Ψ Stability: A New Equilibrium Concept for n-Person Game Theory

R. Duncan Luce
Columbia University[1]

BACKGROUND

Those aspects of game theory which are probably of most interest to the social scientist are the theories of coalition formation for games with a transferable utility which are based on the notion of a characteristic function of a game. Two major theories have been presented in this area: von Neumann and Morgenstern's solutions (8) and Shapley's value (7). Criticisms of the former concept have been based on the grounds that it gives only a normative theory and that, with respect to specific situations, it is not entirely clear what the theory asserts will or should happen. In addition, the definition of a solution has led to such serious mathematical difficulies that it has not yet proved possible to characterize all solutions of any broad classes of games. Shapley's value, on the other hand, was not put forward as either a descriptive or a normative theory, but as a means of estimating the *a priori* value of the game to each of the players.

In this paper I shall describe a theory which attempts to be more descriptive than either of these two. In its present form it is not clearly an adequate descriptive theory, but it suggests a new approach which may ultimately be adapted to a suitable descriptive theory.[2] Before presenting it, however, it may be useful to summarize briefly the notions of a characteristic function, S-equivalence, imputations, etc.

Let a game have n players 1, 2, $\cdots$, n and denote the set of them by I_n. von Neumann and Morgenstern (8) observed that for $n > 2$ there is the distinct possibility that subsets of players, called *coalitions*, may agree on systems of related strategies in order to improve their joint outcome over that expected if they were to act individually. If the set of players S forms a coalition, then the worst thing that can happen to S is for the remaining set of players, $-S$, also to form a coalition, and for the game to be played between these two opposed coalitions. This results in the familiar 2-person situation, which under broad conditions has been solved by the minimax theorem. From this observation it is possible to derive a measure v of coalition strength in terms of the utility units of the underlying normalized game. von Neumann and Morgenstern showed that v is a real-valued set function defined over the subsets of I_n which satisfies

1) $v(\phi) = 0$, and
2) if R and S are disjoint subsets of I_n,
   $v(R \cup S) \geqslant v(R) + v(S)$.

If, in addition, the game is constant-sum,

3) $v(S) + v(-S) = v(I_n)$ for every $S \subset I_n$.

[1] Behavioral Models Project, Bureau of Applied Social Research. This paper was prepared through support extended to the Bureau by the Office of Naval Research, Contract Nonr-266(21).

[2] Since this paper was prepared an analogue of the Ψ-stability concept has been devised for games in which the utility is not transferable. This notion generalizes Nash's definition of an equilibrium point of a non-cooperative game to games in which coalitions are allowed.

The function v is called the *characteristic function* of the game. It can be shown that any real-valued set function satisfying conditions (1) and (2) is the characteristic function of some game, and so the study of all games with a transferable utility becomes the study of all characteristic functions.

In words, the first condition represents the strategic inconsequence of the null set, and the second simply states that a coalition composed of disjoint R and S can do everything R and S can do separately, and possibly more. It is certainly true that an *a priori* discussion of coalition strength, assuming that it can be represented by real numbers, would not lead to less than conditions (1) and (2); indeed, there would be a temptation to try to require more. Thus, one might expect that the theory of games based on the characteristic function could be applied to many situations where the game substructure is either absent or obscure. This seems to be a very important point for social science, for it may be possible to determine empirically a characteristic function for a conflict-of-interest situation without ever knowing, say, the normal form of the underlying game (see section on Experimental Data).

A game is called *inessential* if $v(I_n) = \sum_{i \epsilon I_n} v(\{i\})$; otherwise it is called *essential.* It follows that if a game is inessential, $v(S) = \sum_{i \epsilon S} v(\{i\})$ for every $S \subset I_n$, so in an inessential game there is no gain in forming coalitions; hence, its coalition theory is trivial.

Two games, v and v′, over I_n are called S-*equivalent* if there exist a positive constant c and constants a_i such that $v(S) = cv'(S) + \sum_{i \epsilon S} a_i$, $S \subset I_n$. In effect, two such games have the same strategic character, for the constant c represents a change in the "monetary" scale and the constants a_i are payments determined independently of the outcome of the game, and may, if we choose, be paid to (or by) the players before the game begins. The relation of S-equivalence divides the set of games into equivalence classes, the members of each class involving the same strategic considerations. This being the case, it is adequate to develop any theory in terms of one representative from each class. von Neumann and Morgenstern (8) showed that there is one and only one characteristic function in each class of essential games satisfying $v(\{i\}) = -1$ and $v(I_n) = 0$; this they called the reduced form, and we shall call it the *—1,0 reduced form.* Equally, one can show that there is one and only one characteristic function in each class, which we shall denote by m, which satisfies $m(\{i\}) = 0$ and $m(I_n) = 1$; this we shall call the *0,1 reduced form.* If v′ is a characteristic function of an essential game, the transformation

$$m(S) = \frac{v'(S) - \sum_{i \epsilon S} v'(\{i\})}{v'(I_n) - \sum_{i \epsilon I_n} v'(\{i\})}$$

yields the 0,1 reduced form of the same equivalence class, and

$$v(S) = nm(S) - |S|,$$

where $|S|$ = number of elements in S, yields the S-equivalent —1,0 reduced form.

In n-person game theory it is assumed that payments are made in a numerical utility which is comparable and transferable between players, thus any payments made at the end of the game and any side payments between players can all be combined into a final payment of, say, x_i to player i. It is assumed that no player will accept a final payment less than he can insure for himself and it is also assumed that "rational" players will divide the total value of the game among themselves, so we have the conditions

1) $x_i \geq v(\{i\})$, $i \in I_n$, and

2) $\sum_{i \in I_n} x_i = v(I_n)$.

Any real n-tuple satisfying conditions (1) and (2) is called an *imputation* of the game with characteristic function v.

It is worth observing that when a game is put in 0,1 reduced form, the set of imputations is simply the set of all discrete probability distributions over the n players. The conditions met by the 0,1 reduced form are also similar to those for a probability measure over the subsets of a finite set, but instead of being additive it is superadditive. If for a given X we define

$$P_X(S) = \sum_{i \in S} x_i,$$

then P_X is a probability measure over the subsets of I_n. It is intuitively clear that if, for some S, m(S) is much larger than $P_X(S)$, then there will be strong "forces" tending to cause the coalition S to form and to command a new outcome, say, X′ such that $P_{X'}(S)$ is close to m(S). Thus, the equilibrium problem of n-person game theory involves finding a probability measure P_X which in some sense approximates the normalized superadditive measure m. The heart of the problem is to determine a suitable definition for "in some sense approximates."

Ψ-STABILITY OF GAMES

The von Neumann and Morgenstern solution to the above problem involves finding sets of imputations which together are stable in the sense that for any S one of them approximates m(S) at least as well as any other imputation. [We need not go into the exact details of their definition of a solution for it is well known (5, 8)]. They were driven to using sets of imputations because they allowed any possible coalition to form at any time; had they looked for a single imputation X, then P_X would have had to approximate m for all subsets of I_n. Intuitively, it is clear that this is not possible using only one X, though they have shown it to be possible, in a sense, if certain sets of X's are used.

An alternative procedure, the one I am proposing, assumes that only some of the possible coalitions can be considered for formation at any particular instant. Such an assumption seems to be in accord with many observations about economic and social situations to the effect that changes generally seem to be small modifications of the current situation; such changes over time may result in radical reorganizations, but each step of the process is relatively small. We shall characterize the state of coalition formation in a game by a pair (X,τ), where X is the imputation tentatively agreed upon and τ is the tentative partition of the players into coalitions, i.e., $\tau = (T_1, T_2, \cdots, T_t)$, where the T_i are non-overlapping coalitions which exhaust I_n We shall wish to say that a coalition S can be formed, if there is any reason to do so, provided that S is not too different from some T_i of τ. It will be adequate to suppose that this information is given in the following manner: For each possible τ let $\Psi(\tau)$ be a given set of coalitions which includes all the coalitions in τ. We shall interpret Ψ as follows: S is a coalition which can arise when the players are arranged according to τ if and only if $S \epsilon \Psi(\tau)$. Any $S \epsilon \Psi(\tau)$ will be called a *Ψ-critical coalition* of τ.

Suppose Ψ is given, then a pair (X,τ) is called *Ψ-stable* if the following two conditions are met:

1) if $T \epsilon \tau$ and $|T| > 1$, then for any $i \epsilon T$,
   $x_i > v(\{i\})$, and
2) if S is a Ψ-critical coalition of τ,
   $v(S) \leqslant \sum\limits_{i \epsilon S} x_i$.

The first of these conditions reflects the intuition that a player i will not trouble to participate in a coalition of two or more members if they do not agree that he shall receive more than he can insure for himself when operating alone. The second condition can be best understood if we suppose, on the contrary, that for some Ψ-critical coalition S, $v(S) > \sum\limits_{i \epsilon S} x_i$. In this case, S is a coalition which can be formed and there is a positive profit in so doing; thus each player $i \epsilon S$ can be made to profit from the change, by giving him, say,

$$x_i + \frac{v(S) - \sum\limits_{i \epsilon S} x_i}{|S|}$$

and so the change might well be effected. It will, of course, have to compete with other possible and profitable changes. If, on the other hand, condition (2) is met, no profit is assured by the change and we shall assume that the change will not occur. Thus, with respect to changes limited according to Ψ, such a Ψ-stable pair will be in a state of equilibrium.

A game will be called *Ψ-stable* if it has at least one Ψ-stable pair, otherwise it will be called *Ψ-unstable*. A discussion of the uniqueness of Ψ-stable pairs will be postponed until later.

It is not difficult to show that Ψ-stability is a concept invariant under S-equivalence, so it is adequate to state all results in terms of one or the other of the reduced forms.

Observe that for each τ, $\Psi(\tau)$ is a set (of coalitions) and so inclusion relations may hold between two different Ψ's. If $\Psi(\tau) \subset \Psi'(\tau)$ for every τ, then if the game is Ψ'-stable, it is Ψ-stable, and if it is Ψ-unstable, it is Ψ'-unstable.

Without any further specification as to the nature of Ψ it is unlikely that any results can be proved, so the remainder of this paper will be devoted to an examination of certain special cases. We wish to make specifications which reflect the intuition that a Ψ-critical coalition of τ should be "near" at least one of the coalitions of τ. We shall define three special cases of Ψ.

The first case will be denoted V_k. Let k be any integer in the range 1 through $n - 2$. A coalition S is a member of $V_k(\tau)$ if and only if there exists a coalition $T_i \epsilon \tau$ such that

$$|(S - T_i) \cup (T_i - S)| \leqslant k.$$

Put in another and possibly more revealing way, S is a member of $V_k(\tau)$ if and only if there exists a coalition $T_i \epsilon \tau$, a subset H of T_i, and a subset G of $- T_i$ such that

$$S = (T_i \cup G) - H \text{ and } |G \cup H| \leqslant k.$$

In words, we suppose S can be considered by the players T_i if by adding the players G to T_i and expelling the players H from T_i, then S is formed and no more than k players have been involved in changes of alliance. While we have defined the concept for the full range of k, probably only $k = 1$ and 2 will be used in social science applications.

The second special case arises from the following consideration: Suppose players i and j are in coalitions and that they find that by bolting their respective coalitions and banding together as the coalition $\{i, j\}$ they can each profit, then it is reasonable to suppose that this relatively simple change would occur. In general, this change is not admitted by V_k, so we define a new Ψ, call it W_k. A coalition S is a member of $W_k(\tau)$ if and only if either

1) $S \epsilon V_k(\tau)$, or
2) $|S| \leqslant k + 1$.

The final special case of Ψ is the one in which all possible coalitions are admissible for each τ; this universal case we denote by $E(\tau)$.

It is easily seen that the following inclusion relations hold:

$$V_k(\tau) \subset W_k(\tau) \subset E(\tau)$$
$$V_{n-2}(\tau) = W_{n-2}(\tau) = E(\tau)$$
$$V_k(\tau) \subset V_{k'}(\tau) \text{ if } k \leqslant k'$$
$$W_k(\tau) \subset W_{k'}(\tau) \text{ if } k \leqslant k'$$

THEOREMS

The results presented in this section will not be proved. Proofs may be found in (4), where only the V_k case is studied using the terminology "k-stability" rather than "V_k-stability." The extension of those results to the ones given here is not difficult.

If a game is constant-sum, then it is E-unstable. This theorem is not true, as can be shown by an example, if we drop the constant-sum requirement. From this theorem it follows that the 3-person constant-sum game is both V_1- and W_1-unstable. In a sense, this means that the 3-person game is "absolutely unstable," for so long as some change from the existing coalitions is always admitted as a possibility, V_1 is the Ψ allowing the least change. And even so the 3-person constant-sum game is unstable, hence our description as "absolutely unstable."

Since this result about the constant-sum 3-person game seems obscure to some, it may be worth spending a moment examining the result in some detail. The $-1,0$ reduced form of the game is

$$v(S) = \begin{cases} 0 \\ 1 \text{ if } |S| = \\ -1 \end{cases} \begin{cases} 3 \\ 2. \\ 1 \end{cases}$$

Clearly, if each player is alone, there will be a tendency for a coalition to form. Let us suppose without loss of generality $\{1, 2\}$ forms. It can command a payment of 1, which by some means is split up as x_1 and $1 - x_1$, and player 3 gets -1. There is no loss in assuming $x_1 \leqslant 1/2$. Player 3 can come to player 1 and suggest they form the coalition $\{1, 3\}$ in which player 1 will receive $x_1 + \delta$, where $0 < \delta < 1 - x_1$, and player 3 will receive $1 - x_1 - \delta > 0$. Thus both can be made to profit by the change and so we may suppose it will be made. But clearly, player 2 can approach player 3 with a similar offer, and so on. We mean, of course, "and so on" when we say the game is V_1-unstable.

It is known (8) that all 4-person constant-sum games can be given in the following $-1,0$ reduced form:

$$v(S) = \begin{cases} 0 \\ -1 \\ 1 \\ 0 \end{cases} \text{when } |S| = \begin{cases} 0 \\ 1 \\ 3 \\ 4 \end{cases} \text{and } \begin{array}{l} v(\{1,4\}) = 2y_1 = -v(\{2,3\}) \\ v(\{2,4\}) = 2y_2 = -v(\{1,3\}) \\ v(\{3,4\}) = 2y_3 = -v(\{1,2\}) \end{array}$$

where the numbers y_i all lie in the closed interval $[-1, 1]$. All such games have the following property: there exists a set of n constants q_i such that

1) $\sum_{i \epsilon I_n} q_i = v(I_n)$, and

2) $v(\{i, j\}) = q_i + q_j$, $i \neq j$.

Shapley (6) has called any game with this property a *quota game* and the n-tuple $Q = ||q_1, q_2, \cdots, q_n||$ is called the *quota*. It is not difficult to show that the quota for any constant-sum 4-person game is given by

$$Q = ||y_1 - y_2 - y_3,\ y_2 - y_1 - y_3,\ y_3 - y_1 - y_2,\ y_1 + y_2 + y_3||.$$

By the first result we stated and the second inclusion relation at the end of the last section, we know that all 4-person constant-sum games are V_2- and W_2-unstable. It can be shown that such a game is V_1-stable if and only if it is W_1-stable; this stability may be characterized as follows: A 4-person constant-sum game is V_1-stable if and only if the quota is an imputation. In that case the pair $[Q, (\{1\}, \{2\}, \{3\}, \{4\})]$ is always V_1-stable. The pair $[Q, (\{1, 2\}, \{3\}, \{4\})]$ is V_1-stable if and only if $y_1 = -y_2$, $-1 < y_1 < 1$, and $y_3 = -1$. In addition, all permutations of I_4 lead to conditions for the V_1-stability of τ's having one 2-element coalition. Finally, the pair $[Q, (\{1, 2, 3\}, \{4\})]$ is V_1-stable if and only if $y_1 + y_2 + y_3 = -1$ and the first three $q_i > -1$. Again, one must consider all permutations of I_4.

Thus the V_1- and W_1-stability picture of the 4-person constant-sum games is completely known.

I turn now to the stability conditions for a rather broader class of games. A game is called *simple* if, in the 0,1 reduced form, $m(S) = 0$ or 1 for every coalition S. This generalizes, and simplifies, the definition of simple constant-sum games given by von Neumann and Morgenstern. A coalition S such that $m(S) = 0$ is called *losing*, and one for which $m(S) = 1$ is called *winning*.

It can be shown that a simple game is V_k-stable if and only if it is W_k-stable. Furthermore, a simple game is V_k-stable if and only if either

1) there is no $(k+1)$-element winning coalition, or
2) the intersection of all $(k+1)$-element winning coalitions is non-empty.

I have not yet been able to give a complete characterization of the V_k-stable pairs for $k > 1$, but for $k = 1$ a characterization is known. If there are no winning 2-element coalitions, then any pair (X, τ), with $x_i > 0$ for $i \epsilon T$, where $T \epsilon \tau$ and $|T| > 1$, is V_1-stable if either

1) every V_1-critical coalition of τ is losing, or
2) $\tau = (\{1\}, \{2\}, \cdots, \{s\}, T)$, for an appropriate relabeling of the players,

and

a) every V_1-critical winning coalition is of the form $T \cup \{i\}$, $i \epsilon I_n$, and
b) if $T \cup \{i\}$, $i \epsilon I_n$, is winning, then for every $j \epsilon -(T \cup \{i\})$, $x_j = 0$.

If, however, there is a winning 2-element coalition, then let R be the intersection of all 2-element winning coalitions. We may always relabel the players so that $R = \{1\}$ or $\{1, 2\}$. If $R = \{1\}$ the only V_1-stable pair is

$[\|1, 0, 0, \cdots, 0\|, (\{1\}, \{2\}, \{3\}, \cdots, \{n\})]$.

If $R = \{1, 2\}$ then both

$[\|p, 1-p, 0, \cdots, 0\|, (\{1\}, \{2\}, \{3\}, \cdots, \{n\})]$, where $0 \leqslant p \leqslant 1$ and

$[\|p, 1-p, 0, \cdots, 0\|, (\{1, 2\}, \{3\}, \cdots, \{n\})]$, where $0 < p < 1$ are

V_1-stable.

While the above results give some indication as to the nature of the V_1-stable pairs, they do not make clear just how many simple games are V_1-stable. To examine this point further, we need the notion of a decomposable game (8) which, in words, means that there is a set T such that there is no strategic interaction between what happens on T and what happens on $-T$. Formally, a game m over I_n is *decomposable* into games on T and on $-T$ if for every coalition S of I_n:

$$m(S \cap T) + m(S - T) = m(S).$$

It can then be shown that a game is both simple and V_1-unstable if and only if it is decomposable into the essential 3-person constant-sum game and the $(n-3)$-person inessential game. Since the coalition theory of inessential games is trivial, this result means that the theory of all simple V_1-unstable games is the same as that of the 3-person constant-sum game; hence almost all constant-sum games and all non-constant-sum simple games are V_1-stable.

These are the only major classes of games for which any detailed Ψ-stability results are known;[3] however, one further result of some mathematical interest may be mentioned. It might be thought that the imputation X of a V_k-stable pair would always be one of the imputations in a von Neumann-Morgenstern solution, but an example can be given to show this is not the case. An open mathematical problem is to decide when such an X is a member of a solution, and when it is not.

DISCUSSION

One of the first questions people seem to ask about any new concept in n-person game theory is whether it leads to a unique outcome for a game; possibly this interest in uniqueness stems in part from the singular non-uniqueness of the solutions of most games. It is clear from the above results that the concepts of V_k- and W_k-stability lead to far fewer possible outcomes than does the solution notion, but not in general to a unique outcome. In the 4-person constant-sum games the imputation is unique (the quota) but for some 4-person games, though not many, there is more than one possible arrangement of the players into coalitions. This may not be a serious difficulty, for presumably in this equilibrium theory, like so many in physics, the decision as to which equilibrium point will occur depends on the initial conditions of the system and that to predict it requires a full dynamic theory, not just an equilibrium theory. It is not as intuitively acceptable that a dynamic theory plus initial conditions will lead to a unique outcome for simple games. It will be recalled that in a simple game having exactly one 2-element winning coalition, V_1-stability leads to any imputation of the form

$$||p, 1-p, 0, \cdots, 0||, \text{ where } 0 < p < 1.$$

The theory did not select the value of p. In the same spirit as von Neumann and Morgenstern's discussion of some of the outcomes in the theory of solutions, we may suppose that the choice of p depends on the relative bargaining abilities of the two players—but this does not really resolve the difficulty. Whether it will prove possible to incorporate such a notion into a dynamic theory in such a way as to predict a unique outcome is certainly not at the moment clear.

While the present theory does not have the virtue of uniqueness, it does have at least two points in its favor. First, the predictions which it makes are not only for the resulting payments to the players, but also for the expected arrangement of the players into coalitions. In some situations, it may be much easier to observe the latter than the former. Second, the stability definition, at least for some assumptions about Ψ, is mathematically much easier to cope with than is, say, the solution concept; as evidence, we have stated rather complete V_1- and W_1-stability results for all 4-person constant-sum games and for all simple games, whereas complete results on solutions are known for only certain special cases of both classes of games.

If we consider Ψ-stability as a predictive theory, then certain serious questions can be raised about the assumption embodied in our use of the function Ψ. I know of no situation, except when no coalitions are allowed, where Ψ is explicitly known

[3] Since the preparation of this paper, work has been completed on the V_k-stability of quota games and of symmetric games; see Luce, R. D. K-stability of symmetric and of quota games. *Ann. Math.*, in press.

to the players, though in many situations something approximately like Ψ appears to be operative. We generally ascribe this to limitations in people's perceptions or to their inability to organize complicated changes in alliances. But however we describe the intuitions behind Ψ, it is fairly clear that the sharp dichotomy we have assumed is not reasonable. It may be possible to modify the theory so that it is more realistic by assuming as known the probability $p(S, \tau)$ that the coalition S will be considered as a possible change when the players are tentatively arranged into coalitions according to τ. Such a theory should lead to assertions of the form "a pair (X, τ) has a certain probability of being an equilibrium point." Of course, as with the Ψ function we have discussed here, it will be necessary to make assumptions about the probabilities before any detailed results can be expected. The development of such a probabilistic stability theory seems to be one major open problem.

A second and possibly more important problem is the development of a suitable dynamic theory. In the spirit of our equilibrium theory, we would suppose Ψ given and assume that a change from (X, τ) to τ' will be a possibility only if for each $S \epsilon \tau'$ either

1) $S \epsilon \Psi(\tau)$ and $v(S) > \sum_{i \epsilon S} x_i$ or
2) there exists a $T \epsilon \tau$ and $U \subset T$ such that
   a) $S = T - U$, and
   b) for each $i \epsilon U$, the coalition S' of τ' containing i is a member of $\Psi(\tau)$ and $V(S') > \sum_{i \epsilon S'} x_i$.

There are at least two major difficulties in developing such a theory. First, suppose the change to τ' is made, what imputation will be agreed upon? Second, suppose that there exist two overlapping coalitions S and S′ such that both $v(S) > \sum_{i \epsilon S} x_i$ and $v(S') > \sum_{i \epsilon S} x_i$, then which of the two changes will occur, or, better, what determines the probability that a particular change will be made? The solution of these difficulties may very well lead to a Markov chain where the states are pairs (X, τ) and the transition probabilities are derived from those probabilities mentioned under the second difficulty; if this is the case then much of the theory of Markov chains will be applicable.

The implications of an adequate dynamic theory to social science seem great. Not only would such a theory allow a decision as to whether the initial conditions under which coalition formation begins uniquely determine the stable pair which ultimately results, but it will also allow a discussion of situations which are not yet in equilibrium. It may be that most situations of conflict-of-interest which we observe in society are of this type.

EXPERIMENTAL DATA

I know of only one game experiment with which I can compare the predictions of any Ψ-stability theory; it was performed at RAND by Kalisch, Milnor, Nash, and Nering (3). In it subjects were given the characteristic function of a game, which described money payments to the various possible coalitions, and they were given 10 minutes in which to form agreements. As agreements were reached, the subjects reported them to an umpire who in turn announced them to the entire group, and if there were no objections he recorded them and rigidly enforced them at the end of the experiment. In addition, the experimenters mention that there were a number of informal arrangements not reported to the umpire but which were adhered to in all cases. The experiment was run with 4, 5, and 7 players; however, because the subjects were face-to-face about a table there were serious geometrical obstacles to the formation of certain coalitions. They report, for example, that all coalitions in the 7-person game were among players who were adjacent at the table. Such an effect is certainly not a part of the theory of games, and the present special cases of Ψ-stability theory offer no predictions under these asymmetric conditions, though by a careful and complicated choice of Ψ it might be made to do so.

If the experiment were to be run again, it might help to use an adaptation of the Bavelas partitioned table for small group experiments (2). This simple apparatus allows no significant geometrical effects, but it does entail the use of written messages. The latter may not be the disadvantage it seems for it has the effect, if the subjects are identified by letters or numbers, of preserving anonymity, of slowing down the experiment, and of giving a permanent record of the bargaining. The first effect would be desirable since in the RAND experiment it was felt that the subjects did not approach the problem with the ruthlessness assumed in the theory; the second because it would give the subjects more time to consider the logic of the situation; and the third because it would allow a more careful analysis of the details of the bargaining and thus might suggest an appropriate Ψ to consider.

Only for the case $n = 4$ did they feel that the geometrical effects were unimportant, so we shall restrict our attention to those cases. They studied two different 4-person constant-sum games, each in what amounted to the 0,1 reduced form and in an S-equivalent form. The data are based on 8 groups in each case. It is clear from figures 1 and 2 that the subjects did not perceive the strategic identity of S-equivalent games; it appears, rather, that they reacted to the numbers assigned to coalitions and not to the relative differences. This suggests that the 10-minute time limit may have been a mistake; possibly unlimited time using written communication would change the results. The imputations predicted by V_1-stability and by Shapley's value (7) are shown in each of the figures. In the symmetric game both theories predict 0 payment to each of the players, which is approximately what occurred in the 0,1 reduced form, though not in the S-equivalent form. In the non-symmetric game the theories differ appreciably. The V_1-prediction accords reasonably closely to the 0,1 reduced form data and Shapley's value to the S-equivalent data.

In addition to predicting the imputation, V_1-stability theory also states what coalition arrangements should be expected. In the non-symmetric game one expects $\{A, B, C\}$ to form; actually this occurred on only two out of eight trials. In the symmetric case no non-trivial coalitions nor one grand one is predicted. This occurred only once in eight trials, and in three other trials two opposing

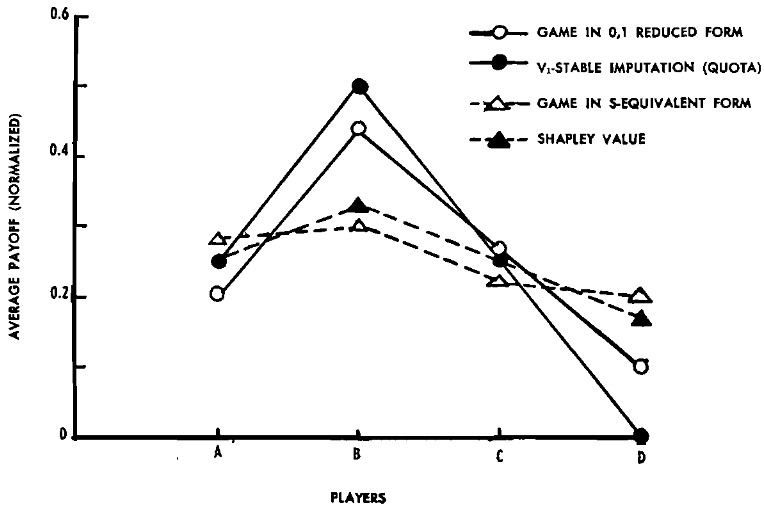

Figure 1.

Four-person constant-sum game: $y_1 = y_3 = -\frac{1}{2}$, $y_2 = 0$.
Adapted with permission from Kalisch *et al.* (3).

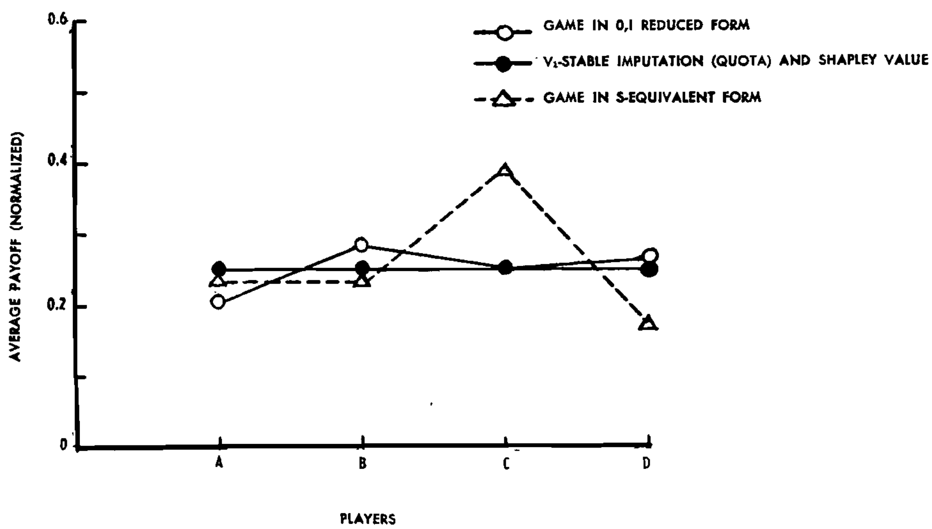

Figure 2.

Four-person constant-sum game: $y_1 = y_2 = y_3 = 0$.
Adapted with permission from Kalish *et al.* (3).

2-element coalitions formed and zero payments for everyone were agreed upon. Four times a 3-element coalition formed, and the isolated player i was given $v(\{i\})$ and the others divided $-v(\{i\})$ with considerable discrimination against the third addition to the coalition, except in one case when the 3-element coalition was formed without a 2-element intermediate stage.

It is somewhat difficult to say exactly what these data mean for Ψ-stability theory, but I do not believe them to be very crucial one way or the other. Some aspects of the experimental technique seem questionable considering the underlying rationality assumptions of game theory, and this tends, in my opinion, to make the results of limited reliability. Nonetheless, our prediction of the imputation is close enough to the data of the 0,1 reduced form to be interesting. It may be that a slight modification of Ψ would improve our prediction of the coalition formation. For example, one obtains the impression from their discussion that additions to coalitions were generally made one member at a time and that sometimes two coalitions coalesced into one (the latter was not important for $n = 4$), but that once a person was in a coalition he did not leave it or get expelled from it. I have not yet examined what effect such a modified Ψ would have on the predictions.

Two empirical problems seem worthy of attention so far as the present Ψ-stability theory is concerned. First, more experiments, more carefully designed but in the same spirit as the RAND one, should be carried out. This program may suggest certain other special cases of Ψ which should be studied.

Second, sociologists and social psychologists might give attention to the estimation of the subjective characteristic functions of people in certain existing conflict-of-interest situations. It is certainly quite impossible to determine the characteristic function for most real situations by obtaining the normal form of the underlying game and then performing the complex calculations involving the minimax theorem to get v. To get a practical theory some means of empirical estimation is needed. I have only one suggestion, which is due to Ernest Adams of the Behavioral Models Project and myself (1). Let K be a set which consists of all the subsets of players in a given situation and all the risk alternatives of the form "coalition S with probability α and coalition R with probability $1 - \alpha$." Now, let an observer, possibly one of the players, impose a preference relation over K by deciding for each pair of alternatives from K which he prefers if he imagines that he will receive the payment an average member of the coalition of his choice, R, receives in the game of R vs. $-R$. In a risk situation he is asked to imagine that he will, with probability α, receive what an average member of R gets when R is opposed by $-R$ and, with probability $1 - \alpha$, what an average member of S gets when S is opposed by $-S$. He is to treat the null set as the alternative of not participating at all. It does not seem unreasonable to suppose that the preference relation should satisfy the von Neumann utility axioms (8), though it is another matter how well it actually does. If it does then we know that there is a numerical utility u, unique up to a linear transformation defined over K such that

1) S is preferred to R if and only if $u(R) < u(S)$, and
2) $u[\alpha R + (1 - \alpha)S] = \alpha u(R) + (1 - \alpha)u(S)$.

If one makes the added and not too unreasonable assumption that for disjoint

R and S, subsets of I_n with r and s players respectively, the coalition $R \cup S$ is at least as desirable as the alternative of being in R with a probability of $r/(r+s)$ and of being in S with a probability of $s/(r+s)$, then u also satisfies:

3) if R and S are disjoint subsets of I_n,

$$u(R \cup S) \geqslant \frac{r}{r+s} u(R) + \frac{s}{r+s} u(S).$$

Now, if one makes the change of variable $v(S) = s\,[u(S) - u(\phi)]$ then it is easy to see that v is a characteristic function and that while u is determined up to a linear transformation, v is determined up to a change of scale. So the class of v's determined by a given preference relation is a subset of an S-equivalent class of games. It can be shown that if the characteristic function of a game is known and if our observer expresses his preferences rationally according to the given charasteristic function, then the function determined by the above procedure is S-equivalent to the actual characteristic function. The task of empirically determining the characteristic function is, if this procedure works, no more difficult than determining the numerical utility of a comparable set of alternatives. This, of course, is difficult, but does not compare with determining the normal form of a game.

REFERENCES

1. Adams, E. W., and Luce, R. D. *A proposal for the empirical determination of the characteristic function of a game.* Behavioral Models Project, Bureau of Applied Social Research, Columbia University, Technical Report Number 6, 1954.
2. Christie, L. S.; Luce, R. D., and Macy, J., Jr. *Communication and learning in task-oriented groups.* Research Laboratory of Electronics, Massachusetts Institute of Technology, Technical Report Number 231, 1952.
3. Kalisch, G.; Milnor, J. W.; Nash, J, and Nering, E. D. Some experimental n-person games. In R. M. Thrall, C. H. Coombs and R. L. Davis (Eds.), *Decision processes.* New York: John Wiley and Sons, 1954. Pp. 301-327.
4. Luce, R. D. A definition of stability for n-person games. *Ann. Math.*, 1954, *59*, 357-366.
5. McKinsey, J. C. C. *Introduction to the theory of games.* New York: McGraw-Hill Book Company, 1952.
6. Shapley, L. S. Quota solutions of n-person games. *Contributions to the theory of games, II.* Annals of Mathematics Studies, No. 28, Princeton University Press, Princeton, 1953. Pp. 343-359.
7. Shapley, L. S. A value for n-person games. *Contributions to the theory of games, II.* Annals of Mathematics Studies, No. 28, Princeton University Press, Princeton, 1953. Pp. 307-317.
8. von Neumann, J., and Morgenstern, O. *Theory of games and economic behavior.* Princeton: Princeton University Press, 1947.