m_k$ for all $j \in T$. By our choice of $k \in K$, it must be the case that $m'_j \leq m_j$ for all $j \in T$. But then $m'_k \geq m_k$, which is a contradiction. □

The characterization of no-delay equilibria can be applied immediately to deduce that there must be delay in the "delay example" above. Simply calculate the **m**-vectors. For any two-player set with worth x, it is easy to see that

$$m_i(S, \delta) = \frac{\delta x}{1+\delta} \text{ for } i \in S,$$

while for the grand coalition N, it is easy to check that

$$\begin{aligned} m_2(N, \delta) &= m_3(N, \delta) = \frac{100\delta}{1+\delta} \\ m_1(N, \delta) &= \delta \left[50 - \frac{100\delta}{1+\delta} \right]. \end{aligned}$$

[You can verify all this by simply checking that the equality version of (7) is satisfied; after all, we already know that the solution is unique.]

Now $m_1(N, \delta)$, while always positive, converges to 0 as $\delta \to 1$. So this *cannot* be the equilibrium payoff for player 1. If he deviates by maling an unacceptable proposal, then simply use the assumed continuation along the no-delay equilibrium to conclude that player 1 can get a payoff bounded away from 0 (as $\delta \to 1$).

4. Immediate Agreement

The delay example suggests that the following condition is sufficient for no-delay:

[M] If $S \subseteq S'$, then for all $i \in S$, $m_i(S, \delta) \leq m_i(S', \delta)$.

I reiterate that it isn't the "no-delay" that we are after, as the delay is of a trivial sort anyway, but we want the ability to characterize equilibria using the **m**-vector. This condition does the trick.

Theorem **3.** *Under* [M], *equilibria can be fully described. At any set of active players S a proposer i chooses T to maximize*

$$v(T) - \sum_{j \in T-k} m_j(S, \delta)$$

and delivers $m_j(S,\delta)$ to each $j \in T$. Any responder j accepts if and only if $y_k \geq m_k(S,\delta)$ for k equal to j and all those after him.

Proof. First we show that the description in the theorem constitutes an equilibrium. Inductively, assume that it is true for all active player sets of cardinality $s-1$ or less (trivially true when $s=1$). Now pick a set S of cardinality s. Complete the verification, first for the proposer, and then for the responder, using condition [M].[3]

To show that nothing else can be an equilibrium, proceed inductively once again. Say the description is complete for all active player sets of cardinality $s-1$ or less (trivially true when $s=1$). Now pick a set S of cardinality s. Study any equilibrium response vector $\mathbf{y}$ on S. Let

$$K \equiv \{i \in S | m_i(S,\delta) \neq y_i\}$$

and pick an index i such that either m_i or y_i is the *biggest* of the values featured in K. If it is y_i, note that $y_i > m_i(S,\delta) \geq m_i(S',\delta)$ for all $S' \subseteq S$ (by condition [M]), so using the induction hypothesis, *i must be making an acceptable offer.* So pick T such that

$$y_i(S,\delta) = \delta \left[v(T) - \sum_{k \in T-i} y_k(S,\delta) \right], \tag{10}$$

and observe that

$$m_i(S,\delta) \geq \delta \left[v(T) - \sum_{k \in T-i} m_k(S,\delta) \right]. \tag{11}$$

By (10) and Lemma 2, $y_j \geq y_i$ for all $j \in T$. So $m_j(S,\delta) \leq y_j$ for all $j \in T$, $j \neq i$. But then using (11), we see that $m_i(S,\delta) \geq y_i$, which is a contradiction.

Alternatively, if at the index i in the "maximal set" K we have $m_i(S,\delta) > y_i$, then note that there exists T such that (10) holds with equality with m_i in place of y_i, and then follow the same argumemt as above with the roles of m_i and y_i interchanged. □

5. Are Equilibria Efficient?

Start with some intuition. Everyone is free to make or reject offers, so why should equilibria be inefficient anyway? But there is an implicit externality here: when someone makes an offer, he has to compensate the responders. Otherwise they can seize the initiative. This means that "at the margin" when a proposer is choosing a coalition, part of the surplus from that coalition has to be "given away". This drives a wedge between the "private surplus" and the "social surplus" and may cause an inefficient choice to be made.

To see this even more clearly, consider the dictator version of this game in which only one player gets to make offers and everyone else can only say yes or no. In that case the outcome is efficient in all equilibria because the entire social surplus is appropriated by the dictator

[3] There is a relatively unimportant technical matter here. When equality holds for the responder he should be free to go either way, rather than accept as we assert he must. But it can be shown that "accept" is the only decision that will work, otherwise the proposer's best response is not well-defined.

who therefore maximizes that surplus. The outcome may not be very welcome from an equity point of view but that is another matter.

Let us begin the formal analysis with the strongest possible notion of efficiency: no matter *who* begins the game, the outcome is efficient. We will call this notion *strong efficiency.*

THEOREM **4.** *A coalitional bargaining game with a strictly superadditive characteristic function is strongly efficient for all discount factors close to 1 if and only if*

$$\frac{v(N)}{|N|} \geq \frac{v(S)}{|S|} \text{ for all } S. \tag{12}$$

Proof. Our proof will reply on the following lemma:

LEMMA **3.** *Fix an equilibrium. For any* (S,δ)*, suppose that* (7) *holds with strict inequality for some* i*:*

$$y_i(S,\delta) > \delta \max_{T:i\in T\subseteq S}\left[v(T) - \sum_{j\in T-i} y_j(S,\delta)\right].$$

Then there exists a strict subset S' *of* S *such that* $y_i(S',\delta) \geq y_i(S,\delta)$*.*

Proof. Let S^* be some *minimal* subset of S (which could be S itself) such that $y_i^*(S,\delta) \geq y_i(S,\delta)$. Then I claim that

$$y_i(S^*,\delta) = \delta \max_{T:i\in T\subseteq S^*}\left[v(T) - \sum_{j\in T-i} y_j(S^*,\delta)\right].$$

For if not, then i must make an unacceptable proposal at S^*. But after that he can get *at most* $y_i(S',\delta)$ for some $S' \subset S^*$. But by construction, $y_i(S',\delta) < y_i(S^*,\delta)$, which is a contradiction. So equality does hold, which means from our premise that S^* must be a strict subset of S. □

Now return to the main proof. First we show that (12) implies strong efficiency. Pick any person i and look at $y_i(N,\delta)$. Then, using Lemma 3, there exists S (could be N itself) and $T \subseteq S$ with $i \in T$ such that

$$y_i(N,\delta) \leq y_i(S,\delta) = \delta\left[v(T) - \sum_{j\in T-i} y_j(S,\delta)\right]. \tag{13}$$

We know from Lemma 2 that $y_j(S,\delta) \geq y_i(S,\delta)$ for all $j \in T$, so using (12),

$$y_i(S,\delta) \leq \frac{\delta v(T)}{1+\delta(t-1)} \leq \frac{\delta v(N)}{1+\delta(n-1)} \tag{14}$$

where t and n are the cardinalities of T and N respectively.

Note that *strict* inequality must hold in the second inequality of (14) whenever $T \neq N$ (why?).

So, combining (13) and (14), we have shown that

$$y_i(N,\delta) \leq \frac{\delta v(N)}{1+\delta(n-1)} \tag{15}$$

for all i, with strict inequality whenever i does not propose acceptably to the grand coalition.

At the same time, we know that

$$\begin{aligned} y_i(N,\delta) &\geq \delta\left[v(N) - \sum_{j\in N-i} y_j(N,\delta)\right] && (16)\\ &\geq \delta\left[v(N) - \frac{\delta v(N)(n-1)}{1+\delta(n-1)}\right] && (17)\\ &= \frac{\delta v(N)}{1+\delta(n-1)}, && (18)\end{aligned}$$

where the second line uses (15). Together, (15) and (18) prove that i must make an acceptable offer to the grand coalition, which proves efficiency.

Conversely, suppose that we have strong efficiency for all discount factors close to 1. Then each starting proposer makes a proposal to the grand coalition, so that $y_i(N,\delta)$ is some constant y for all i, *and*

$$y_i(N,\delta) = \delta\left[v(N) - \sum_{j\in N-i} y_j(N,\delta)\right],$$

so that

$$y = \frac{\delta v(N)}{1+\delta(n-1)}. \tag{19}$$

Moreover,

$$y_i(N,\delta) \geq \delta\left[v(S) - \sum_{j\in S-i} y_j(N,\delta)\right]$$

for all S, so that

$$y \geq \frac{\delta v(S)}{1+\delta(s-1)}. \tag{20}$$

Combine (19) and (20), and send δ to 1. □

Notice that the condition above is closely related to the condition for a nonempty core in symmetric games. It's still sufficient, but not necessary, in general games. In any case, we see that we are getting a theory that tends to generate inefficiency for empty cores *but still gives us a prediction using the* $\mathbf{m}$*-vector*. Indeed, as we shall see, it might even give inefficiency when the core is nonempty. Let us go on.

What if, instead of insisting on strong efficiency, we simply ask for efficiency for *some* initial proposer? Call this *weak efficiency.*

Theorem **5.** *Suppose that (N, v) is a strictly superadditive characteristic function and suppose that we have weak efficiency for some sequence of discount factors converging to one. Let $\mathbf{z}(\delta)$ be the corresponding sequence of efficient equilibrium payoff vectors. Then any limit point of $\mathbf{z}(\delta)$ lies in the core of (N, v).*

Proof. Suppose without loss of generality that player 1 is the "efficient proposer" along the sequence.[4] By strict superadditivity, he must be making a proposal to the grand coalition. Then, if $\mathbf{y}(N, \delta)$ is the equilibrium response vector, we have

$$z_1(\delta) = \frac{y_1(N, \delta)}{\delta}, \tag{21}$$

while

$$z_j(\delta) = y_j(N, \delta) \tag{22}$$

for all other j. Now pick any coalition S and any $i \in S$. By (7),

$$y_i(N, \delta) \geq \delta \left[v(S) - \sum_{j \in S-i} y_j(N, \delta) \right],$$

or

$$\frac{y_i(N, \delta)}{\delta} + \sum_{j \in S-i} y_j(N, \delta) \geq v(S).$$

Pick a subsequence such that $\mathbf{z}(\delta)$ converges, say to $\mathbf{z}^*$. Send δ to 1 along this subsequence and note that both $y_i(N, \delta)$ and $y_i(N, \delta)/\delta$ converge to the z_i^*. Therefore

$$\sum_{i \in S} z_i^* \geq v(S).$$

Because S was arbitrarily chosen, we are done. □

So now the connection with the core starts to become even clearer. For discount factors close to 1, games with empty cores will *never* have efficient stationary equilibria no matter who proposes first.

What about the converse? If equilibria are inefficient, must the core be empty? Consider the following example:

Employer-Employee Game. $N = \{1, 2, 3\}$, $v(i) \simeq 0$ for all i, $v(23) \simeq 0$, $v(12) = v(13) = 1$, $v(N) = 1 + \mu$ for some $\mu > 0$. The interpretation is that player 1 is an employer who can produce an output of 1 with any one of the two employees 2 and 3. He can also hire both employees in which case output is higher. No other combination can produce anything.

Exercise. Show that if $\mu \in (0, 0.5)$, then the equilibrium always involves the coalition $\{12\}$ or $\{13\}$, with the remaining player left "unemployed". Yet observe that the core of this game is empty as long as $\mu > 0$.

[4] Of course the efficient proposer may change along the sequence, but then simply take an appropriate subsequence.

So an empty core will generate inefficiency for sure, but a nonempty core won't guarantee efficiency! Recall the intuition at the start of this section to understand why. Player 1 would love to include both players 2 and 3 and pocket the resulting surplus, but the *very act* of including both players gives them bargaining power (they can reject) and so both players will have to be compensated, at rates that do not justify the gain of the extra μ. [The rates will be justified if $\mu > 0.5$.]

To try and characterize inefficiency, or at least to find a weaker sufficient condition for it, we have to dig deeper. One approach is to try and characterize the limit of the $\mathbf{m}$-vectors as δ goes to 1. Ideally, we would like to do so in a way that only depends on the parameters of the model, and so that we can algorithmically calculate it. The following observation takes the first step.

THEOREM **6.** *There exists a unique vector* $\mathbf{m}^*$ *with the property that for each* i*,*

$$m_i^* = \max_{i \in S \subseteq N} \left[v(S) - \sum_{j \in S - i} m_j^* \right], \tag{23}$$

and the property that $m_j^* \geq m_i^*$ *for all* $j \in S$*, for some* S *that attains the maximum above.*

Proof. To establish existence, recall that $\mathbf{m}^*(N, \delta)$ is well-defined for every δ, by Theorem 2. Pass to any limit point as δ goes to 1; call it $\mathbf{m}^*$. Lemma 2 and a trivial continuity argument assures us that $\mathbf{m}^*$ has both the properties claimed in the statement of the theorem.

The uniqueness of $\mathbf{m}^*$ is obtained by following exactly the proof of Theorem 2 (without any need to invoke Lemma 2).

Suppose, on the contrary, that there are two solutions $\mathbf{m}$ and $\mathbf{m}'$ meeting the conditions of the theorem. Define K to be the set of all indices in N in which the two solutions differ; i.e., $K \equiv \{i \in N | m_i \neq m_i'\}$ and pick an index k such that one of these m-values is maximal; wlog:

$$m_k = \max\{z | z = m_i \text{ or } m_i' \text{ for } i \in K\}.$$

By definition, $m_k > m_k'$. Choose S such that

$$m_k = \delta \left[v(S) - \sum_{j \in S - k} m_j \right].$$

and with $m_j \geq m_k$ for all $j \in S$. We know that

$$m_k' \geq \delta \left[v(S) - \sum_{j \in S - k} m_j' \right].$$

Because $m_j \geq m_k$ for all $j \in S$, it follows from our choice of k that $m_j' \leq m_j$ for all $j \in S$. But then $m_k' \geq m_k$, which is a contradiction. □

The following result is an immediate corollary:

THEOREM **7.** *If* $\sum_{i=1}^n m_i^* > v(N)$*, then no sequence of equilibria can be weakly efficient for all discount factors approaching 1.*

We end, then, by describing an algorithmic way to calculate $\mathbf{m}^*$. The idea is simply to exhibit a vector that satisfies all the conditions of Theorem 6. The uniqueness result there assures us that we then have the correct vector.

Algorithm.

Step 1. Begin by maximizing $v(C)/|C|$. Let C_1 be the *union* of all maximizers of this expression, and let a_1 be the maximum value attained. Define $m_i^* \equiv a_1$ for every $i \in C_1$.

Step 2. Recursively, suppose that we have defined sets $\{C_1, \ldots, C_K\}$ for some index $K \geq 1$, corresponding values $\{a_1, \ldots, a_K\}$, and have defined $m_i^* = a_k$ whenever $i \in C_k$. Now define W to be the union of all the C_k's, and consider the problem of choosing sets $C \subset N - W$ and $T \subseteq W$ to maximize

$$\frac{v(C \cup T) - \sum_{i \in T} m_i^*}{|C|}.$$

Define C_{K+1} to be the union of all sets C such that (C, T) maximizes this expression for some $T \subseteq W$, and let a_K be the maximum value attained. Define $m_i^* \equiv a_{K+1}$ for every $i \in C_{K+1}$.

Continue in this way until $\mathbf{m}^*$ is fully defined on the set N of all players.

Theorem **8.** *The vector* $\mathbf{m}^*$ *as constructed in the algorithm satisfies all the conditions of Theorem 6.*

Proof. Fix $\mathbf{m}^*$ as given by the algorithm. Consider any player i, and the problem of choosing S — with $i \in S$ — to maximize

$$v(S) - \sum_{j \in S - i} m_j^*.$$

Suppose that $i \in C_k$, given by the algorithm. Then the following is true:

Claim. If $S \subset C_1 \cup \cdots \cup C_k$, then

$$m_i^* \geq v(S) - \sum_{j \in S - i} m_j^*, \tag{24}$$

with equality holding when $S = C \cup T$, where (C, T) is an algorithmic maximizer in Step 2 (at stage k).

To prove the claim, note that if S is of the form $C \cup T$, where $C \subseteq C_k$ and $T \subseteq C_1 \cup \cdots \cup C_{k-1}$,[5] then by Step 2 of the algorithm,

$$m_i^* \geq \frac{v(S) - \sum_{i \in T} m_i^*}{|C|} = v(S) - \sum_{j \in S - i} m_j^*.$$

Moreover, equality must hold when (C, T) is an algorithmic maximizer in Step 2 (at stage k). Thus proves the Claim.

[5] If $k = 1$, then treat $C_1 \cup \cdots \cup C_{k-1}$ as the empty set.

We actually want to show that (24) holds for *every* set S, not just S of the form described in the Claim. Suppose this is false. Then there is S such that $i \in S$ and

$$v(S) - \sum_{j \in S-i} m_j^* > m_i^*. \tag{25}$$

Pick $j \in S$ such that j belongs to the "highest" set index in the algorithm, say C_ℓ. Then, using (25) and the Claim, it must be that $\ell > k$. Rearranging terms in (25), we see that

$$m_j^* < v(S) - \sum_{k \in S-j} m_k^*,$$

but this violates the Claim for individual j!

To finish off the proof, for each i (say in C_k) pick $S = C \cup T$, where (C, T) is the algorithmic maximizer in Step 2 (at stage k). Then (24) holds with equality (by the Claim), and $m_j^* \geq m_i^*$ for every $j \in S$. □