

A MODIFICATION OF HARSANYI'S BARGAINING MODEL

BY J. R. ISBELL¹

Communicated by A. W. Tucker, December 11, 1959

1. Introduction. J. C. Harsanyi has described [1] a bargaining model for an arbitrary game, which he treats by an extension of the bargaining theory of Nash [3] to obtain a unique imputation called the *solution*. I believe there are very serious objections to Harsanyi's model, of which the following is the most convincing. One can describe a three-person game (see later) in which any one player, if faced with a coalition of the other two, can get 10; but the Harsanyi solution is (15, 15, 9).

Harsanyi's model, treated by the bargaining theory advanced in my thesis [2], still yields paradoxical solutions. Perhaps there is an irremediable defect in the model. However, on the other side, Harsanyi has shown that applying the model to a game with linearly transferable utility and constant sum yields the Shapley value. This result is more or less independent of the choice of bargaining theory, and suggests that the model is at least worth considerable further study.

This note describes a bargaining theory which can be applied to the Harsanyi model to yield a unique imputation associated with every game; this imputation gives every player at least as much as he could get from playing the game against a coalition of all other players. The main ingredient in the theory is the arbitration scheme (in the technical sense [5]). This is the simple scheme which associates to the threat point (x_1, \dots, x_n) the highest feasible point of the form $(x_1 + a, x_2 + a, \dots, x_n + a)$.

I include a list of axioms which suffice to characterize this arbitration scheme, and descriptions of two test examples. Details will be published elsewhere if the theory survives criticism.

2. The modified model. For simplicity I shall treat only games in which there is no problem of strategy; the methods for passing from this case to the general case, since the work of Nash [4] and others [1; 2; 5], are routine, and since the foundations of the theory are to be examined it is desirable not to bury them too deeply. Precisely, consider an *end game* \ni in the sense of [2]. A pidgin version of the definition of an end game will suffice. We have a nonempty finite set N of players i ; with each player i there is associated a *utility space*

¹ Supported by Office of Naval Research Contract Nonr-1100(12).

I_i , which is a bounded convex set of real numbers. In [2] each I_i is supposed to be $[0, 1)$. What we need of this normalization is that *all I_i have the same length*. This will of course be used implicitly to justify interpersonal comparison of utility.

For each set S of players there is a utility space I_S , the product of all I_i , i in S . When S contains T there is a projection $\pi_T^S: I_S \rightarrow I_T$. The end game \ni , finally, is a function associating to each $S \subseteq N$ a non-empty subset $\ni(S)$ of I_S , satisfying

Convexity: Each $\ni(S)$ is compact and convex.

Domination-closure: If x is in $\ni(S)$, y in I_S , and $x_i \geq y_i$ for all i , then y is in $\ni(S)$.

Superadditivity: If S and T are disjoint, $x \in \ni(S)$, $y \in \ni(T)$, then $\ni(S \cup T)$ includes that $z \in I_{S \cup T}$ whose projections in I_S and I_T are x and y .

The treatment will involve computations with the imputations x in the various utility spaces I_S . It will be convenient to consider each I_S as embedded in the appropriate Euclidean space E_S .

The formal treatment consists of determining for each $S \subseteq N$ an imputation σ^S and vectors Δ^S and β^S , all in E_S , as follows. If S has a single element i , we do not define β^S ; $\sigma^S = \Delta^S$ is the largest element of $\ni(S)$. Inductively, having determined β^T , σ^T , and especially Δ^T for all proper subsets T of S , we define each coordinate β_i^S of β^S as $\sum [\Delta_i^T: T \subset S]$. Now consider the line l in E_S which passes through β^S and has all direction cosines equal; there is a largest common element of l and of $\ni(S)$, if there is a common element at all. We shall prove below that there is. Assuming this, let σ^S be the largest common element of l and of $\ni(S)$, and let $\Delta^S = \sigma^S - \beta^S$. Completing the induction, we obtain an imputation σ^N in $\ni(N)$ which we call the *bargaining value* of the game \ni .

The rationale for this treatment is fairly transparent. We suppose the players to examine successively the profitability of each possible coalition S . In a coalition S consisting of two players, i, j , who can separately get σ^i and σ^j , it is typically possible for the pair to get $(\sigma^i + a, \sigma^j + a)$ for some $a > 0$. The players agree *provisionally* that i is entitled to $\sigma^i + a$ and j to $\sigma^j + a$ for the largest possible a . These agreements must be provisional, for not all two-player coalitions could actually form if the game were played, and it may be impossible to fulfill all these promises. What is permanently agreed at this stage is that the final utilities of the players will be sums of increments including these numbers a . The negotiations continue in the same manner. When a coalition S of three or more players comes up for bargaining, it may turn out that the proper subsets of S have already

promised too much, i.e. that β^S is not in $\ni(S)$. Then the necessary correction, or loss, is to be divided equally among the members of S .

Harsanyi's description of the process of bargaining is somewhat more complicated, and it does not seem to be reducible to a sequence of steps linearly ordered in time. If one adheres to the rule of equal division of profits and losses, the sequence of events described above can be shown to be equivalent to the pattern of events described by Harsanyi [1].

It remains to establish the

THEOREM. *Every end game has a well-defined bargaining value σ^N , and $\sigma_i^N \geq \sigma^i$ for each player i .*

PROOF. This is trivially true for one-player games. By induction, we may assume σ^S is defined and suitable for each proper subset S of N . Let $H(N)$ be the unbounded subset of E_N consisting of all x dominated by (\leq) elements of $\ni(N)$. The line l through β^N with all direction cosines equal certainly meets $H(N)$ in a largest element τ^N ; it remains to show $\tau_i^N \geq \sigma^i$ for each i , which will imply at once $\tau^N \in \ni(N)$.

Write $i=1$, $M=N-\{1\}$. Define γ by $\gamma_1=\sigma^1$, $\gamma_j=\sigma_j^M$ for $j \in M$. By superadditivity, γ is in $\ni(N)$. Then τ_1^N will be $\geq \gamma_1=\sigma^1$ if we have $\beta_1^N - \gamma_1 \geq \beta_j^N - \gamma_j$ for each $j \in M$. But $\beta_j^N - \gamma_j$ is the sum of all Δ_j^S , S a proper subset of N containing 1 and j . $\beta_1^N - \gamma_1$ is the sum of the same numbers $\Delta_j^S (= \Delta_1^S)$ and additional terms Δ_1^T , where T runs through the subsets of $N - \{j\}$ properly containing 1. Writing $P=N - \{j\}$, we have $(\beta_1^N - \gamma_1) - (\beta_j^N - \gamma_j) = \sigma_1^P - \sigma^1 \geq 0$ by the inductive hypothesis; and the proof is complete.

3. Axioms and examples. It is not especially difficult to show that the arbitration scheme employed above is characterized by: the usual conditions of symmetry, weak optimality, and invariance under a common change of scale; the controversial Axiom A discussed at the end of [2]; and a concavity assumption concerning probability mixtures of games. Explicitly, let us consider functions T which associate to each ordered pair $G=(x, E)$ consisting of a point x in a utility space I_N and a compact convex subset E of I_N containing x and every element of I_N dominated by an element of E , a point $T(G) \in E$. The function T_0 for which $T_0(G)$ is the largest element common to E and the line l through x with all direction cosines equal has the following properties.

I. Any permutation of coordinates leaving x fixed and taking E onto itself leaves $T_0(G)$ fixed.

II. There is no $y \in E$ such that $y_i > T_0(G)_i$ for all coordinate indices i .

III. If G' is obtained from G by multiplying all coordinates by a positive real number t , then $T_0(G')$ is obtained from $T_0(G)$ by the same transformation.

IV. If $G = (x, E)$, $G' = (x, E')$, where E' contains E , then $T_0(G')$ dominates $T_0(G)$.

V. If $G^* = (x^*, E^*)$, $G_i = (x^i, E^i)$ for $i = 1, \dots, n$, and for certain positive numbers t_1, \dots, t_n we have $x^* = \sum t_i x^i$ and E^* contains the set of all $\sum t_i y^i, y^i$ in E^i , then $T_0(G^*)$ dominates $\sum t_i T_0(G_i)$. No other function has these properties.

For an example of the difficulties raised by Harsanyi's model as it stands, consider the following end game \ni on players 1, 2, 3 with utility spaces normalized as $[0, 40]$. For $i = 1, 2, 3$, $\ni(i)$ is the interval $[0, 10]$. $\ni(\{1, 2\})$ is the set of all imputations (x_1, x_2) dominated by convex combinations of $(27, 0)$, $(18, 10)$, $(10, 18)$, and $(0, 27)$. $\ni(\{1, 3\})$ is determined in the same manner by the imputations $(27, 0)$, $(18, 10)$, $(10, 12)$, $(0, 13)$, and $\ni(\{2, 3\})$ is the same geometrical figure as $\ni(\{1, 3\})$. $\ni(N)$ is the set of all imputations (x_1, x_2, x_3) such that $x_1 + x_2 + x_3 \leq 39$. One can easily compute the Harsanyi solution, which is $(15, 15, 9)$.

The bargaining value of this example as defined above is $(13.8, 13.8, 11.4)$, which seems to me far more reasonable.

Finally consider the following almost trivial example. There are three players, 1, 2, 3, and for each coalition S the form of $\ni(S)$ is $\{x: \sum x_i \leq v(S)\}$. The numbers $v(S)$ are .5 for each one-player set, 1.6 for $S = \{1, 2\}$, 1.0 for $\{1, 3\}$ and $\{2, 3\}$, 2.1 for $N = \{1, 2, 3\}$. The bargaining value, Harsanyi solution, and Shapley value of this game coincide at $(.8, .8, .5)$. (The three values will always coincide for a game with linearly transferable utility and constant sum, but the example is not constant-sum.) This certainly seems reasonable off-hand, but note the considerably different analysis of essentially the same example in my thesis [2, p. 385].

REFERENCES

1. J. Harsanyi, *A bargaining model for the cooperative n-person game*, Annals of Mathematics Studies, no. 40, Princeton, 1959, pp. 325-355.
2. J. Isbell, *Absolute games*, Annals of Mathematics Studies, no. 40, Princeton, 1959, pp. 357-396.
3. J. Nash, *The bargaining problem*, Econometrica vol. 18 (1950) pp. 155-162.
4. ———, *Two-person cooperative games*, Econometrica vol. 21 (1953) pp. 128-140.
5. H. Raiffa, *Arbitration schemes for generalized two-person games*, Annals of Mathematics Studies, no. 28, Princeton, 1953, pp. 361-387.

PURDUE UNIVERSITY