Econometrica, Vol. 57, No. 1 (January, 1989), 81–95

BARGAINING FOUNDATIONS OF SHAPLEY VALUE[1]

BY FARUK GUL

A transferable utility economy in which each agent holds a resource which can be used in combination with the resources of other agents to generate value (according to the characteristic function V) is studied using a dynamic model of bargaining. The main theorem establishes that the payoffs associated with efficient equilibria converge to the agents' Shapley values as the time between periods of the dynamic game goes to zero. In addition it is demonstrated that an efficient equilibrium exists and is unique when an additivity condition is satisfied. To demonstrate the sensitivity of the solution to the institutional detail we modify the model to allow for partnerships and show that the Shapley value is no longer achieved.

KEYWORDS: Noncooperative bargaining, stationary subgame perfect Nash equilibrium, games in characteristic function form, coalition formation, random matching.

1. INTRODUCTION

THE COOPERATIVE APPROACH to the bargaining problem has been criticized for not having strategic foundations. Nash (1953) himself viewed his bargaining solution as a tentative step which needed to be supported within a noncooperative framework. The program of establishing noncooperative foundations for cooperative solution concepts has also been pursued by Binmore (1980), (1982), (1983), Binmore, Rubinstein, and Wolinsky (1985), Harsanyi and Selten (1980), and Herrero (1985). The noncooperative approach, on the other hand, has also received severe criticism. The choice of the particular (extensive form) game, the choice of the equilibrium concept, and the multiplicity of equilibria have often been the source of controversy.

Our purpose is to study the relationship between the cooperative and noncooperative approaches by establishing a framework in which the results of the two theories can be compared.

In Section 2, we study an economy in which the underlying opportunities for gains from trade are determined by a characteristic function. Within this economy, we examine a natural extension of the discrete time extensive form games used in various existing models on noncooperative bargaining. We use the equilibrium concept stationary subgame perfect Nash equilibrium (SSPNE)[2] and show that the set of equilibria is finite. Of these equilibria, we focus on a particularly attractive one, the unique efficient equilibrium. This equilibrium, when it exists, is the only one in which trade (or agreement) occurs at each

[1] This paper constitutes the first chapter of my Ph.D. dissertation. The problem of analyzing bargaining games in characteristic function form was suggested to me by my advisor, Hugo Sonnenschein. I am also grateful to him for many other comments and criticisms including the observation that the Shapley value equilibrium discussed in Section 2 is Pareto-efficient. I would also like to thank Ken Binmore, Roger Myerson, Barry Nalebuff, Ariel Rubinstein, and Larry Samuelson for helpful discussions.

[2] Our stationarity requirement is a direct application to our framework of the stationarity defined in Rubinstein and Wolinsky (1985). It is equivalent to the familiar condition that strategies depend only on payoff relevant histories. See also Binmore (1985) for a discussion of the relationship between stationarity and continuity.

meeting of the agents. We show that, if time intervals are arbitrarily small, the (expected) utility of each agent in this equilibrium is his Shapley value. We also provide a necessary and sufficient condition on the characteristic function for the existence of such an equilibrium.

In Section 3, we modify the extensive form game of Section 2 so that agents make offers of partnership rather than sales. We show that, if the number of agents is three, there is a unique equilibrium for the partnership game. Furthermore, we prove that the limit of the equilibrium expected utilities of each agent is equal to the Shapley value computed from a transformation of the characteristic function, rather than the ordinary Shapley value. Hence we conclude that our results are sensitive to the details of the extensive form structure.

Section 4 summarizes our conclusions. In particular, we discuss the role of the institutional structure and the nature of the characteristic function in determining the applicability of the Shapley value as a reasonable prediction of behavior in a noncooperative setting.

2. THE BARGAINING GAME

Consider a transferable utility game in characteristic function form $G=(V, N)$ where $N=\{1,2,3,\ldots,n\}$, $V: 2^N \to IR_+(V(\phi)=0)$. For our purposes, V has the following interpretation: Each agent i owns a valuable resource, and various combinations of these resources (called resource bundles) produce utility according to the function V. Hence, the resources of agents $M \subset N$, when pooled together, produce a flow of utility which has a discounted present value of $V(M)$. Agents can buy and sell these resources in exchange for payment of utils. When agent i sells his resource to j, he leaves the market. After such a transaction, we will sometimes denote j as $\{i, j\}$ to emphasize the fact that he now owns the initial resources of both i and j.

Exchange takes place within the framework of the following extensive-form game: At each period $t=0,1,2,\ldots$, a random meeting occurs, say between players i and j, with probability $(2/n_t(n_t-1))$, where n_t is the number of agents still in the game at time t. Hence, each possible meeting is equally likely. Next, one of the two parties, i and j, is chosen randomly (with probability $\frac{1}{2}$) to make an offer $r_t \in IR_+$. An offer r_t is an offer of utility (or a numeraire good which gives the same level of utility to each agent). Say i is chosen and offers r_t to j; then one of the following occurs: (i) j accepts the offer, which means that he sells his resource to i, and leaves the market; (ii) j rejects the offer, and the meeting dissolves.

In either case, the next period begins with a new meeting. It should be noted that the probability of a meeting between two players in period t, given that they are still in the game, is independent of whether either or both of them participated in any previous meetings.

The game continues as long as there are two or more agents still in the market. This particular extensive form aims at mimicking the bargaining process in markets with very little friction and communication costs. Hence, all agents are constantly making offers, accepting or rejecting offers, and looking for new bargaining partners. To capture this, we assume all future meetings are equally likely and focus on the case where the time between offers is small. We take the point of view that the resources produce streams of utility, hence, each agent derives $(1-\delta)V(M)$ from holding the initial resources of the agents in M for one period, where δ denotes the common per period discount factor. Therefore, the utility of agent i associated with a given outcome of this game is

$$U^i = \sum_{t=0}^{\infty} \left[(1-\delta)V(M_t^i) - r_t^i\right]\delta^t,$$

where $M_t^i \subset N$ is the set of all agents whose initial resources are owned by i at the end of period t, and r_t^i is the payment made by agent i in period t. A payment to an agent is considered a negative payment by i. (Henceforth, we will simply say resource M whenever we mean the sum of the initial resources of the agents in M.)

Since this is a game of perfect information, a tth period strategy for agent i, $\sigma_t^i \in \Sigma_t^i$, is a mapping from each $t-1$ period history $h_{t-1} \in H_{t-1}$ and each possible meeting involving i, to a pair of real numbers (a, b) with the following interpretation: $\sigma_t^i(h_{t-1}, j) = (a, b)$ means that, if after history h_{t-1} (which includes up to time $t-1$, a full record of all the events in the game) i meets j, he will offer j, a if i is chosen to make the offer and accept any offer greater than or equal to b if j is chosen to make the offer.[3]

We define a strategy for i, $\sigma^i \in \Sigma^i$, by $\sigma^i = (\sigma_t^i)_{t=0}^{\infty}$, where $\sigma_t^i \in \Sigma_t^i$ for all t. Therefore, $\Sigma = \Sigma^1 \times \Sigma^2 \times \ldots \times \Sigma^n$ is the set of strategy profile σ.

The noncooperative extensive form game $B = \{(U^i, \Sigma^i)_{i=0}^{n}\}$ is called a Bargaining Game.

Our notion of equilibrium is stationary subgame perfect Nash equilibrium (SSPNE). A strategy profile σ is a SSPNE if and only if: (i) The strategy of each agent i is optimal after every history, given the strategies of all other agents. (This is the standard notion of subgame perfection due to Selten (1975).) (ii) σ is stationary—that is, if after two histories h_t and h_r, (a) player i owns the same resources; (b) the distribution of the remaining resources is the same; specifically, for all $M \subset N$, if there exists an agent who owns exactly M at time t, then there exists some agent (possibly a different one) who owns exactly M at time r, implies $\sigma_t^i(h_{t-1}, M) = \sigma_r^i(h_{r-1}, M)$. Hence, i's strategy depends only on what he owns and how the remaining resources are partitioned.

Notice that our stationarity assumption does not require that if, after a history h_t, i buys j's resource, then he follows the same strategy for the remainder of the game as j would if j had purchased i's resource. However, the strategies of all agents other than i and j must be independent of whether i or j ended up with the combined resources of i and j.

[3] Allowing agents to accept or reject arbitrary sets of offers, rather than just closed intervals would not alter any of our results. We restrict attention to interval strategies only for notational convenience.

One story behind the notion of Nash equilibrium is that people know the strategies of opposing players because they have previously observed third parties playing the same game. It is inconceivable that such a model of learning will enable the communication of arbitrarily complicated strategies in infinite horizon games.

Hence, if Nash equilibria are to be viewed, in some sense, as the limits of a dynamic adjustment process, then some restriction on the strategies of the agents seems to be desirable. Stationarity enables us to concentrate on a particularly simple class of strategies.

We will use E to denote the set of all SSPNE. Henceforth we will mean a SSPNE whenever we say equilibrium. Sometimes we will index our game and equilibria by δ, that is, use $B(\delta)$ and $E(\delta)$.

The state $q \in Q$, $q = (M_1, M_2, \ldots, M_k)$ refers to a situation in which, after various rounds of trading, there are k players left in the game and each player (who is still in the game) i owns resources $M_{k(i)} \subset N$. Q is the set of all possible states (partitions of N). Given our weak stationarity requirement, a state $q \subset Q$ does not summarize all of the relevant history. Since we have not required that i and j have the same strategy if they are faced with the same situation, the identity of the owners of the various combinations of resources could conceivably influence the outcome. Obviously Theorems 1 and 2 will hold even with the stronger stationarity requirement.

$\bar{N} \in Q$ denotes the finest partition of N. For all $L, M \in q$, let $R(q, L, M) \in Q$ denote the partition obtained from q by replacing the elements L and M by $L \cup M$. That is, $R(q, L, M) = (q \setminus \{L, M\}) \cup \{L \cup M\}$.

Let $F \equiv \{f_q | f_q : q \to N \text{ for some } q \in Q\}$. Hence, the pair q, f_q summarizes all of the relevant history; q determines how the resources are distributed and f_q determines the identity of the owners.

In what follows, we will assume that marginal product of agents' resources are always positive. Hence, V is *strictly* super-additive.

Strict super-additivity:[4]

$$(SS): \quad L \cap M = \phi, \quad L \neq \phi \neq M \quad \text{implies} \quad V(L) + V(M) < V(L \cup M)$$
$$\text{for all } L, M \subset N.$$

$U(M, q, f_q, \sigma)$ denotes the expected utility of the continuation of the game for player $f_q(M)$, according to strategy profile σ given state q. Note that $U(M, q, f_q, \sigma)$ does not reflect payments made or utilities enjoyed before the occurrence of q. When the relevant f_q is obvious (for example, at the start of the game) or if the payoffs of the agents can be shown to be independent of their identities (as is the case in the efficient equilibrium), we omit f_q and write $U(M, q, \sigma)$ instead of $U(M, q, f_q, \sigma)$. We also write $U(i, q, \sigma)$ rather than $U(\{i\}, q, \sigma)$.

[4] Larry Samuelson has pointed out that all of our conclusions would remain valid if we substituted $A \subseteq N$ and $A \neq N$ implies $V(A) < V(N)$ and super-additivity for strict super-additivity. For this case we would need to make a minor change in the proof of Theorem 2.

Let

$$S(M,q)=\sum_{A\subset q}\frac{(\mu(A)-1)!\,(\mu(N)-\mu(A))!}{\mu(N)!}\cdot\left[\bar{V}(A)-\bar{V}(A\setminus\{M\})\right],$$

where $\mu(A)$ is the number of elements of the partition q which are included in A and $\bar{V}(X)=V(\bigcup_{\alpha\in X}\alpha)$ for all $X\in q$. Of course $S(M,q)$ is only defined for $M\in q$.

Observe that $S(i,\bar{N})$ is the Shapley value of $i\in N$ in the game G.[5] In general, $S(M,q)$ is the Shapley value of M given state q. That is, $S(M,q)$ is the Shapley value of the player who owns the resource bundle M in a game in which initial endowments are distributed according to q. Since S enables us to compute Shapley values with respect to arbitrary partitions of N rather than just the finest partition, we call the function S the generalized Shapley value.

Our main result deals with the efficient strategy profiles $\sigma\in\Sigma$ in the bargaining game B. Since σ determines a stochastic outcome path, the appropriate definition of efficiency involves the use of expected utilities.

DEFINITION 1: $\sigma\in\Sigma$ is efficient if and only if

$$\sum_{i\in N}U(i,\bar{N},\sigma)\geq\sum_{i\in N}U(i,\bar{N},\sigma')\quad\text{for all }\sigma'\in\Sigma.$$

Given (SS) it is clear that an equilibrium is efficient if and only if it prescribes every possible meeting to end in agreement.

THEOREM 1: *Let $\sigma(\delta_k)\in E(\delta_k)$ for some $\{\delta_k\}_{k=0}^{\infty}$ such that $\lim_{k\to\infty}\delta_k=1$. If $\sigma(\delta_k)$ is efficient for all k, then $\lim_{k\to\infty}U(i,\bar{N},\sigma(\delta_k))=S(i,\bar{N})$.*

PROOF: See Appendix.

Hence, Theorem 1 states that, as the time interval between meetings becomes arbitrarily small, the expected payoff of each player at an efficient equilibrium converges to his Shapley value. Furthermore, it is clear from the proof of Theorem 1 that there can be at most one efficient equilibrium in $E(\delta)$. That equilibrium will involve a trade at each meeting. Observe that our proof also shows that in an efficient equilibrium, for δ close to 1, the expected utility of continuation for an agent who owns M at state q is approximately given by $S(M,q)$. But if a trade is to take place at each meeting, it must be the case that an agent who owns M at state q must offer to an agent who owns L approximately $S(L,q)$ and benefit from his acceptance; that is,

$$S(L\cup M,R(q,L,M))-S(L,q)\geq S(M,q).$$

[5] Hence, $S(i,\bar{N})=\sum_{M\subset N}((|M|-1)!(|N|-|M|)!/|N|!)(V(M)-v(M\setminus\{i\}))$ where $|N|$ and $|M|$ denote the cardinalities of the set N and M respectively.

Therefore, it is clear that the following condition, which we call value additivity of V, (VA), is necessary for an efficient equilibrium to exist for δ close to 1:

$$(VA) \qquad q \in Q, \quad L, M \in q, \quad L \neq M \quad \text{implies} \\ S(L,q) + S(M,q) \leq S(L \cup M, R(q,L,M)).$$

It is interesting to note that, if $N = \{1,2,3\}$ and V is normalized, then (VA) is equivalent to the following simple condition:

$$(VA') \qquad V(\{1,2\}) + V(\{1,3\}) + V(\{2,3\}) \geq 1.$$

Theorem 2 establishes that (VA) is also a sufficient condition for the existence of an efficient equilibrium.

THEOREM 2: *If V satisfies (VA), then for δ sufficiently close to* 1, *there exists a unique efficient equilibrium.*

PROOF: See Appendix.

From Theorem 2, it can be noted that all equilibria of the game can be computed as follows:

First, we postulate which meetings will result in trade at each state. Then we solve, using backward induction as described in the proof of Theorem 1, for the equilibrium (expected) payoffs for each agent, given the postulated set of trades. Finally, we check whether or not this solution corresponds to an equilibrium. More specifically, the solution corresponds to an equilibrium if and only if, at each state q and each meeting $\{L, M\} \subset q$, the continuation payoff of owning $L \cup M$ is no less (no greater) than the sum of the computed payoffs of the agents who own L and M at state q whenever a trade is (is not) postulated between L and M at state q. (Note that this condition corresponds to (VA) for the case of efficient equilibria and δ close to 1.)

We now provide an example of a game which has an inefficient equilibrium:

Let $N = \{1,2,3\}$ and $V(A) = 0$ for $A \subset N, A \neq N$, and $V(N) = 1$. Furthermore, let $a(i, j)$ denote the offer of agent i to agent j at state $\overline{N}$ and $r(j, i, a)$ denote the response of agent j to an offer of a by agent i at state $\overline{N}$.

Consider the strategy profile σ described below:

$$a(1,2) = a(2,1) = \frac{\delta}{4(3-2\delta)}$$

and

$$r(2,1,a) = r(1,2,a) = \text{Yes if } a \geq a(1,2), \\ = \text{No otherwise,}$$

$$a(1,3) = a(2,3) = 0$$

and

$$r(3,1,a) = r(3,2,a) = \text{Yes if } a \geq \frac{\delta^2}{2(3-2\delta)},$$
$$= \text{No otherwise},$$
$$a(3,1) = a(3,2) = 0$$

and

$$r(2,3,a) = r(1,3,a) = \text{Yes if } a > \frac{\delta^2}{4(3-2\delta)},$$
$$= \text{No otherwise}.$$

Note that:

$$U^1(\sigma) = \tfrac{2}{3}\delta U^1(\sigma) + \tfrac{1}{6}\left[\delta/2 - \delta U^2(\sigma)\right] + \tfrac{1}{6}\delta U^1(\sigma) = \frac{\delta}{4(3-2\delta)},$$
$$U^2(\sigma) = U^1(\sigma),$$
$$U^3(\sigma) = \tfrac{2}{3}\delta U^3(\sigma) + \tfrac{1}{3}\delta\tfrac{1}{2} = \frac{\delta}{2(3-2\delta)}.$$

It is easy to verify that σ is a SSPNE.

The equilibrium has the following structure: If player 3 meets 1 or 2 before 1 and 2 have traded between themselves, the meeting results in disagreement. If 1 and 2 meet they trade. Note that the equilibrium payoffs converge to $\frac{1}{4}$, $\frac{1}{4}$, $\frac{1}{2}$ respectively, as δ approaches 1. Also note that since (VA) fails, this game has no efficient equilibrium. It is easy to show that with 3 players no inefficient equilibrium exists when (VA) is satisfied. It is not known whether this is true in general, but it seems unlikely that (VA) rules out all possible inefficient equilibria when the number of players is large.

We have taken the point of view that the underlying (cooperative) possibilities implied by G reflect the possibilities for gains from trade. Hence, our extensive form game has led us to a theory of bargaining. Alternatively, we can interpret the outcomes of meetings as the formation of coalitions rather than trades (i.e., during the meeting agents buy and sell services, binding contracts, and/or voting rights rather than private goods); in this case our model provides a simple, noncooperative model of coalition formation. One of the main topics of discussion in cooperative game theory has been the formation of stable coalition structures. Our approach yields the following conclusion:

A coalition structure (set of postulates about which coalitions will form) is "stable" if the payoffs computed using that structure constitute an equilibrium of the associated bargaining game. In cases where (VA) is satisfied we find that a structure that prescribes the merger of any two coalitions whenever the opportunity of such a merger arises, is indeed stable.

3. THE PARTNERSHIP GAME

One possible criticism of the coalition formation interpretation of the bargaining game discussed in Section 2 is the following:

If the result of an accepted offer at a meeting is the formation of a coalition (rather than the realization of a trade, as in the bargaining interpretation) no one leaves the market after an accepted offer. Hence, is it not reasonable to expect coalitions with a large number of agents to have greater bargaining power than coalitions with a small number of agents? If one believes that large coalitions can commit to longer delays in case of disagreement or that it is somehow more difficult to reach an agreement with large coalitions, then the answer to previous questions must be yes. The following game, which we call the partnership game, is a modification of the bargaining game and gives larger coalitions more power in the bargaining process. It is not meant to support another solution concept but rather to demonstrate the nonrobustness of the game to the choice of the extensive form game.

Let $G = (V, N)$ be a game in characteristic function form as in Section I. Furthermore, let $N = \{1, 2, 3\}$. We define a Partnership Game, $P(\delta)$ for $\delta \in (0, 1)$ as follows:

In period 0, a random meeting between two agents takes place. Every possible meeting has probability $1/3$. Let us assume, without loss of generality, that a meeting occurred between 1 and 2. Then each player is chosen with probability $1/2$ to make an offer. Let's also assume that agent 1 is chosen. Agent 1 will make an offer $\alpha \in [0, 1]$. If 2 rejects this offer, then the meeting dissolves and a new (random) meeting takes place, and the process is repeated. If 2 accepts the offer α, it means that 1 and 2 agree to pool their resources and split whatever utility they obtain from continuation of the game so that 1 and 2 get respectively $1 - \alpha$ and α proportion of the total utility.[6] In the next round the partners $\{1, 2\}$ meet 3. Each player is chosen with probability $1/3$ to make an offer.

Say 3 is chosen and makes an offer $\beta \in [0, 1]$. If agents 1 and 2 both accept this offer,[7] then agents 1, 2, and 3 obtain utilities

$$\delta(1 - \alpha)\beta V(N) + (1 - \delta)(1 - \alpha)V(\{1,2\}),$$

$$\delta\alpha\beta V(N) + (1 - \delta)\alpha V(\{1,2\}),$$

and

$$\delta(1 - \beta)V(N),$$

respectively. If either 1 or 2 rejects the offer, a new offerer is chosen with probability $1/3$, and he makes a new offer. Notice that after a rejected offer, the partnership between 1 and 2 does not dissolve.

[6] Permitting more general types of offers which specify shares and transfers does not change the result. This is clear from the Proof of Theorem 3.

[7] We assume that players 1 and 2 announce their acceptance or rejection according to a prespecified order and *not* simultaneously. Hence subgame perfection will rule out equilibria which have 1 rejecting only because 2 is rejecting and 2 rejecting only because 1 is rejecting.

Let $\bar{E}(\delta)$ denote the set of SSPNE of the partnership game.

THEOREM 3: (i) *$\bar{E}(\delta)$ is a singleton for all $\delta \in (0,1)$.* (ii) *If $U(\delta) \equiv (U^1(\delta), U^2(\delta), U^3(\delta))$ are the utilities associated with $\sigma \in \bar{E}(\delta)$, then*

$$\lim_{\delta \to 1} U^i(\delta) = \tfrac{1}{3}V(N) + \tfrac{1}{9} \sum_{\substack{j \in N \\ j \neq i}} (V(N \setminus \{j\}) - 2V(j)) + \tfrac{4}{9}V(i) - \tfrac{2}{9}V(N \setminus \{i\}).$$

PROOF: See Appendix.

Computing the equilibrium for the partnership game if the number of agents is greater than 3 is very complicated. However, it is clear that a condition essentially weaker than (VA) will suffice to guarantee the existence of an efficient equilibrium for the general case.

The *complement* of a characteristic function is a function $\bar{V}: 2^N \to IR_+$ defined by

$$\bar{V}(M) = V(N) - V(N \setminus M) \quad \text{for all} \quad M \subset N.$$

Charnes, Rausseau, and Sieford (1978) define a *homomollifier* $h: 2^N \to IR$ as follows:

$$h(M) = \frac{|M|}{|N|}\bar{V}(M) + \frac{|N| - |M|}{|N|}V(M)$$

where $|M|$ denotes the cardinality for all $M \subset N$ of the set M. It is easy to verify that $\lim_{\delta \to 1} U^i(\sigma)$ is equal to the composition of Shapley value and the homomollifier. That is, the limits of the equilibrium utilities of the partnership game for the case of $N = \{1, 2, 3\}$ correspond to the Shapley values of the homomollifier. In cooperative game theory, the homomollifier reflects an attempt to incorporate the size of coalitions into characteristic function representation. Theorem 3 provides a similar noncooperative interpretation for the homomollifier.

4. CONCLUSION

Our investigation of the extensive form bargaining models described in Sections 2 and 3 have enabled us to isolate the two main sources of impatience in a multiperson bargaining problem. The first source is well known from the studies on the two person bargaining problem: Delay is costly because agents discount the future. The second source of impatience is the desire to realize the gains from trade before others do. Hence, if a player rejects an offer (or makes an unacceptable one), he will give the other players a chance to meet and reach an agreement, so that by the time he gets another chance to bargain, there will be smaller gains from trade left to realize.

As for establishing noncooperative criteria for the appropriateness of the Shapley value, our analysis suggests that the condition (VA) is key in determining whether a given characteristic function game constitutes a suitable framework for the application of the Shapley value. Furthermore, the model of Section 3 shows us the importance of institutional detail. Comparing the results of the bargaining game with the results of the partnership game enables us to conclude that the validity of a noncooperative interpretation of the Shapley value relies on a decentralized and sparse institutional framework.

From a cooperative standpoint, our analysis yields the following alternative interpretation of the Shapley value. (i) If there exists a function S which associates with each coalitional structure (state) and coalition, a value, and (ii) if coalitions occur randomly, such that given any coalitional structure, the probability that two coalitions L and M will meet and join to form the new coalitions $L \cup M$ is the same for all pairs of coalitions (L, M), and (iii) if (L, M) share the surplus created by their meeting equally, and (iv) if S associates with each M at each state q, an expected payoff over all possible future meetings, then S is the generalized Shapley value.

In particular, at the initial state, S associates with each agent his Shapley value.

The reliance of the above described program on the outcomes of "subgames" and the equal division of surplus implied by (iii) may suggest that this particular characterization of the Shapley value is similar to the preservation of differences principle characterization discussed in Hart and Mas-Colell (1985) and Myerson (1980). This, however, is misleading since their notion of a "subgame" for G entails the restriction of V to subsets of N, whereas our use of the term "subgame" refers to coarser partitions of N.

Finally, it may be interesting to consider the implications of the model under consideration for the problem of value allocation in the nontransferable utility framework, and to see if it can shed some light on the controversy over solution concepts in that area. Unfortunately, one is immediately confronted with the following problem:

The nonlinearity of the utility possibility frontier which is the defining characteristic of the NTU model implies that expectations of uncertain final allocations are in general inefficient.[8] This makes random matching models of the type we have been considering unsuitable for the analysis of NTU situations.

Roth (1980) offers an example of a 3-person NTU game in which there is a unique outcome that is strictly preferred to any other feasible outcome by two of the agents. Furthermore the two agents can, by themselves, guarantee this outcome. He argues that this outcome is indeed the only reasonable solution for this game. Unfortunately neither of the two most popular cooperative value concepts include the Roth outcome in their set of solutions. While it is not clear how the current model can, in any reasonable way, be generalized to the NTU

[8] Aumann (1985) makes this observation.

setting, establishing a satisfactory noncooperative model that yields equilibria other than the Roth outcome for this game appears to be a formidable task.

Graduate School of Business, Stanford University, Stanford, CA 94305-5015, U.S.A.

Manuscript received October, 1986; revision received January, 1988.

APPENDIX

In proving the theorem, we will use the following lemmas:

LEMMA 1: *Assume that, after various rounds of trade, the economy has reached a situation in which only two agents remain. Hence, we are at a stage where the continuation game is a 2 person bargaining game with randomly selected offerers. At this stage $q = \{M, N\backslash M\}$ for some $M \subset N$, $\varnothing \neq M \neq N$. Then the expected utilities of the continuation game for these agents are given by*

$$U(M,q,\sigma) = \frac{V(N) - V(N\backslash M) + V(M)}{2},$$

$$U(N\backslash M,q,\sigma) = \frac{V(N) - V(M) + V(N\backslash M)}{2}.$$

PROOF: By stationarity there has to be a single offer made by each player. Call these offers a and b for players M and $N\backslash M$ respectively. Using $V(M) + V(N\backslash M)$, it is easy to show that both players have been making acceptable offers. The expected utility of each agent, in equilibrium, is given by:

$$U(M,q,\sigma) = \tfrac{1}{2}(V(N) - a) + \tfrac{1}{2}b,$$

$$U(N\backslash M,q,\sigma) = \tfrac{1}{2}a + \tfrac{1}{2}(V(N) - b).$$

Furthermore if agent M refuses the offer b, his utility will be $\delta U(M, a, \sigma) + (1-\delta)V(M)$. Hence the agent knows that M will accept any offer greater than $\delta U(M, a, \sigma) + (1-\delta)V(M)$ and will not accept any offer less than this amount. So $b = \delta U(M, q, r) + (1-\delta)V(M)$. A symmetric argument establishes $a = \delta U(N\backslash M, q, \sigma) + (1-\delta)V(N\backslash M)$. Substituting for a and b in the above two equations and solving the $U(M, q, \sigma)$ and $U(N\backslash M, q, \sigma)$ yields

$$U(M,q,\sigma) = \frac{V(N) + V(M) - V(N\backslash M)}{2} \quad \text{and}$$

$$U(N\backslash M,q,\sigma) = \frac{V(N) + V(N\backslash M) - V(M)}{2}.$$

We have completely described the unique candidate for an equilibrium. Proving that this is indeed an equilibrium is straightforward.[9]

LEMMA 2: *Let $\sigma^*(\delta)$ be an equilibrium in which trade takes place at each meeting. If $N = \{1,2,\ldots,n\}$, $n > 3$, then $\lim_{\delta\to 1} U(M,q,\sigma^*(\delta)) = S(M,q)$ for $q = R(\overline{N}, i, j)$ for some i, $j \in N$ and $i \neq j$ implies $\lim_{\delta \to 1} U(i, \overline{N}, \sigma^*(\delta)) = S(i, \overline{N})$.*

Lemma 2 says that if, after one transaction occurs, $\sigma^*(\delta)$, for δ close to 1, yields an expected payoff equal to his Shapley value for every remaining player (relative to a new distribution of

[9] In fact, this Lemma does not require the stationarity assumption. See Binmore (1982) or Binmore, Rubinstein, and Wolinsky (1985).

resources), then for δ close enough to 1, $\sigma^*(\delta)$ will yield each agent expected payoff equal to his Shapley value in the original game before any transaction takes place. But this is equivalent to saying that, if the equilibrium $\sigma^*(\delta)$ yields expected payoffs according to the Shapley values in all $n-1$ player games, then $\sigma^*(\delta)$ will yield payoffs according to the Shapley values in all n player games.

PROOF: If the assumptions of Lemma 2 are satisfied, then we can easily show that

$$\begin{aligned} U(i,\bar{N},\sigma^*(\delta)) = {} & \tfrac{1}{2}\frac{1}{\binom{n}{2}}\sum_{\substack{j\in N\\ j\neq i}}\left[\delta U(i,\bar{N},\sigma^*(\delta))+(1-\delta)V(i)\right] \\ & +\tfrac{1}{2}\frac{1}{\binom{n}{2}}\sum_{\substack{j\in N\\ j\neq i}}\Big[\delta U\big(\{i,j\},R(\bar{N},i,j),\sigma^*(\delta)\big) \\ & \qquad +(1-\delta)V(\{i,j\}) \\ & \qquad -\delta U(j,\bar{N},\sigma^*(\delta))-(1-\delta)V(j)\Big] \\ & +\frac{1}{\binom{n}{2}}\sum_{\substack{\{j,k\}\subset N\\ j\neq i\neq k}}\left[\delta U(i,R,(\bar{N},j,k),\sigma^*(\delta))+(1-\delta)V(i)\right]. \end{aligned}$$

To see this, note that in $\sigma^*(\delta)$, i may (a) meet j in the initial round and not be chosen as the offerer, or (b) meet j in the initial round and be chosen as the offerer, or (c) two players other than i may meet in the first round.

The first term in the above equation involves case (a). $1/\binom{n}{2}$ is the probability that i will meet j and $\frac{1}{2}$ is the probability that he will be making the offer given he has met j. Since i is to accept j's offer by the argument used in Lemma 1, j must offer $\delta U(i,\bar{N},\sigma^*(\delta))+(1-\delta)V(i)$ to i. Similarly, the second term accounts for case (b), and the third term accounts for case (c).

After multiplying both sides by $n(n-1)$ and collecting terms, the corresponding equations for the other agents yield the following system of linear equations: $A_\delta U_\delta = C_\delta$, that is

$$\begin{pmatrix} a & \delta & \delta & \dots & \delta \\ \delta & a & \delta & & \vdots \\ \delta & \delta & & a & \vdots \\ \delta & \delta & \dots & \dots & a \end{pmatrix}\begin{pmatrix} U_\delta^1 \\ U_\delta^2 \\ \vdots \\ U_\delta^n\end{pmatrix} = \begin{pmatrix} C_\delta^1 \\ C_\delta^2 \\ \vdots \\ C_\delta^n\end{pmatrix}$$

where $a=(n-1)(n-\delta)$ and

$$\begin{aligned} C_\delta^i = {} & \sum_{\substack{j\in N\\ j\neq i}}\left[\delta U\big(\{i,j\},R(\bar{N},i,j),\sigma^*(\delta)\big)-(1-\delta)V(j)+(1-\delta)V(i)\right] \\ & +2\sum_{\substack{\{j,k\}\subset N\\ j\neq i\neq k}}\left[\delta U(i,R(\bar{N},j,k),\sigma^*(\delta))+(1-\delta)V(i)\right]. \end{aligned}$$

Furthermore, since trade takes place at each meeting and $\delta=1$, $\sum_{i\in N}U^i=V(N)$.
Hence, the following system of linear equations is equivalent to $A_\delta U_\delta = C_\delta$ (for $\delta=1$):

$$\begin{pmatrix} e & 0 & 0 & \dots & 0 \\ 0 & e & 0 & \dots & \vdots \\ 0 & 0 & e & \dots & \vdots \\ 0 & \dots & \dots & \dots & e \end{pmatrix}\begin{pmatrix} U^1 \\ U^2 \\ \vdots \\ U^n\end{pmatrix} = \begin{pmatrix} C^1-V(N) \\ C^2-V(N) \\ \vdots \\ C^n-V(N)\end{pmatrix},$$

where $e=n(n-2)$. Therefore, $\lim_{\delta\to 1}U(i,\bar{N},\sigma^*(\delta))=U^i=(C^i-V(N))/(n(n-2))$. To complete the proof, we will show that $(C^i-V(N))/(n(n-2))=S(i,\bar{N})$. First, note that in the first term of

C^i, $V(N)$ appears in each $S(\{i, j\}, R(\overline{N}, i, j))$ with a $1/(n-1)$ coefficient. Since there are $n-1$ $S(i, R(\overline{N}, i, j))$'s, the first term of C^i yields one $V(N)$. Similarly, the second term has $2\binom{n-1}{2}$ $S(i, R(N, j, k))$'s and hence contributes $2\binom{n-1}{2}(1/(n-1)) = (n-2)$ $V(N)$'s. This implies that the coefficient of the term $V(N)$ in the expression $(C^i - V(N))/(n(n-2))$ is $1/n$.

Similar computations for other $M \subset N$ also yield the Shapley value coefficients establishing the desired result.

PROOF OF THEOREM 1: It follows from (SS) that the only efficient strategy profiles are those which specify trade at every possible meeting. But, then for $n = 2$, Lemma 1 establishes the desired result. For $n \geq 3$, the result is established inductively by using Lemma 2.

PROOF OF THEOREM 2: First, we establish that there is an equilibrium $\sigma^*(\delta)$, such that trade takes place at each meeting if (VA) is satisfied.

By following the backward induction program described in Lemma 2 for all $q \in Q$, we can compute the solution of the linear system

$$A_\delta U_\delta = C_\delta$$

where the game at state q is treated as a game in which the number of players is the cardinality of $q, |q|$, and A_δ is a $|q| \times |q|$ matrix and C_δ is computed by solving all similar matrices for states $R(q, L, M)$ for all $L, M \in q$ and $L \neq M$. The solution U_δ will be referred to as $U(M, q, \sigma^*(\delta))_{M \in q}$. Hence, by Lemma 2, we know that for all $\varepsilon > 0$, there exists δ such that for all $q \in Q$, $L, M \in q$,

$$|S(L, q) - U(L, q, \sigma^*(\delta))| < \varepsilon,$$

$$|S(M, q) - U(M, q, \sigma^*(\delta))| < \varepsilon,$$

$$|S(L \cup M, R(q, L, M)) - U(L \cup M, R(q, L, M), \sigma^*(\delta))| < \varepsilon.$$

Hence, (VA) implies that, for small enough ε and the appropriate δ,

$$\delta U(L, q, \sigma^*(\delta)) + (1-\delta) V(L) + \delta U(M, q, \sigma^*(\delta)) + (1-\delta) V(M)$$
$$\leq U(L \cup M, R(q, L, M), \sigma^*(\delta)).$$

Then the strategy profile $\sigma^*(\delta)$ that states that, at every state q, every agent M accepts only those offers that are no less than $\delta U(M, q, \sigma^*(\delta)) + (1-\delta)V(M)$, and at every q, every agent L that meets M offers exactly $\delta U(M, q, \sigma^*(\delta)) + (1-\delta)V(M)$ constitutes an equilibrium. To verify this, note that the optimality of making the offer $\delta U(M, q, \sigma^*(\delta)) + (1-\delta)V(M)$ follows from the fact that no lower offer is accepted and by the inequality above, L does not benefit from making an unacceptable offer. Furthermore, clearly this profile involves trade at each meeting and therefore is efficient. By Theorem 1, it is the only efficient equilibrium.

PROOF OF THEOREM 3: Assume that we are in a situation σ in which 1 and 2 are partners with shares α and $1-\alpha$ respectively. Let $U^1(\alpha), U^2(\alpha)$, and $U^3(\alpha)$ denote the expected utilities of the agents at this state. Let γ^1, γ^2, and γ^3 be the offers made by agents 1, 2, and 3 respectively.

Reasoning as in Lemma 1, we can establish that

(1) $$\gamma^1 = \gamma^2 = [\delta U^3(\alpha) + (1-\delta) V(3)]/V(N),$$

(2) $$\gamma^3 = [\delta(U^1(\alpha) + U^2(\alpha)) + (1-\delta) V(\{1,2\})]/V(N),$$

(3) $$U^1(\alpha) = \tfrac{1}{3}\gamma^2 \alpha V(N) + \tfrac{2}{3}\alpha(1-\gamma^1) V(N),$$

(4) $$U^2(\alpha) = \tfrac{1}{3}\gamma^3(1-\alpha) V(N) + \tfrac{2}{3}(1-\alpha)(1-\gamma^1) V(N),$$

(5) $$U^3(\alpha) = \tfrac{2}{3}\gamma^1 V(N) + \tfrac{1}{3}(1-\gamma^3) V(N).$$

Substituting the values of γ^1 and γ^3 from (1) and (2) into (3), (4), and (5) and solving the resulting simultaneous system yields

$$U^1(\alpha) = \frac{2\alpha}{3}\big(V(N) - V(3) - V(\{1,2\})\big) + \frac{\alpha}{3}V(\{1,2\}),$$

$$U^2(\alpha) = \tfrac{2}{3}(1-\alpha)\big(V(N) - V(3) - V(\{1,2\})\big) + \frac{1-\alpha}{3}V(\{1,2\}),$$

$$U^3(\alpha) = \tfrac{1}{3}\big(V(N) - V(3) - V(\{1,2\})\big) + \tfrac{2}{3}V(3).$$

Observe that the equilibrium expected payoff of agent 3 does not depend on α. Let $A(1,2)$ denote the total expected utility of players 1 and 2 if they meet in the initial period. Then, by the previous argument,

$$A(1,2) = \delta\big(U^1(\alpha) + U^2(\alpha)\big) + (1-\delta)V(\{1,2\}) \tag{1}$$
$$= \delta\big[\tfrac{2}{3}\big(V(N) - V(3) - V(\{1,2\})\big) + \tfrac{1}{3}V(\{1,2\})\big] + (1-\delta)V(\{1,2\}).$$

Similarly, $B(3)$, the total expected payoff of agent 3 if 1 and 2 from the first partnership, is given by

$$B(3) = \delta U^3(\alpha) + (1-\delta)V(3) + \delta\big[\tfrac{1}{3}\big(V(N) - V(3) - V(\{1,2\})\big) + \tfrac{2}{3}V(3)\big] \tag{2}$$
$$+ (1-\delta)V(3).$$

Since the total payoff of the partnership that is established first is independent of the shares, agent 2 will offer exactly that α which keeps 1 indifferent between accepting and rejecting. That is, 2 will offer 1 the α that yields the utility $\delta U^1 + (1-\delta)V(1)$ where U^1 is the expected utility of the game for agent 1 in equilibrium σ. Hence, the equilibrium expected utility of 1 satisfies

$$U^1 = \tfrac{1}{6}\big[A(1,2) - \delta U^2 - (1-\delta)V(2)\big] + \tfrac{1}{6}\big[A(1,3) - \delta U^3 - (1-\delta)V(3)\big] \tag{3}$$
$$+ \tfrac{1}{3}\big[\delta U^1 + (1-\delta)V(1)\big] + \tfrac{1}{3}B(1).$$

The first term arises from the contingency which entails 1 meeting 2 and being chosen as the offerer. Similarly, the second term results from the possibility that 1 may meet 3 and be chosen as the offerer. The third term accounts for the possibility that, although 1 may participate in the first meeting (either with 2 or 3), he may not be chosen as the offerer. The final term results from the contingency that 2 and 3 may form the first partnership. Substituting the values of $A(1,2)$ and $B(1)$ from equations (1) and (2) into equation (3) and solving the three equation system that arises from the counterparts of (1), (2), and (3) for players 2 and 3 yields $U^1(\delta), U^2(\delta), U^3(\delta)$. Taking the limit as δ approaches 1 establishes the desired result. (Observe that our argument establishes that every equilibrium must yield expected payoffs $U^1(\delta), U^2(\delta)$, and $U^3(\delta)$. Choosing the unique set of appropriate shares to yield these payoffs is a trivial task.)

Note that the problem of deciding which partnerships will actually form does not arise since it is clear that every meeting will result in an agreement. Hence, the equilibrium is unique.

REFERENCES

AUMANN, R. J. (1985): "An Axiomation of Non-Transferable Utility Value," *Econometrica*, 53, 599–612.

BINMORE, K. G. (1980): "Nash Bargaining Theory II and III," ICERD, LSE, D.P., 80/14, 80/15.

——— (1982): "Perfect Equilibrium in Bargaining Models," ICERD, LSE, D.P., 82/58.

——— (1983): "Bargaining and Coalitions," in *Game Theoretic Models of Bargaining*, ed. by A. Roth. Cambridge: Cambridge University Press.

——— (1985): "Modeling Rational Players," Mimeo, LSE.

BINMORE, K. G., A. RUBINSTEIN, AND A. WOLINSKY (1986): "The Nash Bargaining Solution in Economic Modelling," *Rand Journal of Economics*, 17, 176–188.

CHARNES, A., J. RAUSSEAU, AND L. SEIFORD (1978): "Complements, Mollifiers, and Propensity to Disrupt," *International Journal of Game Theory*, 7, 37–50.

GALE, D. (1984): "Equilibrium in a Market with Sequential Bargaining and No Transaction Costs Is Walrasian," ICERD, LSE, D.P., 84/104.
HARSANYI, J., AND R. SELTEN (1980): "A Noncooperative Solution Concept with Cooperative Applications," Center for Research and Management, Berkeley.
HART, S., AND A. MAS-COLELL (1985): "The Potential: New Approach to the Value in Multi-Person Allocation Problems," Harvard University, HIER, D.P. 1157.
HERRERO, M. (1985): "A Strategic Bargaining Approach to Market Institutions," London University, Ph.D. Thesis.
MYERSON, R. B. (1980): "Conference Structures and Fair Allocation Rules," *International Journal of Game Theory,* 9, 169–182.
NASH, J. (1953): "Two-Person Cooperative Games," *Econometrica,* 21, 128–140.
ROTH, A. E. (1980): "Values for Games Without Sidepayments: Some Difficulties with Current Concepts," *Econometrica,* 48, 457–465.
RUBINSTEIN, A. (1982): "Perfect Equilibrium in a Bargaining Model," *Econometrica,* 50, 128–140.
RUBINSTEIN, A., AND A. WOLINSKY (1985): "Equilibria in a Market with Sequential Bargaining," *Econometrica*, 53, 1133–1150.
SELTEN, R. (1975): "Reexamination of the Perfectness Concept for Equilibrium Points in Extensive Games," *International Journal of Game Theory,* 4, 2–55.
SHAPLEY, L. S. (1953): "A Value of n-Person Games," in *Contributions to the Theory of Games II* (Annuals of Mathematics Studies 28), ed. by H. E. Kuhn and A. W. Tucker. Princeton: Princeton University Press, 307–317.