AD NUMBER

AD032467

LIMITATION CHANGES

TO:

Approved for public release; distribution is unlimited.

FROM:

**Distribution authorized to U.S. Gov't. agencies and their contractors;
Administrative/Operational Use; APR 1954. Other requests shall be referred to Office of Naval Research, Arlington, VA 22203.**

AUTHORITY

ONR ltr 9 Nov 1977

Behavioral Models Project
(NR 042-115)

TECHNICAL REPORT NO. 5

A Survey of the Theory of Games

by

R.Duncan Luce and Howard Raiffa

April 1954

Part I: Introduction

Part II: in preparation

Part III: n-Person Games

Part IV: in preparation

CU-5-54-NONR 266(21)-BASR
Bureau of Applied Social Research
New York 27, N.Y.

The accompanying paper on game theory consists of two parts of a report which in final version will have four parts. The omitted parts will cover 2-person theory and the relation between game theory and various topics such as statistics, linear programming, etc. I am sending you this incomplete version because I should like to receive criticisms of the work as soon as possible; to this end let me cite the ultimate purpose of the report.

It is one of a series of reports being written by various people on mathematical models in the behavioral sciences. They are expository in nature and they are designed primarily for two audiences:

1. social scientists with some, but limited, mathematical training who wish to find out some of the structure and of the conclusions of the various mathematical models, but who have neither the interest nor the mathematical sophistication to follow detailed formal proofs;

2. mathematicians interested in mathematical applications in the social sciences who want a quick survey of the area and who can, if they become interested, obtain the mathematical details from books and articles referred to in the exposition.

When these reports are finally issued as a unit, they will be accompanied by a short exposition of some basic mathematical concepts and notations. For example, terms like set, function, relation, product space, etc., and notations like $\in$, $\cup$, $\cap$, $\subset$, etc., will be explained. It was felt that it was better to do this in one place rather than to try to make each report completely self-contained.

I would very much appreciate it if you can spare the time to give this partial report a critical reading with these aims and facts in mind, and you may be sure that in preparing the final draft I shall put to good use any (preferably detailed) comments you care to make.

Sincerely,

R. Duncan Luce

Bureau of Applied Social Research
427 W. 117th Street
New York 27, New York
April, 1954

A SURVEY OF THE THEORY OF GAMES

by

R. Duncan Luce and Howard Raiffa

Behavioral Models Project
Columbia University
1954

Part I

General Introduction

In all of man's written record there has been a preoccupation with situations in which there is a conflict of interest; possibly only the subjects of God, love, and inner struggle have received comparable attention. The scientific study of interest conflict, in contrast to its description or its use as a dramatic vehicle, comprises a small, but growing, portion of this literature; as a reflection of this trend we find today that conflict of interest, both among individuals and among institutions, is one of the more dominant concerns of at least several of our academic departments: economics, sociology, political science, and other areas to a lesser degree.

It is not difficult to characterize in an imprecise way the major aspects of the problem of interest conflict: An individual is in a situation from which one of several possible outcomes will result with respect to which he has certain personal preferences. However, though he may have some control over the variables which determine the outcome, he does not have full control. Sometimes this is in the hands of several individuals who like him have preferences among the possible outcomes, but who in general do not agree in their preferences. In other cases, chance events (which are sometimes known in law as "acts of God") as well as other individuals (who may or may not be affected by the outcome of the situation) may influence the final outcome. The types of behavior which result from such situations have long been observed and recorded, and it is a challenge to devise theories to explain the observations and to formulate principles which should guide

intelligent action.

The literature on such problems is so vast, so specialized, and so rich in detail that it is utterly hopeless to attempt even a sketch of it. However, the attempt to abstract a certain large class of these problems into a mathematical system forms only a small portion of the total literature; in fact, aside from sporadic forays in economics, where for the most part attempts have been made to reduce it to a simple optimization problem which can be dealt with by the calculus, or in more sophisticated formulations by the calculus of variations, the only mathematical theory so far put forth is the theory of games, our topic here. In some ways the name 'Game Theory' is unfortunate, for it suggests that the theory deals with only the socially unimportant conflict of interest found in parlor games, whereas it is far more general than that. Indeed, von Neumann and Morgenstern entitled their now classical book The Theory of Games and Economic Behavior, presumably to forestall that interpretation, although this does not emphasize the even wider applicability of the theory.

The modern mathematical approach to interest conflict - game theory - is generally attributed to von Neumann in his papers of 1928 and 1937 [ ]; although recently Frechet has raised a question of priority by suggesting that several papers by Borel in the early '20's really laid the foundations of game theory. These papers have been translated into English and republished with comments by Frechet and von Neumann [ ]. While Borel gives a clear statement of an important class of game theoretic problems, it is pointed out by von Neumann that he did not obtain one crucial result - the minimax theorem - without which no theory of games can be said to exist. In fact, Borel conjectured that the minimax theorem is

false in general, although he did prove it is true in certain special cases. von Neumann proved it true under general condtions.

Of more interest than a debate on priority is the fact that neither group of papers - the one in France and the other in Germany - attracted much attention on publication. There are almost no other papers than those mentioned before the publication in 1944 of the book by von Neumann and Morgenstern, and those were confined to the mathematical journals. Apparently no interest was stimulated in the empirical sciences most concerned with conflict of interest. Fortunately, von Neumann and Morgenstern attempted to write their book so that a patient scientist with limited mathematical training could absorb the motivation, the reasoning, and the conclusions of the theory; judging by the acclaim and interest evidenced in non-mathematical journals, as well as in the mathematical ones, they were not without success in this aim. Only a very few scientific volumes as mathematical as this one have attracted as much attention and admiration, and yet we know that much of the material had lain dormant in the literature for two decades. One can only speculate on the sociological factors at work to alter the response, but presumably the recent war may have been one of the most important. During that period there developed a considerable interest in a scientific, or at least systematic, approach to problems which previously had been considered the exclusive province of men with "experience." These include such topics as logistics, submarine search, air defense, etc. Game theory certainly fits into this trend, and it is probably the most sophisticated theoretical structure so far resulting from it. The sustained activity and interest in game theory is in some considerable measure attributable to

the RAND Corporation, which is itself very much a product of the same war and postwar phenomena.

It is also of interest, though not directly relevant to the theory itself, that game theory is primarily a product of mathematicians and not of scientists from the empirical fields. In large part this results from the fact that the theory was originated by a mathematician and was, to all intents and purposes, first presented in book form as a highly formal (though, for the most part, elementary) structure, thus tending to make it accessible as a research vehicle only to mathematicians. Indeed, the total impact of game theory has been greater in mathematics than in the empirical sciences, where its techniques, though no longer its results, have caused a not inconsiderable revolution in the formulations of mathematical statistics.

Game theory does not, and probably no mathematical theory could, encompass all the diverse problems which are included in our brief characterization of conflict of interest. In this introduction we shall try to cite the main features of the theory and to present some substantive problems included in its framework. The reader will easily fill in examples not now in the domain of the theory, and as we discuss our examples we shall point out some other important cases which are not covered.

First, with respect to the possible outcomes of the given situation, it is assumed that they are well specified and that each individual is able, either directly or indirectly, to assign a numerical utility (to all intents and purposes a money value) to each of them in such a fashion that one with a larger numerical utility is preferable to one with a smaller utility. Thus, the assumed individual desire for the preferred outcomes becomes, in game theory, a maximization problem with respect to a numerical utility

defined over all possible outcomes.

Second, the variables which control the possible outcomes are also assumed to be well specified, that is, one can precisely characterize all the variables and all the values which they may assume. Actually, one may best think of these variables as grouped in n! classes if there are n individuals in the situation, or in the terminology of the theory, if it is an n-person game. To each person is associated one of the classes, which represents his domain of choice, and the one left over is within the province of chance.

As we said earlier, in this type of conflict situation we are interested in only some of the resulting behavior. Actually, our curiosity may encompass all of it - the tensions resulting, suicide rates or frequency of nervous disorder, aggressive behavior, withdrawal, changes in personal or business strategy, etc. - but of these, any one theory will, presumably, deal with only a small subset. At present, game theory deals with the choices people may make, or, better, the choices they should make in a sense to be defined, in the resulting equilibrium outcomes, and in some aspects of the communication and collusion which may occur among sets of players in their attempts to improve their outcomes. While much of what is socially, individually, and scientifically interesting is not a part of the theory, certain important aspects of our social behavior are included.

A theory such as we are discussing cannot come into existence without assumptions about the individuals with which it purports to be concerned. We have already stated one: each individual strives to maximize his utility. Care must be taken in interpreting this assumption, for a person's utility function may not be identical with some numerical measure given in the game.

For example, poker is a game with numerical payoffs assigned to each of the outcomes when it is played for money, and one way to play the game is to maximize one's expected outcome, but there are players who enjoy the thrill of bluffing for its own sake and they do so with little or no regard to the expected payoff. Their utility functions cannot be identified with the game money payments. Indeed, there are those who feel that the maximization assumption itself is tautological, and that the empirical question is whether or not a numerical utility exists in a given case. Assuming maximization of a numerical utility, it is quite another question how well the person knows the function, i.e., the numerical utility, he is trying to maximize. Game theory assumes he knows it in full. This, and the kindred assumptions about his ability to perceive the game situation, are often subsumed under the phrase "the theory assumes rational players." Though it is not apparent from some writings, the term "rational" is far from precise, and it certainly means different things in the different theories which have been developed, but loosely, it seems to include any assumption one makes about complete knowledge on the part of the player in a very complex situation, where it is known from experience that any human being would be far more restricted in his perceptions. The immediate reaction of the empiricist seems to be that such assumptions are so at variance with known fact that there is little point to the theory, except possibly as a mathematical exercise. We shall not attempt a refutation so early, though we feel we have given some defense in the body of the report. Usually added to this criticism is the patient query: why does the mathematician not use the culled knowledge of human behavior found in psychology and sociology when formulating his assumptions? The answer is simply that, for the most part, this knowledge is not in a

sufficiently precise form to be incorporated as assumptions in a mathematical model. Indeed, it is to be hoped that the unrealistic assumptions and the resulting theory will lead to experiments designed in part to improve the descriptive character of the theory.

In summary, then, one formulation of the game theoretic situation is the following: There are n players which we may label by the integers 1,2,...,n. Each player i will be required to make one choice, let us call the one he makes s_i, from a set S_i of possible choices, and these choices will be made without any knowledge of the choices of the other players. The set S_i, the domain of possible actions of player i, may include as elements such things as "playing an ace of spades," or "to produce tanks instead of automobiles," or, more important, a strategy covering the actions to be taken in all possible eventualities (see below). Now, given the choices of each of the players, i.e., the elements $(s_1,s_2,\ldots,s_n)$ of the product space $S_1xS_2x\ldots xS_n$, then there is a certain outcome, utility, for each of the players. Clearly, the outcome is a function of the element selected in the product space $S_1xS_2x\ldots xS_n$ and so it may be denoted $M_i(s_1,s_2,\ldots,s_n)$, where i runs from 1 through n. The function M_i is real-valued and it prescribes the utility to player i of the outcome of the situation. This characterization of the game we shall come to know as the normalized form of the n-person game. Two other forms - the extensive and the characteristic function form - will play important roles in our subsequent discussion; but there is no need to go into that now.

Next we should consider what significant problems of conflict of interest are included in this formulation. Our brief examination will cover four areas: economics, parlor games, military problems, and politics. One

basic economic situation involves several producers, each attempting to maximise his profit but each having only a limited control over the variables which determine it. One producer will not have control over the variables controlled by another producer, and yet these variables may very well influence the outcome for the first producer. One may object to treating this as a game on the grounds that the game model supposes that each producer i makes one choice from a domain S_i of possible choices, and that from these single choices the profits are determined. But it is obvious to all that this is not the case, else industry would have little need for boards of directors and the many elaborate executive apparatus. Rather, there is a series of decisions and modifying decisions which depend on the choices and the timing of other members of the economy. However, in principle, it is possible to imagine that an executive foresees all possible contingencies and that he describes in detail the action to be taken in each case instead of meeting each problem as it arises. By "describe in detail" we mean that the further operation of the plant can be left in the hands of a clerk or a machine and that no further interference or clarification will be needed from the executive. For example, in the game tic-tac-toe, it is perfectly easy to write down all different possible situations which may arise and to specify what shall be done in each case (and for this reason it is considered by adults to be a dull game). Such a detailed specification of actions is called a (pure) strategy. There is, of course, no reason why the domains of action S_i need be minor decisions; they may have as elements the various pure strategies of the players. Looked at this way, a player chooses a strategy which covers all possible specific circumstances which may arise. For practical reasons, it is generally not possible to specify economic strate-

gies in full, and as a result a business strategy is usually in practice only a guide to action with respect to pricing, production, advertising, hiring, etc., which does not state in detail either the conditions or the actions to be taken. The game theory notion of strategy is an abstraction of this ordinary concept in which it is supposed that no ambiguity remains with respect to either the conditions or the actions, and it serves the function of eliminating the apparent difficulty in applying the game theoretic model to economic problems. The notion of a pure strategy, and some related concepts, will receive considerably more discussion in part III.

A more important difficulty obtains in most economic problems which prevents them from being put in game form, except approximately. In general, it is not possible to specify the spaces S_i, the strategy spaces. This is not merely a practical difficulty, as suggested above, but it is in many cases not even possible in principle, for tomorrow's new invention or scientific discovery may open a whole new range of activities to one producer. How can such a possibility be imbedded in a theory? One can only hope to obtain limited prediction when such a possibility exists, using the present spaces S_i. This seems to be regarded by many social scientists as a terrible inadequacy, and yet it is a common difficulty in all of physical science. It is analogous to a physical prediction based on a physical theory and certain boundary conditions, which is surely invalidated if the boundary conditions are changed, either externally or through the very process which is being predicted. In many ways, social scientists seem to want from a mathematical model more comprehensive predictions of complex social situations than have ever been possible in applied physics or engineering; it is almost certain that their desire will never be fulfilled, and so either their aspira-

tions will be changed or formal deductive systems will be discredited as far as they are concerned.

To turn from economics, it is well known that for parlor games there is always a clear-cut scoring procedure. In some games which are played for money, such as poker, there is a finely graded numerical scale assigned to the outcomes. In others, such as chess, the outcome is simply winning or losing, but one can assign a more or less arbitrary numerical scale, such as 0 or 1. Very often it is the aim of the player to maximize his expected gain as described by the numerical score of the game; but, as we pointed out earlier, there are cases when this score function cannot be identified with the person's utility, such as when an adult purposely loses to a child.

In a parlor game, as in our economic example, each player makes not one choice but a whole series whose order and nature depend upon the previous choices both he and the other players have made, that is, on the previous play of the game. In exactly the same way as in the economic situation one is able to show that the strategy notion allows the reduction of this extensive form to the above-mentioned normal form. In part III we shall do this in some detail. It should be pointed out here that while parlor games have been characterized in extensive form and while it has been rigorously shown that any such game can be put in normal form, the corresponding statement of the extensive form of an economic situation and its reduction to normal form has not been given. The argument that game theory is applicable to such economic situations is therefore by analogy and so is no more than heuristic. Apparently a major difficulty in describing the extensive economic model is the role played by time and the timing of decisions.

There is not, as in parlor games, any fixed sequence of decisions; timing in a production situation is often as important as the decision itself.

Another difference between the parlor game and the economic problem is of the utmost importance in the theories developed. It is almost always a part of the rules, or at least of the social mores, that there shall be no collusion among the players of a parlor game. In economics, the concept of a coalition, i.e., of collusion among some of the producers so that each betters his position at the expense of the other producers or the consumer, is widely recognized in theory, in the law, and in everyday discourse. It thus behooves a theory of games which purports to have application beyond parlor games to be concerned with this common phenomenon of conflict situations.

A military conflict is, by definition, a conflict of interest in which neither side has complete control over the variables determining the outcome, and in which the outcome is determined through a series of battles. We may naively take the outcome to be winning or losing, to which we might assign the numerical values 1 and 0. More subtle interpretations of the outcomes are obviously possible, based on, say, the degree of destruction, etc. Again we have the same two difficulties as in the economic problem: there is actually a series of decisions on each side, the timing of which is of vital importance, and the domain of choices for these decisions is not usually well specified. The first problem can be surmounted as before by the notion of a strategy, and indeed the concept of a military strategy is common, even if it is not always clearly formulated. The second problem is again more profound, and it appears to prevent a game theoretic analysis

of many important military situations; but certainly other important ones are subject to the theory. One of the simplest is the "duel", which in its simplest form consists of two players 1 and 2 having p and q "shots" respectively. For each player i there is a given function $p_i(t)$ which gives the probability density that a shot fired at time t will result in a "hit", let us suppose a fatal hit. We may suppose that the domain of t is limited, as it would be in an air engagement by fuel supply. The problem is then to determine when each player best take each of his shots, assuming that he knows how many shots his opponent has already taken, so as to maximize the probability that he will hit his opponent before being hit. For most duel situations of interest, $p_i(t)$ is a monotone increasing function, as, for example, in the classical duel of two men walking towards each other with guns leveled.

It is hardly necessary to labor the point that political situations involve conflicts of interest. In addition to the difficulties of the economic and military problems with respect to ill-defined domains of action, we know that here there is considerable ambiguity as to the outcome, or payoff, function even over a known domain of possible actions. This is to some extent true in the other situations we have described, but it is overwhelmingly obvious in the political realm, where, for example, the defeat of a candidate has sometimes been attributed (after the fact) to a single sentence out of the thousands he spoke in a campaign. (There is a case of an American orator reading, one supposes for the first time, a speech in which 1776 came out "one thousand, seven hundred, and seventy-six".)

From the above comments we see that there is some hope that the normalized form of a game includes some socially important phenomena, but it

is clear that with respect to many situations there are serious difficulties. This, however, is not the entire picture. In developing the n-person game theory, von Neumann transformed the normal form into a mathematically simpler structure, simpler in that much of the detail of the normal form is condensed, which, it appears, will allow a broader application of the theory than the above discussion suggests. This is more appropriately discussed in part III than here, and we shall content ourselves with remarking that to attain this application approximate estimates of the "characteristic function" will have to be obtained, presumably by empirical techniques. This does not appear to be beyond the scope of some of the techniques under development in social psychology and sociology, and it is to be hoped that some empiricists will be attracted to this problem. However, this is conjectural, and we have the historical fact that many social scientists have become disillusioned with game theory. Initially there was a naive bandwagon feeling that game theory solved innumerable problems of sociology and economics, or that, at the least, it made their solution a practical matter of a few years' work. This has not turned out to be the case.

What then is the significance of game theory to the social scientist? First, because there has not been a plethora of applications in 10 years, it is not clear that it will not ultimately be vital in applied problems. Judging by physics, the time scale for the impact of theoretical developments to be felt is often measured in decades. Second, while the present form of the theory may not be totally satisfactory - in part, presumably, because of its so-called normative character - this does not necessarily mean that abandonment by the social scientist is the only possible course. Much of the theory is of very general importance, but some revision may be

required for fruitful applications. Attention to the theory is needed; and not attention from the mathematician alone, as is now the case. Third, game theory is the first example of an elaborate mathematical development centered solely in the social sciences. The conception derived from non-physical problems, and the mathematics - for the most part elementary in the mathematical sense - was developed to deal with that conception. The theory draws on known mathematics according to need - on set theory, on the theory of convex bodies, etc; furthermore, new mathematics was created when it was not already available. Most other attempts at mathematization (with the exception of statistics which plays a special role) have tended to take over bodily small fragments of the mathematics created to deal with physical problems. If we can judge from physics, the main developments in the mathematization of the social sciences will come - as in game theory - with the development of new mathematics, or significantly new uses of old mathematics, suited to the problem. No one of these theories should be expected to be a panacea, but their cumulative effect promises to be revolutionary.

It is the singular genius of the von Neumann and Morgenstern book that in this, the first major publication on the subject, we find a clearly formulated abstraction of considerable breadth, drawn from the relatively vague social sciences, and an elaborate and subtle superstructure developed with masterful scope - a rarity in science. The depth of their contribution can be appreciated, in part, from the fact that today the material still must be presented according to their outline; there have been additions, true, but the main conceptions are unchanged.

A word about the organization of this report. The main body is divided into three parts, the first devoted to games having two players,

the second to general games, and the third to miscellaneous topics in one way or another closely related to game theory. The division of the first two parts is dictated by von Neumann's development of the theory, in which the study of n-person games rests on the already completed study of the 2-person games. In addition, the subsequent contributions to the theory have in the main continued this dichotomy, and so the material is most easily presented in two parts.

We do not in any way intend this report to serve as a text on the subject, nor as a research reference; rather, we hope to lay bare, with a minimum of mathematical notation, the main structure of the theory, the assumptions, and the conclusions. A consequence of this aim was our decision to omit all proofs. It is a report directed toward the social scientist who has found the long chains of argument in von Neumann and Morgenstern too tedious and the crisper style of McKinsey too spare, but who still would like to know the principal features of this important theory. Anyone interested in pursuing research in game theory, or in its applications, will have to consult at least one of these two books and some of the research papers referred to, but his task may be simpler - at least we hope it will be - for having read this less technical outline.

A SURVEY OF THE THEORY OF GAMES

by

R.Duncan Luce and Howard Raiffa

Part III: n-Person Games

Behavioral Models Project
Bureau of Applied Social Research
Columbia University
1954

Table of Contents

1. Introduction

The theory of games would be a very incomplete edifice, both esthetically and practically, if it were restricted to the 2-person case. It is not. In this part of the report we therefore turn to an examination of the n-person theory, which is, in the main, very different from the 2-person theory.

Intuitively, it is reasonable to suppose that the two most significant notions of the 2-person theory - strategies and equilibrium points - can be extended to games with more than two players. This we shall discuss in section 3. Were this extension of definitions and the resulting theorems the totality of n-person theory, we should have presented it in a unified manner for all $n \geq 2$. However, it has long been recognized in sociology, and in practical affairs, that between two-person situations and those involving three or more persons there is a qualitative difference which is not as simple as the difference between 2 and 3. Georg Simmel writes, "The essential point is that within a dyad, there can be no majority which could outvote the individual. This majority, however, is made possible by the mere addition of a third member." [26, p. 137] And again, "The typical difference in sociological constellation, thus, always remains that of two, as over against three, chief parties." [26, p. 144] The recognition of this feature - that of coalitions in the language of von Neumann and Morgenstern [21] - has resulted in an n-person theory markedly different from 2-person theory.

A major obstacle to developing a satisfactory theory of coalitions is that in the present formalizations of a game no explicit provisions are

made about communication and collusion among the players. Thus any theory of collusion, or of coalition formation, has a distinctly ad hoc flavor. The difficulties in making explicit assumptions about communication appear, at least superficially, to stem from the variety of rules which are found in empirical situations. Collusion in parlor games is prohibited by social sanctions and by a sense of sportsmanship; that the rules are well heeded is, one supposes, because so little is at stake. Of course, there are known exceptions in the history of gambling. In the economy one finds the whole gamut from no rules at all, through moral sanctions, to elaborate legal codes as in the anti-trust laws. In international affairs, coalitions and their disruption bulk large in the history of at least the past 300 years, the rules obeyed have been few.

One point of view which has been presented, and which we shall discuss in section III.3, is to the effect that non-cooperative games are theoretically basic, and that cooperative ones can and should be subsumed under that theory by making communication and bargaining formal moves of the non-cooperative game. This view has never been fully elaborated and so criticism is difficult, but McKinsey has pointed out, "It is extremely difficult in practice to introduce into the cooperative games the moves corresponding to negotiations in a way which will reflect all the infinite variety permissible in the cooperative game, and to do this without giving one player an artificial advantage (because of his having the first chance to make an offer, let us say)." [13, p. 359]

In addition to the conceptual complications of collusion, there are inherent practical complications as n gets larger, for the number of pos-

sibilities increases at a fantastic rate; the difficulty of a detailed analysis of a 2-person game such as chess is minor compared to a similar analysis of most n-person games. One of the principal features of the current theory is to by-pass such a detailed analysis. That we can successfully avoid at the conceptual level the combinatorial problem does not seem to solve all empirical problems of verification, for an empirical study must deal with specific games in all their complications. Fortunately, what we suggest does not seem to cover the issue entirely, but we must postpone more discussion to section III.7.

The principal order of our presentation - extensive form of a game, normalized form, characteristic function, and solutions - is essentially that of von Neumann and Morgenstern [21]. The work on n-person games since the publication of their book has been centered primarily within this framework, and while there have been criticisms of this general organization,* no new approach has been commonly accepted. It must, however, be added that the vast majority of the work in the 10 years since the first printing of their book has been devoted to the 2-person game and to extensions of it which are either beyond the scope of this report or are studied in Part IV; the number of papers on the n-person game is less than a score. Several facts may be mentioned which seem relevant to this phenomenon: the relation of the 2-person game to linear programming and to statistics has attracted considerable attention; mathematicians have been intrigued by the current 2-person theory because it draws on more advanced mathematics

---

* Von Neumann and Morgenstern themselves raised objections and questions about the organization to which they were forced, and they suggested that when the theory is more mature we may find it unified for all $n \geq 2$ and find the now important characteristic function only an unnecessary technicality. [21, p. 606 - 608, particularly p. 608]

than does n-person theory; many workers have felt dissatisfied with the present formalization of n-person theory and rather than meeting the conceptual challenge they have withdrawn to other issues. Nonetheless, it is the n-person theory which is of greatest interest in sociology and economics, and it is here, more than in 2-person theory, that game theory as social science, though not as mathematics, will stand or fall. In two principal ways we shall try to show that general game theory can be of interest to social scientists. First, we shall emphasize the independence of the characteristic function theory from its derivation in terms of the normal form of a game, and we shall suggest the possibility of an empirical determination of these functions in real situations where the strategy spaces and the payoffs of the normal form are difficult or impossible to determine. Second, we shall offer the view, with examples as support, that it is possible to devise theories based on the characteristic function which are more relevant to social science than the solution notion, a concept which has not found wide acceptance outside mathematical circles. We must, however, emphasize that the theory is in far from final form and that the social scientist will find as many difficulties unearthed as are solved in any attempt to make game theory an applied theory.

2. Extensive Form

The mathematical abstraction of a game assumes three forms in the presentation of von Neumann and Morgenstern [21]. The first - called the extensive form - is our present topic; it is an attempt to capture the salient features of a game, such as a parlor game. From this is derived

another more compact form known as the normal form, which we shall discuss in section III.3. It is curious that, while the normal form is a special case of the extensive form (to which every extensive form may be reduced), only the normal form has the apparent or psychological generality to encompass many social and economic problems. There are few situations other than parlor games which are games in extensive form, excluding the case of the normal form. The third stage of their development is the derivation of a real-valued set function, called the characteristic function, which represents coalition strengths (section III.4). While in one sense a characteristic function represents a game, it need not, and so it turns out that in a very important sense to social science the theory of characteristic functions is more general than the theory of games in extensive and normal form. The principal mathematical theory is at the level of characteristic functions, and so it could be presented mathematically without reference to the extensive and normal forms, but it is appropriate for us to follow the longer development in order that the reasons for the abstractions should be clear.

Any parlor game is composed of a series of well-specified moves, where each move is a point of decision for a given player among a set of alternatives. The particular choice a player makes at a given choice point is called a choice, but the fact that he must make a choice coupled with the set of alternatives for the choice is called a move. A sequence of choices, one following another until the game is terminated, is called a play. Let us suppose that in one game (at some stage of a play) player 1 has to choose

among playing a king of hearts, a two of spades, or a jack of diamonds, and that in another game a player, also denoted 1, has to choose among passing, calling, or betting. In each case the decision is among three alternatives, which may be abstracted by a drawing as in Fig. 1.

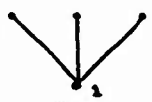

Fig. 1

But how can these two examples be considered the same? Certainly it is clear from common experience that one does not deal with one three-choice situation in the same way as any other three-choice situation. One might, were they given out of context, for there would be no other considerations to govern the choice; but in a game there have been all the choices preceding the particular move, and all of the potential moves following the one under consideration. That is to say, we cannot truly isolate and abstract each move separately, for the significance of each in the game depends on some of the other moves. However, if we abstract all the moves of the game in this fashion and indicate which choices lead to which moves, then we shall know the abstract relation of any given move to all other moves which have affected it, or which it may affect.

Such an abstraction leads to a drawing of the type shown in Fig. 2. The numbers associated with the moves indicate which player is to make the move, and therefore they run from 1 through n. The set of players, i.e., the first n integers, will be denoted I_n. In the example of Fig. 2, $n = 4$. But we have

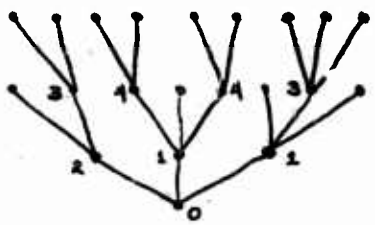

Fig. 2

denoted the first move 0. In addition to moves by players, some games have chance moves, as, for example, the shuffling of cards prior to a game of poker. Such moves, which need not be the first move of the game, are assigned to the "player" 0, which stands for chance. For the 0 moves we must be given a probability distribution, or weighting, of the several alternative choices.

A drawing such as Fig. 2, when considered abstractly as a mathematical system, is known as a graph. A graph consists of a collection of points (called nodes) and branches (the lines between some pairs of nodes drawn in on the figure) between certain pairs of nodes. If there is at most one branch between any pair of nodes, as in the case of a game, a graph is isomorphic to a symmetric relation over the set of nodes. A graph may have closed loops of branches, such as abc or abdec in Fig. 3. A graph with no such loops of branches is called a tree. The graph of a game is a tree, which is called the game tree. This may not seem reasonable for in such games as chess one can arrive at the same arrangement of pieces on the board by several different routes, which appears to mean that closed loops of branches can arise. However, in game theory we choose to distinguish two moves as different if they have different past histories, even if they have exactly the same possible future moves and outcomes. In games like chess this distinction is not really important and to make it appears to be an arbitrary decision, but in many ways the whole conceptualization and analysis of games is simplified if it is made. The tree character of a game

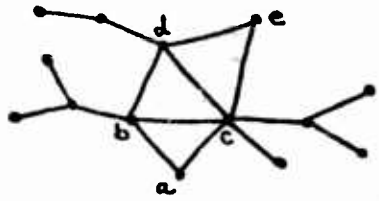

Fig. 3

is not unrelated to the common sinking feeling one often has after making a stupid choice in a game, for, in a sense, each choice is irretrievable, and once it is made, there are parts of the total game tree which can never again be reached.

The tree is assumed to be finite, in the sense that a finite number of nodes, and hence of branches, is involved. This is the same as saying that there is some finite number σ such that every possible play of the game terminates in no more than σ steps. This is certainly true of all parlor games, for there is always a "stop" rule to terminate stalemates, e.g., as in chess. To say the tree is finite is not to say that it is small and easy to work with. For example, card games often begin with the shuffling of a deck of 52 cards, and so the first 0 move has 52! or approximately 8.07×10^{67} branches stemming from it. Clearly, for such games no one is going to draw the game tree in full detail!

At the end of each play of the game certain rewards and punishments - payoffs - occur. These may be the subjective reward of saying "I won" or the monetary punishment of seeing someone else sweep in the pot - a pot which often includes more of your money than it should - or, as McKinsey says, "the death in Russian Roulette." Each of the end points in the game tree is a possible termination point of the game and it completely characterizes the play of the game which led to that point. We may index these end points and denote a typical one by the symbol α. Now to each α and for each player i we have a payoff function which we may denote $M_i(\alpha)$, that is, a function which has as its domain the plays - or

end points - of the game. The range of the functions can vary a great deal, as we suggested. In some cases the range is the real number system, such as for monetary payoffs. In Russian roulette the range is a space having two elements, one representing death, the other not-death. While in principle it is not necessary to restrict the possible ranges of the functions $M_i(\alpha)$, to make any progress at all in the theory of games it is necessary to assume that the ranges are part of the real number system, usually a finite part. Further, it must be possible to form sums such as $M_i(\alpha) + M_j(\beta)$ and to be able to assign some meaning to them. Von Neumann and Morgenstern [21, p. 617, Appendix] have developed a theory of utility in which they show that certain assumptions about the ability of people to assign preferences among certain alternatives make it possible to assign a numerical utility to the various alternatives. This work is discussed very briefly in section III.7.2. From this they conclude that the restriction of the ranges to the real number system is not really such a serious restriction, after all. But more must be assumed in order to justify forming sums of payoffs. The assumptions are summarized by saying that utility must be both numerical and transferable. The need to form sums will arise later in the theory of coalitions when we wish to allow side payments: A player who will receive a certain amount from the play of the game can be induced by other players to participate in a coalition by their offering him added payments other than those provided by the rules of the game. If the rules of the game provide death for certain players, this payoff is not transferable, even though a numerical value might be assigned. Essentially then, we will have to think of the payoffs in terms of some infinitely divisible extra-player commodity, which to all intents and purposes is money. Without this assumption we would not go far.

The final step in the formalization of a game is to indicate what each player knows when he makes a choice at any move. It is assumed that each player is omniscient in that he knows the entire game tree and all of the payoffs, but there is the possibility that the rules of the game do not provide him with knowledge on any particular move of all the choices made prior to that move. This is certainly the situation in most card games which begin with a chance move, or where certain cards are chosen by another player and are placed face down on the table, or where the cards in one player's hand are not known to the other players. Indeed, it may be that a player at one move does not know what his domain of choice was at a previous move! The most common example of this is bridge where the two partners must be considered as a single player who intermittently forgets and remembers what alternatives he had available on previous moves (see sections III.3 and III.4).

It is possible, in principle, for such a super-intelligent player to ascertain from all the information known to him and from the rules of the game a certain minimum set of moves of which his is one, but which one he is not certain. Since he knows the game tree in advance, it is thus possible from the rules of the game to characterize these indistinguishable moves in advance. Abstractly, there are only two necessary features to these sets of moves - which are known as <u>information sets</u>. Each of the moves in the set must be assigned to the same player, and each of the moves must have exactly the same number of alternatives. For if one move has r alternatives and another s, where $s \neq r$, then he would only need to count the number of alternatives he actually has in order to eliminate

the possibility of being at one move or at the other. In the graphical presentation of a game, the information sets may be indicated by enclosing the nodes comprising each set by a dotted line, as in Fig. 4.

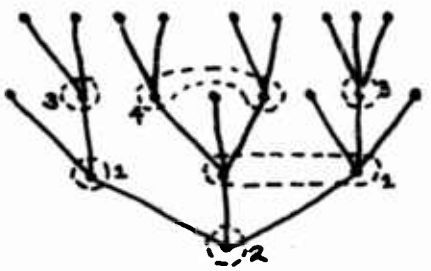

Fig. 4

When an information set consists of a single move, the player is totally informed in that he knows exactly where he is on the tree. When all moves are of this type, we say the game has <u>perfect information</u>. Tic-tac-toe and chess are examples of games with perfect information.

In summary, then, the abstraction of a game which is called the <u>extensive form</u> consists of

i. a finite tree (which describes the relation of each move to all the other moves),

ii. a set of payoff functions $M_i(\alpha)$ (one for each player and defined over the end points - or plays - of the game),

iii. a partition of the nodes of the tree into n+1 sets (which tell which of the n players or chance takes each move),

iv. a probability distribution over the branches of each 0 move,

v. a refinement of the player partition into information sets (which describes the ambiguity each player has when he takes each move).

The original description of a game in extensive form is due to von Neumann and Morgenstern, but it differs somewhat from and is less compact than this one, which was given by Kuhn. [8]

In this extensive description of a game all the subtle differences between games are apparent. No matter how intuitively similar two games are, if there is a formal difference as given by the rules it will show up at this level. At this detailed level the problems of analysis seem over-whelming, and certainly to date there has been very little work done on the problem of games in extensive form. The principal result in this area will be presented in the next section. It is not even clear how much work is desirable at this level, for while such work might result in a very adequate theory of parlor games, very few examples from economics or social science fall into the pattern of a tree, that is, of well-specified, temporally ordered moves. Were the theory entirely at this level of abstraction, it could be of deep interest only to theorists of parlor games and gambling. However, any game in extensive form can always be put into a special extensive form, called the normal form, which, as we shall see in the next section, makes it clear that game theory encompasses problems of more general interest and depth. It is only at the level of the normal form, and at an even more abstract level, that game theory seems to have the potential of far-reaching impact on the social sciences.

So far we have described what we shall mean by a game, but a theory about games can be developed only as an answer to questions about them. One general class of problems has been raised: if we assume "rational" players who are omniscient in that they know in full the game tree, the payoffs, and the information sets, if we assume the payoffs are in a numerical and transferable utility, and if we assume that each player wishes to maximize his expected return (in utility units) from playing

the game, then what happens? This, as we shall come to see, is not the precise problem it may seem to be, and we will have to specify further what we mean. Unfortunately, the further specification appears to take one beyond the assumed structure of the game.

3. Normal Form*

3.1 Strategies

One way to ascertain the outcome of a game in extensive form is to let the players play it and observe the outcome. Indeed, many would say this is the only way, but they would be wrong, for in principle we could cause each player to state in advance what he would do in each situations which might arise in the play of the game. From this information for each of the players, an umpire could carry out the play of the game without further aid from the players and thereby determine the payoffs. Such a prescription of decision for each possible situation is known as a pure strategy for a player.

For many games the actual preparation of a pure strategy in a form an umpire could use without ambiguity is a hopeless task; however, certain simple examples of pure strategies are easily given, though in general they would be poor ways to play. For example, if we suppose that each branch stemming from a move is given a number, $1,2,\ldots,r$, where r is the number of branches, then one pure strategy is always to take branch 1. Another is always to take the branch with the largest number. Indeed, if

---

* The last subsection of this section, and of sections III.4, III.5, and III.6, is a brief summary of the principal concepts of the section. While these summaries are not intended to be intelligible without a first reading of the preceding sections, they may assist some readers to grasp the main line of development in the section as it is being read.

player i has q different information sets, which we may number 1,2,...,q, then any pure strategy can be represented by a set of q numbers, where each number represents the branch chosen when, and if, the play reaches that information set. Thus q-tuples of integers,

$$\|y_1, y_2, \ldots, y_q\|$$

represent pure strategies. For example, the strategy in which branch 1 is always taken is represented by

$$\|1, 1, \ldots, 1\|$$

But each y_i has as its range only a finite number of integers, since each move has only a finite number of branches, and there are only a finite number of y's, namely q, so there are only a finite number of strategies. Without any loss, we may label the strategies by numbers 1,2,...,t, where t is the total number of strategies available to the player. The number t is finite, but it need not be small. A game having but 10 information sets for a player and 10 branches at each set is exceedingly simple, but there are 10 billion different strategies for that player.

We let s_i be a variable which has as its domain the available strategies, or more exactly, the integers which stand for them, of player i. Now, as we pointed out, when each player has selected a strategy s_i, then an umpire is in a position to play the game and to determine the payoffs. That is to say, from the given payoffs of the game in extensive form we may determine a payoff function defined over a domain which is the product space of the n sets of strategies. First, if there are no chance moves in the extensive game, then the selection of the strategies $(s_1, s_2, \ldots, s_n)$

determines a play α and so we define

$$M_i(s_1,s_2,\ldots,s_n) = M_i(\alpha).$$

If, however, there are chance moves, then the selection of the strategies $(s_1,s_2,\ldots,s_n)$ does not uniquely determine a play, but rather there is a probability distribution $p(\alpha)$ over all possible plays (of course, the probability for some plays may be 0). Now, as the payoff associated with the strategies $(s_1,s_2,\ldots,s_n)$ we take the expected value* over all the plays, that is,

$$M_i(s_1,s_2,\ldots,s_n) = \sum_{\alpha} p(\alpha)M_i(\alpha).$$

As we have seen in Part II, for $n = 2$ we may always represent the payoff function as a matrix with the rows representing the strategies of player 1 and the columns the strategies of player 2. Clearly, for $n > 2$ a simple matrix will not do, but we may think of the function as a matrix in n-dimensions, with the i^{th} coordinate giving the strategies of player i.

Observe that by means of the strategy notion every game in extensive form has been reduced to a game of the following form: each player has exactly one move (a choice among his several strategies) and he takes his move in the absence of any knowledge about the choices of the other players. The payoff to the players is determined from the functions M_i and the values of s_i. This is a reduction of every game to a simple standard formwhich is called the <u>normal form</u> of a game.

---

* There has been some misconception that the concept of a numerical utility is not needed at this point but only when the notion of mixed strategies is introduced (III.3.2), and that as far as pure strategies are concerned we may still deal with orderings of preferences. But in the case where there are chance moves numerical utility is necessary if we are to assign a payoff to a selection of strategies, and without such a payoff the development of the theory would be blocked.

What sleight of hand is this? We began by abstracting parlor games and arrived at the extensive form of a game, which, in general, led to an oppressively complex game tree. For games of any reasonable complexity, the number of possible trees and of variations arising from different information sets immediately led us to believe that there is little hope of finding detailed classifications of games in extensive form or of analyzing player behavior at that level. Then by introducing the idea of a pure strategy we have suddenly reduced all games to a comparatively simple standard form. That is, the sleight of hand was to trade the conceptual complexity of a game tree for the numerical, but not conceptual, complexity of listing all available strategies.

The reduction of any specific game, except the simplest, to normal form is a task defying the patience of man; but, because the normal form of all possible games is comparatively simple, there is hope that one may successfully examine mathematically all possible games in normal form. The study of specific games may be close to impossible, but the classification, analysis, and determination of features of all games may be now quite feasible.

Assuming the payoff function is in terms of a numerical and transferable utility, we may make the first important classification of games which will play a role in the following sections. If there exists some constant K such that for every possible choice of the strategies s_i,

$$\sum_{j=1}^{n} M_j (s_1, s_2, \ldots, s_n) = K,$$

then the game is called <u>constant-sum</u>. If the value of K is 0, then the

game is called zero-sum. The significance of the latter notion is that as a result of a play of the game, utility (or money) is neither created nor destroyed, it is only exchanged; what one player wins is compensated for by the loss of others. All parlor games are zero-sum in their money payments, but not necessarily in a utility measure. Many economic processes, if they are games, are not zero-sum, for production carried out during the play of the game (the execution of economic processes) may mean that no player loses, though some may gain more than others. Only rarely will they be constant-sum, for the amount of production will generally vary with the strategies.

We see that the normal form of the game is exactly the general problem which was evolved and discussed in Part I: Each player has some limited control over the variables which determine what he shall receive and each of the players wishes to maximize his return. So we have again returned to our original problem, but many readers may, at this point, feel that a strange psychological trick has been played on them. The extensive form of the game, while apparently a suitable representation of parlor games, did not seem adequate for many other situations. The rigid development of one move following another is not typical of many economic decisions, though it is somewhat analogous. It is not usually possible to state in advance that industry A will make a decision among certain alternatives, and when it does, and only then, industry B will make a decision. The economic importance of timing is too well known to belabor this point, so an economic process in extension will not often be an extensive game; thus, one must ask how we can hope that anything as rigid as a game tree will represent anything other than parlor games. Then suddenly, by means of the strategy notion, we have

reduced all parlor games to the form of n players, each trying to maximize a different function of n variables, only one of which is controlled by each player. It is intuitively clear that this form is suited to socially more interesting problems.

The reduction of a specific economic situation in its "extensive" for, while not a game in extensive form, will entail the use of what are commonly called "strategies". Again, actually finding the s_i's and the M_i's for real situations is a monumental task, but it can be done in principle for a great many different situations which in their detailed or "extensive" structure are not isomorphic, even in some approximate sense, to a parlor game. These practical difficulties, no more than similar ones in the physical sciences, do not cancel the power of a theory to study all possible cases encompassed by the theory.

3.2 Mixed Strategies and Equilibrium Points

Our description above that in the normal form of a game "each player has exactly one move...and he takes his move in the absence of any knowledge about the choices of the other players" tends to be misleading. For while he may have no knowledge of the choices of the other players, it does not follow that the players have not agreed beforehand to make certain choices. It simply means that if such an agreement were reached and a double-cross occurred, none of the other players would know of it when he made his choice. Thus, we may distinguish situations where communication can occur among the players and coalitions can form prior to the play of the game from those

in which no communication is allowed. The vast amount of n-person game theory is devoted to the former case and it will be discussed in sections III.4, III.5, and III.6, and to some extent in III.7.

Certain authors, notably Nash, have felt that this is not the basic case to study, but rather games in which no communication is allowed. One cannot but be sympathetic with this view, for the possible restraints on communication, excluding complete prohibition, at least rival in complexity the rules of the game, and so they afford a problem as complex - or more so - as the one originally tackled. Nash argues [16] that whatever communication is allowable can be introduced as part of the formal structure of the game, with the bargaining as formal moves. This, in the normal form of the game, simply enlarges the domain of the various strategies and extends the payoff function. Were it possible to give an explicit and intuitively acceptable way of enlarging an extensive game so as to include communication, the argument would be more convincing. The difficulty in so doing is not unrelated to the fact that most economic situations in extensive form are not games in extensive form; timing in bargaining can often be of vital importance. In addition, if one were to treat coalition formation as moves in an enlarged extensive game, then one would lose the chance of developing a theory of coalition formation in a game situation, which may be the most interesting aspect of general game theory to the social sciences.

But whether or not we accept the belief that all games should be recast in terms of non-cooperative games, one part of the theory certainly should be devoted to non-cooperative games. Presumably it should be a "natural" extension of the (non-cooperative) 2-person theory. This is not

to say that it will not include more phenomena for general n than for $n = 2$, but only that the theory for arbitrary n should coincide with that already developed when $n = 2$. In actual fact, in its present form, which is due to Nash, the general theory of non-cooperative games does not have characteristics different from the 2-person case; Nash's contribution is an appropriate generalization of an equilibrium point and a proof of the existence of equilibrium points in games. [18]

Suppose that in a game it is possible to find a strategy for each player, say $s_1, s_2, \ldots, s_n$, such that if every player except one, say j, chooses s_j, then the remaining player cannot do better than choose s_j. That is, j's payoff for any other strategy r_j will not exceed what he will obtain by choosing s_j. Formally, we require for every j that

$$M_j(s_1, s_2, \ldots, s_n) \geq M_j(s_1, \ldots, s_{j-1}, r_j, s_{j+1}, \ldots, s_n).$$

If this is the case for every j, then we say that $(s_1, s_2, \ldots, s_n)$ is an <u>equilibrium point</u> in pure strategies. (It is a point in the n-space of pure strategies.) Thus, if the players are at an equilibrium point, it does not behoove any single player to move from it, though if several were to change together they might all improve their lot. But since no communication is allowed, it might be thought that once the players arrived at an equilibrium point they would be in equilibrium, i.e., there would be no resultant forces acting to make anyone change. However, it is not difficult to show that this concept of equilibrium in pure strategies is the same as that defined in Part II when $n = 2$, so all the objections and difficulties raised there apply here without change.

As for the 2-person case, there is no assurance in general that an equilibrium point in pure strategies exists. It is known that a sufficient condition for games to have equilibrium points in pure strategies is that they have perfect information [8] but this is not a necessary condition. Dalkey [4] has given a necessary condition, but a discussion of this would take us beyond the scope of this report.

A second way of dealing with the problem is, as in the 2-person case, to introduce the concept of a mixed strategy. In essence, the player does not tell the umpire which strategy to use, rather his instructions are to choose a strategy by a chance device according to a given probability distribution. Thus, a mixed strategy σ_i for player i is a probability distribution over his set of pure strategies s_i. We may denote the distribution as $p(s_i)$. Of course, the given payoffs $M_i(s_1,s_2,\ldots,s_n)$ are only defined over pure strategies, but we can extend the function to mixed strategies in a natural way, for suppose player 1 uses a mixed strategy σ_i, and each of the others uses pure strategies, then we define

$$M_i(\sigma_1,s_2,\ldots,s_n) = \sum_{s_1} p(s_1) M_i(s_1,s_2,\ldots,s_n).$$

In like manner, we may extend the notion of payoff function to the entire product space of mixed strategies. Of course, the notion of a mixed strategy has the same psychological peculiarities for n players as it did for 2.

The above definition of an equilibrium point in pure strategies can obviously be taken over with a formal substitution of σ_i for s_i to yield a definition of an equilibrium point in mixed strategies. Nash's principal theorem [18] shows that over the domain of mixed strategies every

finite game has at least one equilibrium point. This shows, in effect, that the definition is acceptable, at least in the sense that every game has such a point. It is a matter of intuitive judgment and empirical verification whether it really is acceptable in social science; many feel it is not.

3.3 Perfect Recall and Behavioral Strategies

It has probably occurred to the reader that while these notions of strategies, both pure and mixed, are fine tricks for the mathematical development of game theory, people almost never pick a strategy on such a grand scale. The domain of strategies is just too large ever to have been completely given even for most parlor games; in all the years that chess has been played and analyzed, only a small fraction of partial strategies has ever been discussed and listed, though judging by experience they include most that are really important. Nonetheless, one might wonder about a theory of games with a more limited view of the strategy notion. One of a somewhat special and limited nature has been examined and the results are of interest, for in a certain important class of games they justify a theory based on mixed strategies.

Instead of giving a mixed strategy to the umpire, a player might specify for each of his information sets a probability distribution over the alternatives at each set. Such a set of distributions is known as a behavioral strategy for the player. Now, while it is still a monumental task to list behavioral strategies for most games, it may be felt that in effect a player has such a distribution in his mind when he makes decisions during a play of the game, and by making him play it many times (after learn-

ing has occurred) and observing his choices we could get experimental estimates of these distributions.

It is reasonably clear - and it can be shown - that by using mixed strategies a player can do as well as by using behavioral strategies, and examples can be presented where he can do better with mixed strategies. Therefore, it is of interest to classify those games in which it is possible for the players to do as well using behavioral strategies as using mixed strategies. This problem was posed and solved by Kuhn [8].

The appropriate class of games turns out to be that in which each player remembers everything he did prior to each move, though he may not know what choices the other players made. Such games as bridge are to be excluded by definition, but most parlor games, if played by rational players, are included. Formally, let us suppose that V is any one of the information sets of some player i. Let Q denote any move made by i which is prior to the information set V. If there is but one branch from Q which leads to any of the moves contained in V, then when player i is in V he will recall perfectly what he did on the move Q. If this is the case for every possible i, V, and Q in a game, then we say that it is a game having <u>perfect recall</u>. It will be recalled that if every information set contains but one move, the game is said to have perfect information. Since a game is based on a tree and hence there are no closed circuits of branches, a game with perfect information is one with perfect recall, but the converse is not true, since

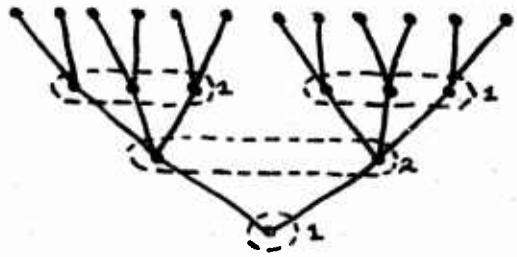

Fig. 5a

player i may not know what player j did prior to some information set U. The game tree of Fig. 5a has perfect recall but not perfect information. The only way in which the game might not have perfect recall is for player 1 to be uncertain on the third move as to what he did on the first move. This is not the case. The game in Fig. 5b does not have perfect recall because if player 1 is in the information set marked V on move three, then he cannot determine which of the two branches he selected on the first move.

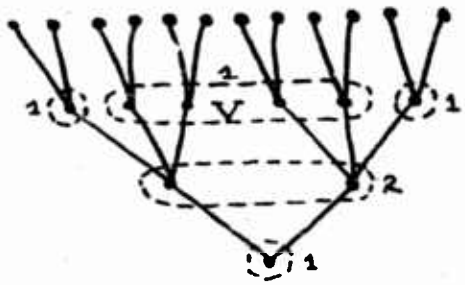

Fig. 5b

The expectation of a player using behavioral strategies is obtained from the original payoff functions, weighted according to the various probabilities in the behavioral strategy, in much the same way as we obtained the payoff for mixed strategies from that for pure strategies. With this definition it is possible to show that in games with perfect recall there are behavioral strategies which have the same expectations as the best mixed strategies. Thus for games with perfect recall it does not matter, as far as theories involving maximum expectations are concerned, whether we use behavioral or mixed strategies.

3.4 Signaling Strategies

Given that by using mixed strategies we can do as well as possible, and that for games with perfect recall the use of behavioral strategies can be as good as using mixed strategies, the question arises whether anything more can be said for games without perfect recall. Thompson [28] attacked this problem and he has given an intuitively very acceptable solution. In

effect, he shows that mixed strategies are required but only over the information sets which prevent there being perfect recall; over the other information sets one may use behavioral strategies.

We want to single out those information sets which prevent a game from having perfect recall: An information set U associated with player i is called a signaling information set if one can find an information set V of player i which follows U such that there exists a move Q in U with a path from Q to a move of V and another move Q' in U with a path to a move of V, but the branch from Q which begins the first path does not correspond to the branch from Q' which begins the second path. Thus, when player i is at the information set V he is unable to know whether he made the choice of one branch at Q or of a non-corresponding branch at Q' when he was at his information set U. As an example, consider Fig. 5b, and let U be the information set consisting of the first move of the game and V the indicated information set. If we take either branch at U, then it is easy to see that paths to V exist in both cases, so U is a signaling information set.

The term 'signaling' used here arises, presumably, from a consideration of bridge, which must be considered a 2-person game with the pairs of partners being single players. The move of one of the partners often serves to signal considerable information to the other partner (and the term 'signal' is part of the vocabulary of bridge), but it is nearly always the case that when the second partner comes to his next move he is not fully certain from what domain of possibilities the first partner made his choice. Thus, the player (=the pair) cannot at any point in

the game have full recall as to what "he" did a little earlier.

Let Δ_i be the set of all signaling information sets for player i. If $\Delta_i = \phi$ (= the empty set) for all i, then it is not difficult to see that the game has perfect recall, and conversely.

Earlier we defined the notion of a pure strategy over the set of all information sets; we can, of course, do the same thing over the set Δ_i and we call this a pure signaling strategy for player i. In like manner, a probability distribution over the pure signaling strategies of player i is called a mixed signaling strategy. These notions are exactly the same as those given in section III.3.1, except that the domain of definition is restricted to the signaling information sets rather than to all information sets.

An associated behavior strategy for player i is a mixed signaling strategy over the set of signaling information sets and behavioral strategies over each of the other information sets of player i. That is, over the information sets having perfect recall we continue to use behavior strategies, and over all other information sets we use mixed strategies. It is easy to see that for games with perfect recall associated behavior strategies are the same as behavioral strategies.

The principal result proved by Thompson [28] is that for any finite game, a player can find an associated behavioral strategy which will result in the same payoff as the best mixed strategy (and of course the converse holds - he cannot do better using an associated behavioral strategy than a best mixed strategy).

This result is of considerable importance in the examination of

specific games without perfect recall. Thompson remarks, "This theorem together with the fact that the normalized form of the game obscures signaling strategies, explains one reason why the normalized form of the game is not always the best form in which to solve actual games." [28, p. 275]

In another paper, which we cannot go into here, Thompson [29] uses the notion of signaling strategies to examine a simplified form of bridge.

3.5 Summary

For a game in extensive form, i.e., described by means of a tree of moves, information sets, etc., we pointed out that rather than making each move as it arises a player could choose a connected system of choices, one for each possible contingency in the game. Such a complete statement of actions is called a pure strategy, and to each player i one may associate the space S_i of all his possible strategies. When each player has selected a strategy, the outcome of the game, or, when there are chance moves, a probability distribution over the possible outcomes, is determined. Thus we defined the payoff function over strategies as the expected payoff over the outcomes arising from the strategies. By this means, any extensive game was reduced to the situation where each player i selects an element from the space S_i of his strategies without any knowledge of the choices of the other players. From these choices the payoffs to each of the players are determined from real-valued functions of the form $M_i(s_1, s_2, \ldots, s_n)$ where $s_i \in S_i$. Such a situation is known as a game in normalized form.

The strategy spaces S_i were then extended to spaces which include

all possible probability distributions over pure strategies, and the payoff functions were extended in the natural manner by weighting the payoffs over pure strategies according to the given probability distribution. An element of this larger space – a probability distribution over pure strategies – is called a mixed strategy.

An n-tuple of mixed strategies $(\bar{\sigma}_1, \bar{\sigma}_2, \ldots, \bar{\sigma}_n)$ is called an equilibrium point if the payoff to each player i is never increased when i selects a mixed strategy different from $\bar{\sigma}_i$ and all the other players take the strategy of the equilibrium point. It can be shown that an equilibrium point exists in mixed strategies for every n-person game, but this theorem does not hold for equilibrium point restricted to pure strategies. An equilibrium point might be expected to arise with very conservative "rational" players, but as a description of player behavior the equilibrium point is subject to the criticisms raised in Part II of this report.

Following this, the concept of a behavioral strategy was introduced. It is a set of probability distributions over the alternatives of each of a player's information sets. It was observed that a player can never do better using a behavioral strategy than using one of the best mixed strategies, but there are some games in which he can do as well by using a behavioral strategy. A class of games, those said to have perfect recall, was defined; it was observed that games of this class have the above property. In the final section, the question was raised as to over which information sets it is necessary to use mixed strategies in order to do as well as using full mixed strategies. This led to introducing signaling strategies which consist of mixed strategies over the class of so-called signaling

information sets and behavioral strategies over the remaining information sets, and the principal theorem is that there is always a signaling strategy for a player which will result in the same payoff as one of his best mixed strategies.

4. Coalitions and the Characteristic Function

4.1 The Characteristic Function

The presentation of the previous two sections has been applicable to all $n \geq 2$; whereas we argued in the introduction that there were reasons for separating the case $n = 2$ from $n > 2$, for in the latter case collusion among some of the players might occur. This section and sections III.5 and III.6 are devoted to theories of coalition formation.

Let us initially restrict our attention to zero-sum games. Suppose S is a subset of the players who have decided to form a coalition in the sense that as a group they shall decide on individual courses of action which together cause the group to do as well as possible. How the individual payments come out does not, for the moment, matter, as long as the summation of them over the members of S is, in some sense, as good as possible. One might object, however, that if it turned out that whenever the coalition did its best one of the players in the coalition did no better, or even worse, than he could have alone, then it might indeed be difficult to persuade him to remain in the coalition. As long as the payoff is in some sort of transferable utility, as we have assumed it is, this is no problem in principle, for the other members of the coalition may extend to him side

payments in order to keep him in the coalition. The extent of the side payment is a difficult problem of prediction, but it presumably depends, in part, on his contributions to the total strength of the coalition. This suggests that it may be sufficient, in developing a theory, to look at the total payment received by a coalition.

The worst possible situation met by a coalition is for the remaining set of players, -S, also to form a coalition. The effects of any other possible combination of players not in S can be achieved by the coalition -S, and in general it can achieve some outcomes not in the province of less unified aggregates of players. Thus, a characterization of the minimum power of a coalition is its expected payoff when the remaining players also act as a coalition, in other words, when the game is played as a 2-person game between two coalitions. This, of course, is the case we have already examined in Part II and for which there is a unique (conservative) value given by the minimax theorem. Let this value be denoted by v(S) for the coalition S. Since this may be computed for each possible coalition, i.e., subset of players, we therefore have obtained a function v with domain the subsets of $I_n = \{1,2,\ldots,n\}$ and range the real numbers, i.e., a real-valued set function. Assuming the normal form of the game is known, the calculations involved in determining v are generally overwhelming. This, however, does not weaken the power of the theory to study all such functions. The function v is not without certain restrictions; it may be shown that in the zero-sum case it satisfies

i. $v(I_n) = 0$,

ii. $v(S) = -v(-S)$, for all $S \subset I_n$,

iii. $v(\phi) = 0$,

iv. if R and S are any two disjoint subsets of players,

$$v(R \cup S) \geq v(R) + v(S).$$

(Note that given condition ii, conditions i and iii are equivalent.) The first two conditions simply reflect the zero-sum character of the game. The third is a formal statement of the obvious fact that the subset involving no players neither loses nor wins anything. The important condition is iv, which, when one thinks about it, is an extremely reasonable one. It says that the whole does not obtain less than the sum of its parts, or, in another way, a coalition composed of the disjoint sets R and S can do anything R and S can do separately, and possibly more.

The function v has been named the characteristic function of a zero-sum game.

It is interesting and important that any real-valued set function v satisfying conditions i through iv is the characteristic function of a zero-sum game. That is, given such a v it is possible to construct a game which has as its characteristic function v.

The extension of this notion to non-zero-sum games is not completely straightforward, but rather it requires a mathematical trick. Suppose we have a non-zero-sum game with n players and to it we add a fictitious player who is not truly a free agent in the game but is so circumscribed that the resulting game on the n+1 players is zero-sum and who does not play a significant role in coalition formation. It is not completely obvious that this can be done, but it can. For the augmented (zero-sum) game one can

obtain a characteristic function, which when restricted to the subsets of the original n players is called the characteristic function of that game. It has only two properties:

i. $v(\phi) = 0$

ii. if R and S are any two disjoint subsets of players,

$$v(R \cup S) \geq v(R) + v(S).$$

If, in addition, the game is constant-sum (not necessarily zero-sum, but not excluding that case) then

iii. $v(S) = v(I_n) - v(-S)$, for all $S \subset I_n$.

Of course, if we assume the game is zero-sum, then $v(I_n) = 0$ and iii becomes the old condition ii.

When we use the term characteristic function we shall mean any real-valued set function satisfying i and ii, for it is again true that to each such function there is a game (no longer zero-sum) for which it is the characteristic function.

As in the case of zero-sum games, the first condition reflects the strategic inconsequence of the null set, and the second that any coalition is at least as potent as any two disjoint sub-coalitions formed from it. Now while these conditions have been derived from game considerations, first, of a game in extensive form, then in normal form, and then using the 2-person theory (with all the difficulties mentioned in Part II), it must be admitted that were we to think about coalition formation removed from any specific theory of games we could not require less. That is, were we to suppose the potency or strength of a coalition could be measured numerically, then we should, at the very least, require that conditions i and ii

be met - indeed, we would probably try to specify more. It is surprising that by restricting our analysis to a game we do not obtain further requirements to be met by characteristic functions. Thus, while we shall make certain criticisms of the characteristic function as an interpretation of the structure of a game, the abstract notion, and the resulting theory, appear to be very generally representative of the power of coalitions in human situations. Of course, the numerical values obtained from a game analysis may well differ from those we might assign by some other considerations. This suggests - and it is easy to confirm - that the study of characteristic functions, which is completely related to game theory, is more general in that situations which are not games in normal form can give rise to such functions; but in conformity to present usage, we shall refer to a finite set and a characteristic function defined over the subsets of the given set as a game. Later, we shall present reasons to suppose the study might be more appropriately called "the theory of finite super-additive measures".

Our next step is to divide games into two classes. It is conceivable that there are games in which no coalition of players is more effective than the several players of the coalition operating alone, in other words, that for every disjoint R and S, $v(R \cup S) = v(R) + v(S)$. Such games are called <u>inessential</u>; any game which is not inessential is called <u>essential</u>. It is not difficult to see that a game is inessential if and only if $v(I_n) = \sum_{i=1}^{n} v(\{i\})$. Since there is no value in forming coalitions in inessential games, it is clear that we cannot expect any

theory of coalition formation in that case, and so we shall be concerned only with essential games from now on.

4.2 S-Equivalence and Reduced Forms

Frequently in mathematics, and in its applications to science, we define a large class of objects all of which satisfy certain conditions, as we have done above with the characteristic functions. It is not uncommon that such a class can be partitioned into a number of non-overlapping subclasses, the elements of each subclass being in some sense equivalent. When this is done, one selects a representative from each class and develops the theory in terms of the representatives, of course always showing that the theory is invariant under the equivalence concept which originally allowed the partitioning. We must turn to this problem for characteristic functions. The intuitive idea of equivalence that we want to isolate may be called "strategic equivalence", i.e., we want to consider as equivalent two characteristic functions which lead to the same strategic considerations on the part of the players.

Suppose that one characteristic function v differs from another v' only by a multiplicative positive constant c, i.e.,

$$v(S) = cv'(S), \text{ for all } S \subset I_n,$$

then the two characteristic functions differ only in the unit whereby we measure the utility. One example would be to transform a characteristic function originally in dollars to one in cents. It is clear that such a change of unit cannot possibly affect the strategic character of the game to rational players.

Next, consider a game with characteristic function v and suppose that each player i is paid (or is caused to pay, depending on the sign) an amount a_i prior to the play of the game. Certainly these payments cannot have an effect on the strategies of the game, and yet it is easy to show that

$$v(S) + \sum_{i \in S} a_i, \quad S \subset I_n,$$

is a characteristic function. We certainly would want to consider this function strategically equivalent to v, since we could always effect the payments a_i before the play of the game. Combining these two conditions, we have the following definition: Two n-person games with characteristic functions v and v' are <u>S-equivalent</u> if it is possible to find n constants a_i and a positive constant c such that

$$v'(S) = cv(S) + \sum_{i \in S} a_i, \text{ for every } S \subset I_n.$$

It may not be obvious at this point that this definition of equivalence is a suitable one, and that no further grouping is needed; but the results we shall cite at the end of section III.5.1 show that it is adequate, at least for the von Neumann-Morgenstern theory of solutions.

Assuming this is so, we must now confront the task of selecting one representative from each class with which we may deal. Two suggestions have been put forward, each of which has certain advantages, primarily in the simplicity of stating certain games and certain definitions. The principle behind both of them is that it is possible to require that part of the representative characteristic function be the same for all of the

equivalence classes. Ignoring the (single) class of inessential games, von Neumann and Morgenstern [21] have shown that there is one, and only one, characteristic function in each of the equivalence classes which satisfy

$$v(\{i\}) = -1, \; i \in I_n, \quad \text{and} \quad v(I_n) = 0.$$

This they called the reduced form of a class of characteristic functions; we shall use the more specific term -1,0 reduced form.

A second reduced form, which we shall call the 0,1 reduced form, exists, since it readily follows that there is one and only one characteristic function in each equivalence class satisfying

$$v(\{i\}) = 0, \; i \in I_n, \text{ and } v(I_n) = 1.$$

We shall use the v notation for the -1,0 reduced form, but for clarity it seems appropriate to use a different symbol for the 0,1 reduced form; we shall use m.

Suppose v' is the characteristic function of an essential game, then the question arises as to how to find either the -1,0 or the 0,1 reduced form of the game. It is not difficult to show that the transformation

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

yields the 0,1 reduced form. The further transformation

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

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

we have a simple procedure to go from any characteristic function to either of the reduced forms.

The following remarks (to section III.4.3) are essentially parenthetical to the main development and so may be omitted if one chooses.

One of the first advantages, and possibly the most important, of the 0,1 reduced form is the emphasis it places on the relation between n-person game theory and the concept of a probability measure over the subsets of a finite set. Let us place side by side the conditions a 0,1 reduced form m and a probability measure p over I_n must satisfy

	0,1 reduced form	probability measure
i.	m is a non-negative real-valued set function	p is a non-negative real-valued set function
ii.	$m(I_n) = 1$	$p(I_n) = 1$
iii.	$m(\phi) = 0$	$p(\phi) = 0$
iv.	if R and S are disjoint subsets of I_n, $m(R \cup S) \geq m(R) + m(S)$	if R and S are disjoint subsets of I_n, $p(R \cup S) = p(R) + p(S)$
v.	$m(\{i\}) = 0$	
vi.	if the game is constant-sum, $m(S) = 1 - m(-S)$, for all $S \subset I_n$	it follows from ii and iv above that $p(S) = 1 - p(-S)$, for all $S \subset I_n$

The resemblance between m and p is marked, the most important differences being the inequality in the former and the equality in the latter for iv, and the lack of a p expression in v. We cannot have $p(\{i\}) = 0$ for all i, for were this the case then by a repeated application of iv we could conclude $p(I_n) = 0$, which contradicts condition ii. We shall return to this correspondence again when we try to characterize the

principal problem of n-person game theory.

Earlier we suggested that the study of general games by means of characteristic functions might well have been entitled "finite super-additive measures". Conditions i, iii, and iv above suggest the name "super-additive measure" and condition ii simply means that we shall deal with normalized measures, as in the theory of probability. However, condition v, $m(\{i\}) = 0$ is most unusual in measure theory. It is worth pointing out, at least for the mathematician, that we may drop this condition when we are studying theories invariant under S-equivalence, since under the transformation

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

m' satisfies $m'(\{i\}) = 0$ even if m does not.

These remarks serve to place the study of characteristic functions in a more general mathematical framework, namely, in the study of arbitrary, finite, normalized, real-valued set functions. If for all disjoint subsets R and S of a finite set,

$$m(R \cup S) - m(R) - m(S)$$

is equal to zero, then the measure is additive and the theory is that of discrete probabilities. If the quantity is always less than or equal to zero then the measure is called sub-additive. Some work has been done on these functions in conjunction with the theory of additive measures. Now game theory completes the area by introducing a theory of finite super-additive measures, which has so far resulted in a theory very different from the sub-additive or additive one; probably this is an inherent difference and not simply a reflection of the game terminology and motivation.

4.3 Imputations and Distributions

So far we have dealt with only one ingredient of the n-person game: the strength of the different coalition possibilities. Distinct from this, though presumably influenced by it, are the payments the players finally receive. Since we have assumed a transferable numerical utility, the direct payments and any side payments resulting from coalition formation can all be additively combined, so that for each player i a final payment x_i is received. Thus the total set of payments is an n-tuple of real numbers, which we may write as $X = \|x_1, x_2, \ldots, x_n\|$.

Whether a player is in a coalition or not, it is hard to imagine that if he is rational he would accept a final payment less than the least he can expect to receive if he were to play alone against a coalition of all other players, so we impose the condition

i. $v(\{i\}) \leqslant x_i$, for every $i \in I_n$.

Further more, we may suppose that rational players, no matter how they constitute themselves into coalitions, achieve a distribution of payments equal to what they would expect to receive if they had formed one grand coalition. For suppose $\sum_{i \in I_n} x_i < v(I_n)$, then each could be made to gain, say, the amount $\dfrac{v(I_n) - \sum_{i \in I_n} x_i}{n}$.

So we have as a second condition

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

Any n-tuple X of real numbers satisfying i and ii is called an imputation

of the game and it is interpreted as a possible set of payments to the players.

For much of the characteristic function theory it is not truly necessary to impose condition ii, specifically, it is shown in section III.5.6 that the restriction is not necessary for the von Neumann-Morgenstern theory of "solutions" (III.5.1).

One game theoretic problem (at this level of abstraction) may be stated as follows: Given the characteristic function of a game, to select from the set of all possible imputations and from the set of all possible arrangements of the players into coalitions those which may reasonably be expected to occur with rational players. The words "may reasonably be expected to occur" are not specified either in the extensive or normal form of the game, and it is the more or less arbitrary specifications that must be made which give parts of the theory the ad hoc character mentioned earlier. So far the interpretations have been as of some sort of stable equilibrium. When this problem is satisfactorily formulated - and many people think it is not - then other problems can be raised, such as, what is the path of changing coalitions and imputations from a non-equilibrium point to an equilibrium point; but such problems have not yet been considered.

We may throw added light on this general problem of game theory if we turn to the 0,1 reduced form. Note that by substituting the conditions for the 0,1 reduced form into the conditions for imputations, we find that

i. $x_i \geq 0$, for $i \in I_n$,

ii. $\sum_{i \in I_n} x_i = 1.$

In other words, the set of imputations corresponding to the 0,1 reduced form is identical to the set of all probability distributions over the elements of I_n. Once again the 0,1 reduced form has thrust us close to probability theory, and indeed this suggests a way to look at the general equilibrium problem of game theory.

Let $X = \| x_1, x_2, \ldots, x_n \|$ be a probability distribution over I_n; then the set function

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

is easily seen to be a probability measure over I_n (see section III.4.2) which simply assigns to each set S the sum of the individual payments to the members of S. Now we interpreted m(S) as characterizing (in the utility units of the game) the strength of the coalition S. It is, of course, by the interplay of coalitions and possible coalitions, by threats to form coalitions if certain agreements as to payments are not accepted, that the final payment X must be determined. The aim of the theory is to determine this outcome by formalizing what the threats must be, and it is clear that if for some S, m(S) is much larger than $P_X(S)$ there will be strong forces for the coalition S to form and to demand a new outcome, say X' such that $P_{X'}(S)$ is close to m(S). Thus, the equilibrium problem of game theory involves finding a probability measure P_X which in some sense approximates the normalized super-additive measure m. The heart of the problem is determining a suitable definition of "in some sense approximates"; the several attempts to do so are discussed in sections III.5 and III.6.

4.4 A Critical Example

Before turning to the theories themselves, we should in fairness point out that the simplification in passing from the normal form of a game to the characteristic function form is not without difficulties. It would indeed be surprising if we were able to make such a radical simplification of the theory of all n-person games without overlooking some of the differences among them. The example discussed in section II. , which is due to McKinsey [13, p. 351] , is sufficient to show that this is the case, even for the 2-person games. It will be recalled that in this game the player 1 has only one strategy and player 2 has two, the payoff matrix being

$$\| (0,-1000) \quad (10,0) \| \ .$$

There is no need to repeat here McKinsey's interpretation nor our discussion of it, except to remark that in general one must consider the normal form of this game to be asymmetrical in the two players. The characteristic function of the game is

$$v(\phi) = 0, \quad v(\{1\}) = v(\{2\}) = 0, \quad v(\{1,2\}) = 10.$$

While in general the normal form is asymmetrical, the characteristic function is always perfectly symmetrical, reflecting no difference between the two players.

Nonetheless, the characteristic function does express some of the aspects of a game, and the example certainly does not invalidate our earlier comment on the representation of coalition strength by characteristic functions independent of the extensive or normal form of game theory. Following a summary we shall turn to examining the resulting features of the theory.

4.5 Summary

The main preoccupation of this section was to reduce the general normalized game to a more tractable mathematical form, the form on which most of n-person game theory is built. It was observed that if coalitions of players are allowed, then the worst thing that can happen to a coalition S is for the coalition -S to form and for the game to be played between the two "players" S and -S. Using the minimax theorem of 2-person theory, a real number $v(S)$ was associated with each coalition S which describes the conservative expected payment to the coalition. It can be shown that the function v, which is known as the characteristic function of a game, satisfies

i. $v(\phi) = 0$,

and ii. if R and S are disjoint subsets of I_n,

$$v(R \bigcup S) \geq v(R) + v(S).$$

Further, it can be shown that any real-valued set function satisfying i and ii is the characteristic function of some game, so one cannot in general derive any further independent properties of characteristic functions. The theory of n-person games is to be based on such functions.

Any game with a characteristic function satisfying

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

is called inessential and it was argued that its coalition theory is trivial; any other game is called essential.

Two characteristic function v and v' (over the same set of n players) are called S-equivalent if there exists a positive constant c and constants a_i such that

$$v(S) = cv'(S) + \sum_{i \in S} a_i .$$

It was argued that such games are subject to the same strategic considerations since c represents only a change of scale and the payments a_i are independent of the outcome of the game. It was observed that S-equivalence is technically an equivalence relation and so it divides the set of all characteristic functions into non-overlapping subsets called equivalence classes. Any two functions in the same equivalence class are S-equivalent, so the characteristic functions of any one class are all subject to the same strategic considerations and it is therefore sufficient to develop any theory in terms of one example for each class. Two possible and closely related choices were given both of which have the property that part of the characteristic function is constrained to be the same in each equivalence class. The first, called the -1,0 reduced form, is the unique characteristic function in each class for which

$$v(\{i\}) = -1, \quad i \in I_n, \quad \text{and} \quad v(I_n) = 0.$$

The second, called the 0,1 reduced form, is the unique characteristic function in each equivalence class such that

$$v(\{i\}) = 0, \quad i \in I_n, \quad \text{and} \quad v(I_n) = 1.$$

The characteristic function can be thought to represent the threat power of the various coalitions and it is hoped that from this it will be possible to determine what happens in the game. One of these events is that each of the players will ultimately receive a certain payment, which consists not only of his payoff as prescribed by the payoff function of the game, but which must also take into account any side payments he receives or pays out in order to preserve a certain advantageous coalition arrangement. This

suggested that the final payment to each player could be represented by a single number x_i and it was argued that such n-tuples should satisfy

i. $x_i \geq v(\{i\})$ for all $i \in I_n$,

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

Such n-tuples are called imputations. In the 0,1 reduced form of the game an imputation is simply a probability distribution over the set of players.

The general problem of n-person game theory was then stated: to find those imputations and those arrangements of the players into coalitions which are in some sense compatible with the given characteristic function of the game. The reader was warned that in the attempt to make more precise what we shall mean by "in some sense compatible" (a concept not prescribed by the formalism of the game) n-person game theory is given an ad hoc flavor to which some authors object.

5. Solutions

5.1 The von Neumann-Morgenstern Definition

In the published literature of n-person games one definition, based on characteristic functions and imputations, has received most attention; this definition, introduced at length by von Neumann and Morgenstern [21], was offered as the "solution" to the n-person game - indeed, it was given the name "solution". Following their exposition, we may first suggest the idea by an example. It is not difficult to see that the -1,0 reduced form of a constant-sum 3-person game is unique, and that it is

$$v(\{i\}) = -1, \quad v(\{i,j\}) = 1, \quad v(\{1,2,3\}) = 0.$$

Suppose, for the moment, that the coalition $\{1,2\}$ forms; according to the characteristic function it may command a payment 1. Since both 1 and 2 have symmetric roles in the sense that if we were to change their labeling so that 1 were 2 and 2 were 1 the characteristic function would be unchanged, it is not unreasonable to suppose that they would divide 1 equally, and player 3 would be forced to accept -1. But arguing by symmetry again, there is no reason to single out the coalition $\{1,2\}$ as superior to $\{1,3\}$ or to $\{2,3\}$, and so any of the three imputations

$$\left\| \tfrac{1}{2},\tfrac{1}{2},-1 \right\|, \qquad \left\| \tfrac{1}{2},-1,\tfrac{1}{2} \right\|, \qquad \left\| -1,\tfrac{1}{2},\tfrac{1}{2} \right\|$$

seem reasonable outcomes. We call this set of imputations F. Suppose we consider any other imputation $\| x_1,x_2,-x_1-x_2 \|$, not one of the above three, then at least two of the entries are less than $\frac{1}{2}$, otherwise the sum of the payments is not zero. Thus, the imputation in the set of 3 having payments of $\frac{1}{2}$ for those two players is superior for both of those players, and since they are in a coalition of two, they may force the better arrangement. Equally important, no imputation of the set F dominates either of the other imputations in F in that fashion. Thus, the set F of three imputations plays a very special stable role in the set of all imputations for the game. The question arises whether the notion can be generalized.

Von Neumann and Morgenstern proposed the following definitions. Let a game be given by its characteristic function v. An imputation Y is said to <u>dominate with respect to a coalition</u> T another imputation X if

i. T is a non-empty set of players,

ii. $v(T) \geq \sum_{i \in T} y_i$,

ii. $y_i > x_i$ for every $i \in T$.

T is called an effective set for this domination. This is exactly the condition met above in the domination of, say,

$\left\| \frac{1}{4}, -\frac{1}{2}, \frac{1}{4} \right\|$ by $\left\| \frac{1}{2}, -1, \frac{1}{2} \right\|$ if we take $T = \{1,3\}$.

If there is some coalition T such that Y dominates X with respect to T, then we can simply say that Y dominates X. It turns out that examples can be given to show that each of the following cases can arise:

i. Y dominates X, but X does not dominate Y,

ii. Both Y dominates X and X dominates Y,

iii. Neither Y dominates X nor X dominates Y.

Now, a solution to a game is any set A of imputations such that

i. for any two imputations X and Y in A neither X dominates Y nor Y dominates X,

ii. and for any imputation Z not in A there is at least one imputation X in A which dominates Z.

It should be pointed out immediately that the definition of solution in no way precludes the existence of imputations not in A which dominate one, or indeed all, of the members of A. This possibility is implicit in statement ii following the definition of domination. We shall return to this point, which is not without complications.

As might be expected, the set F of imputations in the 3-person zero-sum game is a solution.

We mentioned earlier that our theory should be invariant under S-equivalence, that is, that two S-equivalent games should lead to the same results. This has been shown [13, 21] to be the case for the domina-

tion concepts, and so for solutions. The much more subtle converse, that if two games have imputation spaces which are isomorphic under domination then the games are S-equivalent, has recently been shown by McKinsey [14] to hold for zero-sum games.

5.2 Some Remarks about the Definition

Before discussing the mathematical results which have been obtained - and some which have not - certain questions about the intuitive adequacy of the definition of a solution must be considered. The notion of "dominance with respect to a coalition T" is really the conjunction of two notions: "Y is 'better' than X with respect to T" is the meaning of condition iii: $y_i > x_i$, for $i \in T$; and "Y is 'feasible' with respect to T" is the meaning of condition ii: $v(T) \geqslant \sum_{i \in T} y_i$. Of these two, there seems little reason to question the first under any condition, while the latter is open to question. It can be argued that if the theory we desire is normative, then the coalition T can never enforce more than v(T) since rational players will certainly form the coalition -T, and so no imputation Y with $v(T) < \sum_{i \in T} y_i$ is feasible. If, however, one is concerned with a descriptive theory of games, then it is not clear that the feasibility condition is appropriate, for if the players not in T do not form a single coalition, then the members of T may be able to get more than v(T). Just how much more they can get is not easy to say, in fact, saying it would amount to developing a descriptive game theory. It appears that this point precludes our interpreting the solution concept as a descriptive theory, for certainly not all economic, military, and social conflict-of-interest

situations reduce to the opposition of two coalitions.

A close corollary of the last remark, and we believe an important weakness of the solution concept, is that it is concerned only with imputations and does not give any information about the coalition structure of the game when it is in the equilibrium state. Again, whether this is really a suitable objection when a normative theory is desired is not certain, but it is a valid criticism of solutions as a descriptive theory since the coalition structure is probably the most easily observed fact in any real situation, certainly one more easily observed than the imputations.

These remarks strongly suggest that there is little hope that the solution notion can be used in other than a normative way, and this will be confirmed when we examine the resulting theory. Even were we to try to use the theory as a descriptive one, it is not at all clear what we should say the theory asserts to happen. McKinsey remarks, "Although a large part of von Neumann and Morgenstern's book (roughly 400 out of 600 pages) is devoted to games with more than two players, mathematicians generally seem to have been dissatisfied with the theory there developed." [13, p. 303] It is not clear whether he intended this to apply equally to the characteristic function development and to the definition of a solution, or only to the latter. Certainly, there has been warm admiration for the ingenuity of the solution idea, and it has received considerable study - to which we now turn.

5.3 Some Implications of the Definition

The first main point we should make is that a solution does not

generally consist of a single imputation, but rather of several. This is obvious in the example of the 3-person game, and indeed, it can be shown that any game having a solution consisting of but one imputation is inessential. In addition to solutions having a finite number of imputations, such as the three in F, some solutions consist of an infinity - and not necessarily a countable infinity - of imputations. We shall give an example of this in a moment.

Second, aside from the multiplicity of imputations in a solution, there are in some games more than one solution. That this is a possibility was suggested by our earlier comment that there may be imputations not in a solution which dominate imputations of that solution. As an example of the non-uniqueness of solutions, in the zero-sum 3-person game any set of imputations

$$\| x_1, x_2, c \|$$

where c is fixed but such that $-1 \leqslant c < \frac{1}{2}$ and $x_1 + x_2 = -c$, is a solution. Equally well, the two sets of imputations obtained by moving c to player 1 and to player 2 are solutions. We shall denote these solutions $F_i(c)$, $i = 1,2,3$. Since for each possible fixed c in the half-open interval $[-1,\frac{1}{2})$ there are solutions, we have a continuum of solutions, each of which contains a continuum of imputations. Indeed, every possible imputation for the constant-sum 3-person game is included in at least one solution! "Therefore in the case of the essential three-person game we have an embarrassing richness of solutions." [13, p. 339]

This abundance is not restricted to the 3-person game.

The question must immediately be raised as to how these solutions are to be interpreted. Von Neumann and Morgenstern divide the discussion

into two parts. First, they say that of the several solutions, the one which is accepted depends on "standards of behavior" which are moral or conventional rules imposed by society. Thus , they say, if society accepts discrimination, one may find a solution of the type $F_1(c)$ where the position of c in the range -1 to $\frac{1}{2}$ is determined by the degree of discrimination tolerated by the society. Assuming c fixed, there is a question how the other two players will divide -c, and this they say is a problem in bargaining which depends on the relative bargaining abilities of the two players. They do not say how it will be decided which player will be discriminated against, or in the case of the non-discriminatory solution F, which imputation will arise. Apparently this is a chance matter depending on which coalition was first formed, or again, it may depend on the relative bargaining abilities of the three players. It is such discussion which gives this theory the ad hoc character mentioned earlier (III.1).

They argue at some length that solutions are "stable". They point out that while an imputation not in a solution may dominate one in a solution, and although it is "preferable to an effective set of players, [it] will fail to attract them, because it is 'unsound'" [21, p. 265]. And "the attitude of the players must be imagined like this: If the solution [A]... is 'accepted by the players 1,...,n, then it must impress upon their minds the idea that only the imputations... [in A] are 'sound' ways of distribution." [21, p. 265] And "The above considerations make it even more clear that only [A] in its entirety is a solution and possesses any kind of stability - but none of its elements individually. The circular character ... makes it plausible also that several solutions [A] may exist for the

same game. I.e., several stable standards of behavior may exist for the same factual situation. Each of these would, of course, be stable and consistent in itself, but in conflict with all others." [21, p. 266]

The full flavor of their argument is hard to recapture, and it can only be recommended that the reader turn to the discussions of solutions in the book. That some readers have not been completely persuaded by their arguments is indicated by the comment of McKinsey that "Some people have felt dissatisfied with the intuitive basis of this notion, however; and the question has been raised as to whether knowing a solution of a given n-person game would enable a person to play it with greater expectation of profit than if he were quite ignorant of this theory." [13, p. 332]

5.4 Further Implications of the Definition

We have already given solutions to the 3-person constant-sum game, and it is known that these are all of the solutions to that game.

It is also known that every 4-person constant-sum game has at least one solution, and of the triple infinity of 4-person games a few have been studied in detail. The reader is referred to von Neumann and Morgenstern [21] for the full discussion of these cases.

It is not known whether every game possesses a solution; for example, it is not known if every 5-person game has a solution. From the first systematic presentation of n-person game theory to the present, this has been considered the most important unresolved problem.

A game in 0,1 reduced form is called simple if $m(T) = 1$ or 0 for every coalition T. Von Neumann and Morgenstern studied solutions in

some simple games, particularly for $n = 4,5,6,7$, but also in certain more general cases. In addition, Bott [2] introduced the notion of an (n,k), or majority, game defined by

$$m(S) = \begin{cases} 0 & \text{if } |S| \leq k \\ 1 & \text{if } |S| > k \end{cases}$$

which is simple, and he studied the symmetric solutions of such games. Gillies [5] examined non-symmetric or discriminatory solutions to such games. In the words of Kuhn and Tucker, "Dropping symmetry, D.B. Gillies exhibits ... a surprising variety of other solutions of (n,k)-games, all derived from Bott's symmetric solutions. Gillies' solutions are obtained by several methods which may carry over to a more general context: (1) by the addition of 'bargaining curves' (Theory of Games and Economic Behavior, p. 501), (2) by inflation to larger games (ibid., p. 398), (3) by 'discrimination' (ibid., pp. 288-289) in which the non-discriminated players divide their take according to any solution to a smaller game, or (4) by partitioning the players into fixed subsets, assigning the spoils arbitrarily (i.e. in all admissible ways in one solution) among these subsets, and then dividing the spoils in any one subset according to the symmetric solution to a smaller game the players think they are playing." [10, p. 304]

Another class of games which has been studied is the quota games. Shapley [23] calls a game a <u>quota game</u> if it is possible to divide $v(I_n)$ among the n players, i.e., to find ω_i with

$$v(I_n) = \omega_1 + \omega_2 + \ldots + \omega_n,$$

in such a way that

$$v(\{i,j\}) = \omega_i + \omega_j, \text{ for all } i \text{ and } j, \ i \neq j.$$

"Shapley obtains families of solutions for the entire class of quota games, a class that contains some three-person games, all constant-sum four-person games, and a sizeable swath of all games with more than four players. In a typical imputation in one of these solutions, all but two or three of the players receive their 'quotas' ω_i." [10, p. 304-305]

In still another paper, Shapley [25] has presented a class of solutions to a certain simple game, which, as he says, "...provides at one stroke a large fund of 'pathological' examples against which conjectures on the behavior of ... solutions can be tested." [25, p. 1] The solution is based, in part, on an arbitrary closed set C of an (n-3)-dimensional subset of the space of imputations. "The arbitrariness in the choice of C (for example, C may be a Cantor-type discontinuum) makes it easy to dispose of many conjectures concerning the regular behavior of ... [solutions] " [25, p. 2]

There is little reason to present these results in detail here, for they would require considerable space and not a little notational apparatus; the interested reader can refer to the original publications. However, certain summary observations are in order. The variety and complexity of solutions in the games so far studied are overwhelming; their characterization and the corresponding proofs are involved and often subtle. It is doubtful that a mathematician could be found today holding any hope for completely general characterization of solutions; the most optimistic hope is that it will be possible to divide the class of all games into a number of subclasses such that solutions in each can be characterized completely.

We may fairly conclude that in addition to the conceptual difficul-

ties mentioned in III.5.2, there are also mathematical difficulties, or, at least, the mathematical problem is difficult. This is going to prove either so stimulating that it will lead to deep insights or so discouraging that little more will be discovered about solutions. At this stage it is not clear which will occur.

Assuming that at least some people will be discouraged, there appear to be two possibilities: (1) efforts will be made to single out some of the solutions as more important than others and these will be studied, and (2) efforts will be made to introduce new concepts more or less in competition with that of a solution. In the next section we deal with an example of the first approach, in section III.6 with three examples of the second. Of these four, two (III.5.5 and III.6.1) had not yet been published at the time of writing, so it is not possible to give them a critical analysis resting on the observations and work of a number of people; therefore, both must be treated as somewhat tentative approaches to the problem.

5.5 Strong Solutions*

The principal question to be discussed in this section is whether, aside from "standards of behavior" there are game theoretical requirements which impose a greater stability on one solution than on another. This problem and the ideas here discussed were raised by Vickrey [30]. With respect to a specific solution A he calls an imputation of A a conforming imputation, one not in A, non-conforming. Among the non-conforming imputations some dominate one or more conforming imputations; these he calls heretical imputations, and an effective set for such a domination is called a heretical set.

---

* Throughout this subsection, whenever we quote Vickrey, we shall replace his symbols for imputations, coalitions, and solutions so that the notation is in conformity with the rest of this report.

The shift from a conforming imputation to a heretical one is termed a heresy. He remarks, "...there is nothing in the definition of a solution as it stands that makes it dangerous for players to participate in a heretical move. We can, however, observe solutions in which heresies tend to be dangerous to one or more of the members of the heretic set, as well as solutions in which heresies may be quite profitable." [30, p. 7] As an example, consider the imputation $F_3 = \|\frac{1}{2},\frac{1}{2},-1\|$ of the solution F of the 3-person game. Any non-conforming imputation X which is to dominate F_3 with the heretical set $\{2,3\}$ must clearly satisfy

$$x_1 < \tfrac{1}{2},\ x_2 > \tfrac{1}{2},\ x_3 > -1,\ x_2 + x_3 \leqslant 1.$$

It is not hard to show that such an imputation is dominated by one and only one member of F, namely $F_2 = \|\frac{1}{2},-1,\frac{1}{2}\|$ with the effective set $\{1,3\}$. Vickrey writes, "...in this case the movement to a non-conforming imputation X requires the cooperation of a player 2, who though he may gain immediately, finds that although it may have been difficult to move from F_3 to X it is now much easier for the couple $\{1,3\}$ to organize a movement to the conforming imputation F_2 to the great discomfiture of 2... If 2, finding himself now in the excluded position, attempts to negotiate with either 1 or 3 to move away from F_2, not only will 2 have to propose a heresy in which he gets less than the $\frac{1}{2}$ that he started with in F_3, but he will find that 1 and 3, having observed what happened to 2, will be very reluctant to join any such heretical coalition, and in fact may refuse to do so altogether. Either because the players foresee all this, or because after a short time they come to the conclusion as a result of experience that heresy is in the long run likely to lead to disaster for at least one of the heretics, they

eventually will come to stick to the policy of staying at one of the approved imputations..." [30, p. 8]

Of course, this observation would be idle if all solutions had that property. Consider the solution of the 3-person game of the form

$$\| x, -x-c, c \| , \quad \text{where} \quad -1 \leqslant c < \tfrac{1}{2} \quad \text{and} \quad -1 \leqslant x \leqslant 1-c$$

the set we have called $F_3(c)$. Let us suppose the players are at one of these imputations, say

$$X = \| x_1, x_2, c \| .$$

If we exclude $X = \| \tfrac{1}{2}, \tfrac{1}{2}, -1 \|$, we may assume without loss of generality that $x_2 < x_1$. Observe that with respect to $F_3(c)$, the imputation

$$F_1 = \| -1, \tfrac{1}{2}, \tfrac{1}{2} \|$$

is heretical since it dominates X, the effective set being $\{2,3\}$. In turn, it can be shown that any imputation Y of $F_3(c)$ which dominates F_1 must satisfy

$$y_2 > x_2 \quad \text{and} \quad y_3 = c.$$

Thus "...even if there was a return to a conforming imputation after a relatively brief period of heresy at F_1, the players 2 and 3 responsible for the heresy would gain from the excursion, 3 temporarily and 2 more permanently." [30, p. 9]

Vickrey adds the following paragraph. "Even if a return from F_1 to a conforming imputation Y is made indirectly... so that it is possible for player 2 to be worse off in Y than in F_1, it is by no means certain from the characteristics of the game that player 2 will not be able to avoid such an eventual worsening of his position. And even if after one particular heretical excursion player 2 finds his position...worse..., there is now nothing to prevent him from trying another heretical excursion, since player 3 whose

cooperation he needs has nothing to lose by it in any event and stands to gain at least temporarily. In effect any player who is willing to engage in heretical excursions is at an advantage in bargaining for position among the approved imputations, over a player who eschews such tactics. It thus appears that in this case it will take a much stronger social sanction to compel adherence to the approved standard of behavior than where the standard of behavior conforms to the symmetrical solution..." [30, p. 9]

Vickrey proposes the following two definitions. Let A be a solution, X an imputation of A, Y a heretical imputation dominating X with the effective set T, and U the set of elements of A which dominate Y. A is said to be a strong solution if for every such X, Y, and T there is at least one element i of T such that for every Z in U, $z_i < x_i$. On the other hand, A is said to be weak if for every X of A there exists at least one heretical Y with effective set T such that for all Z of A which dominate Y, and all i in T, $z_i \geqslant x_i$.

For the constant-sum 3-person game we have seen that the symmetric solution F is strong, and that all discriminatory solutions $F_i(c)$, $i = 1,2,3$, are weak.

For games with more than 3 players there are solutions which are neither strong nor weak, but rather there are intermediate notions of strength. Primarily, however, one is interested in the strong solutions, for which all heresies are dangerous to some member of its effective set.

To examine specific cases one must, of course, know the solutions, so Vickrey has been restricted to studying such cases as some 4-person games and some simple games, and, in summary, he finds that "For constant-sum games,

the concept of the strong solution has thus far appeared to be fairly effective in narrowing down the number of solutions that have to be accepted. When it comes to the variable-sum games, unfortunately, it appears that much of the selectivity of insistence on strong solutions disappears. For one and two person games, all solutions are already strong, while for three person games, it appears that insistence that ... solutions be strong offers only a relatively small reduction in the range of possible imputations." [30, p. 32]

"No attempt has as yet been made to try out the effect of insisting on strong solutions for variable-sum games of more than three persons, so there is no way of telling whether the concept would prove more restrictive in such cases or not. The complexities and variations possible between the extremes of strong and weak solutions already observed for the four-person constant-sum game indicate that the analysis of such games may prove to be extremely difficult. On the basis of the experience with the three-person games, one is inclined to be not too sanguine. The strong solution, that appears to be such a potent device for the simplification of the results of constant-sum games, may, it appears, be of relatively little value for the variable-sum games, although this tentative hypothesis is hardly more than a conjecture." [30, p. 35]

5.6 Extension of the Solution Concept

In addition to the difficulties we have raised earlier with respect to solutions, Shapley has pointed out [24] that even if the general notions of domination and solution are accepted there is an open problem about the domain of n-tuples over which these notions should be defined. In von Neumann

and Morgenstern and in our presentation here, the domain is that of imputations. "The propriety of this restriction to [the set of imputations] may be challenged on several grounds. In the first place, it is not at all obvious that the notion of group rationality, as exemplified by the solution of an n-person game, must necessarily be a refinement of the principle of individual rationality, as embodied in the inequalities $[x_i \geq v(\{i\})]$. In the second place, it would seem methodologically more correct to study the consequences of the domination process separately from those of the blocking process.* One might even hope that the former, apparently the more powerful, might make [illegible] would be only a technical convenience, and would not prejudice the conceptual substructure of the theory.) Failing this, the restriction to [imputations] might better be applied (if it is desired to exclude 'irrational' solutions) after stability under domination has been secured." [24, p. 3]

To begin with, Shapley weakens the conditions on a characteristic function. He continues to require

$$v(R \cup S) \geq v(R) + v(S) \quad \text{for disjoint R and S in } I_n$$

but he drops the requirement $v(\phi) = 0$. Rather, he assigns to ϕ the (negative) value: the least that the whole group might get minus the maximum they might get. That is, $v(\phi)$ gives the spread of possible profit from playing the game. Of course, in a constant sum game $v(\phi) = 0$. While in this section we shall use this more general definition, in sections III.6 and III.7 the definition of III.4.1 will be employed.

Three different classes of n-tuples have been isolated:

---

* By "blocking process" Shapley means the refusal of a player i to accept a payment less than $v(\{i\})$.

G is the set of n-tuples X such that

$$v(\phi) + v(I_n) \leqslant \sum_{i \in I_n} x_i \leqslant v(I_n).$$

E is the set of n-tuples X such that

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

I is the set of n-tuples which are in E and such that

$$x_i \geqslant v(\{i\}).$$

We observe, first, that I is the set of imputations, and that

$$I \subset E \subset G.$$

If the reader will turn back to section III.5.1 he will see that neither the definition of domination nor that of solution directly employs the fact that we were dealing with imputations; they are concepts defined for any given set C of n-tuples, and at that time we specified $C = I$ ($=$ the set of imputations). Shapley introduces the term C-stable for those sets A of C which satisfy the conditions of a solution, i.e.,

i. no element in A dominates another element in A,

and ii. every element of C not in A is dominated by some element of A.

An I-stable set is therefore another way of speaking of a von Neumann-Morgenstern solution.

Among the theorems proved by Shapley we find that a set A is G-stable if and only if it is E-stable. That is to say, if one is concerned with stable sets, then it is immaterial whether one chooses G or E as the set of n-tuples, for no G-stable set intersects G-E. The relation between G and I is more complicated, but it is somewhat revealing of the effect of the added condition. Suppose a set A is a solution, i.e., it is an I-stable

set, then A is a G-stable set if and only if for each player i it is possible to find an n-tuple X in A such that $x_i = v(\{i\})$.

The significance of the work of Shapley's is that it shows clearly the effect on the von Neumann-Morgenstern theory of solutions of restricting the class of possible payments to the set of imputations. The last result indicates that the restriction to imputations is not redundant in solution theory, but more than that, it shows what effect the restriction has. His other results are of a similar nature, and the reader is referred to [24] for a full exposition.

5.7 Summary

The topic of this section - the solutions of von Neumann and Morgenstern - is the major game theoretic superstructure so far constructed upon the concept of a characteristic function. Initially it was noted that over the space of imputations a relation known as "domination" can be defined. One imputation X is said to dominate another imputation Y if there exists a non-empty coalition T such that every member of T prefers X to Y, or in symbols, if

i. for $i \in T$, $x_i > y_i$,

and if it is reasonable for the members of T to expect the total payment prescribed by X, i.e., if

ii. $v(T) \geq \sum_{i \in T} x_i$.

It was noted that the domination relation need not be asymmetric; in other words, that for imputations X and Y it is possible for both X to dominate Y and Y to dominate X (of course, different coalitions are involved in each case).

A set A of imputations is called a solution if

i. no imputation in A dominates another imputation in A,

and ii. every imputation not in A is dominated by some element of A.

The solutions of the 3-person constant-sum games were given, and from these results it is known that there may be a continuum of different solutions, that any one solution may contain either a finite number or a continuum of imputations, and that every imputation of the 3-person constant-sum game is a member of at least one solution. In contrast to the plethora of solutions in that case, it was noted that one of the major unsolved problems of n-person game theory - some would say the major one - is to prove the existence of at least one solution for every n-person game.

Verbal arguments were presented to defend the point of view that any solution represents a particularly "stable" set of imputations and that rational players will not attempt to deviate from it once it is selected. The selection of one solution from the many possible was ascribed to "standards of behavior" of society which, in the 3-person case, would dictate whether discrimination is allowed and if so how much. The determination of exactly which imputation of a solution will arise in a given situation was attributed to the "bargaining abilities" of the players and/or chance. Doubts exist as to whether such verbal discussions can really be considered a satisfactory resolution of the problem.

In addition to the above conceptual points, it was pointed out that solutions do not generally appear to have very regular properties and that so far it has proved impossible to characterize mathematically all solutions of any broad class of games. Their very irregularity and abundance,

however, are felt by many to be the strength of the theory for they allow it to encompass a wide variety of phenomena. It is argued that this is necessary since human beings seem to organize in a large variety of ways to cope with the same situation.

An attempt to give a formal meaning to the notion that solutions are particularly stable sets of imputations led to the concept of a strong solution. Briefly, a strong solution is one such that each imputation which dominates a heretical one (not in the solution) also actively "punishes" at least one of the players participating in the heresy. It appears that this concept is a very effective restriction on solutions in constant-sum games (in the 3-person case isolating only one), but there are tentative indications that the notion is much less successful for non-constant-sum games.

In the final section it was pointed out that the restriction to imputations is not necessary in order to define the domination relation and to isolate sets of n-tuples analogous to solutions; these are called C-stable sets, where C is the particular set of n-tuples under consideration. One of the central results is that the condition

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

which is required of an imputation, is not essential when solutions are studied. The condition

$$v(I_n) \geqslant \sum_{i \in I_n} x_i$$

coupled with the properties of the solution concept automatically causes the equality to be satisfied by the n-tuples in the solution.

6. Stability, Value, and Reasonable Outcomes

6.1 Stability of Games

Aside from solutions and their ramifications, there appear to be three other topics in n-person (characteristic function) game theory which have received attention. While two of these (sections III.6.1 and III.6.3) continue to be concerned with outcomes which might reasonably be expected to occur in a game, all three differ appreciably from the solution notion. For example, one of the salient differences of the definition we shall present in this section is that it does not deal with imputations or sets of imputations alone, but, following the suggestion of III.4.3, it isolates pairs consisting of an imputation and a corresponding breakdown of the players into coalitions.

Following our familar precedent, we shall use the 3-person constant-sum game as a source of ideas. Suppose the players were to consider an imputation X, where, without loss of generality, we may suppose $x_1 \leqslant x_2 \leqslant x_3$. It follows immediately that $x_3 > -1$, and so $x_1 + x_2 < 1$. Thus, players 1 and 2 might be expected to form a coalition and to split the resulting payment, 1, say by adding half the difference between 1 and $x_1 + x_2$ to the amount each would have received according to X. In this arrangement, player 3 receives only -1, and so it behooves him to go to player 1, who is receiving less than player 2, and to suggest to him that both of them could improve their lot by forming the coalition $\{1,3\}$. This proposal would be acceptable, for 3 can allow 1 to do a little better than he would in the coalition with 2, and at the same time 3 will do better than -1. Of course, this iso-

lates player 2 with an expected payment of -1, but he in turn can approach player 3 with a similar offer, and so on. It might be proposed at some stage that, to counter this infinite regress, a coalition of two players gives the third player enough so that he would not try to disrupt the coalition; but it can be shown, in the 3-person case, that "enough" to satisfy him would cause at least one of the other players to lose as a result of joining the coalition, and so it would not be formed. Looked at in this way there appears to be an inherent instability in the outcome of the 3-person constant-sum game.

Intuitively it appears that this argument could be applied to any game and so every game is unstable in this sense. This, if true, means the analysis must be too gross, for certainly there are some games one simply does not want to pass off as unstable. What is suggested is that, rather than an absolute stability-instability dichotomy for games, we define a notion of degree of stability. Our method of doing this will involve the introduction of an extra-game parameter, and it is this which gives the present theory its ad hoc character.

Let us suppose that in one way or another the players have agreed on a system of coalitions, which we may describe by $\mathcal{T} = (T_1, T_2, \ldots, T_t)$, where the T_i are coalitions which are non-overlapping and which exhaust the set of players. Now, these rational players presumably wish to better their lot, and so we must assume that each of the coalitions T_i is contemplating changes in membership in an attempt to improve its position. In general, the coalition T_i may contemplate the addition of a set of members, say G, and also it may decide to expel some members, H, (who are not carrying their

share of the load, in some sense). If these changes were made, the net result would be the coalition $(T_i \cup G)-H$. It is not difficult to see that if there is no restriction on the choice of G and H, then any possible coalition S can be represented in the form $(T_i \cup G)-H$ by appropriate choices of G and H. If, however, we were to restrict the choice of G and H, i.e., to restrict the coalitions which T_i may consider within its domain of change, then there may be coalitions S which cannot be written in the form $S = (T_i \cup G)-H$.

We shall suppose that the limitations on the choice of G and H are given in the following manner. For each possible system of coalitions $\mathcal{C}$ (= partition of the players into non-overlapping subsets), a distinguished set of coalitions which includes the elements of $\mathcal{C}$ is given; this set of coalitions may be denoted by $\psi(\mathcal{C})$. Each of the coalitions in $\psi(\mathcal{C})$ is called a ψ-<u>critical</u> coalition of $\mathcal{C}$.

Intuitively, we think of ψ being determined so that if S is a ψ-critical coalition of $\mathcal{C}$, then there is a coalition T_i in $\mathcal{C}$ which is not too different from S. Our assumption will be that a change from T_i to S can and will be effected by the players if there is some reason to do so (see below). One might imagine that this assumption would necessitate tagging each ψ-critical coalition according to which T_i of $\mathcal{C}$ may change into it, but for the present equilibrium theory this is not necessary; a mere listing of all the ψ-critical coalitions of $\mathcal{C}$ is adequate.

Given such a ψ, however chosen, our next concept is concerned with those imputations and partitions into coalitions such that there are no "forces" on the players to change their alliances, the degree of allowable

change being given by ψ. Let X be an imputation of a game and γ a partition of the players into coalitions. The pair (X, γ) is called ψ-stable if the following two conditions are met:

i. if $T \in \gamma$ and if $|T| > 1$, then $x_i > v(\{i\})$ for $i \in T$;

ii. if S is a ψ-critical coalition of γ,

$$v(S) \leqslant \sum_{i \in S} x_i .$$

The first of these two conditions simply reflects the intuition that to persuade a player to participate in a coalition of two or more players it is necessary to give him more than he could expect to receive if he were to play alone. To understand the second condition, suppose that on the contrary, $v(S) > \sum_{i \in S} x_i$ for some ψ-critical coalition S. Then if coalition S is formed there is an assured gain in payment to the coalition S above what was arranged in the imputation X, and each of the players in S could be made to profit by giving him, for example,

$$x_i + \frac{v(S) - \sum_{i \in S} x_i}{|S|}$$

Since S is a ψ-critical coalition of γ, the change to S is possible by our assumption, and so, assuming rational players, it would be seriously considered. Whether it would be effected depends, presumably, on other competing possible and advantageous changes. In any case, there would be "positive forces" to disrupt the pair (X, γ). If, on the other hand, condition ii holds for every ψ-critical coalition of γ, then within the limitations on change specified by ψ there is no inducement for any changes from

the pair $(X, \mathcal{C})$; and so it is a point of equilibrium, or, as we have said, the pair is ψ-stable.

One might raise at this point the question of uniqueness of such pairs in a game, which, as we have stressed earlier, is of some importance in a predictive theory. However, this discussion will be easier after we have presented some results.

The definition of ψ-stability of a pair is invariant under S-equivalence, and so it is acceptable from the point of view of section III.4.2. We shall call a game ψ-stable if there exists at least one ψ-stable pair, otherwise it is called ψ-unstable.

These definitions, and the following results, are due to one of the authors of this report. His paper [11] presents definitions and results for only the first special case of ψ which we shall discuss below, but the modifications indicated here are very easily made.

With the function ψ absolutely unspecified, as it is above, little more can be said. If, however, we make certain specific choices for ψ, it is to be expected that certain theorems can be proved. We shall make two closely related assumptions on the form of ψ, both of which lead to the same theorems. In effect, the first specification says that a coalition S is in ψ if there exists a T_i in $\mathcal{C}$ such that S and T_i are not too different. To be precise, let an integer k between 1 and n-2 be given. We shall denote the ψ we are about to define by V_k. Any coalition S is in $V_k(\mathcal{C})$ if and only if there exists a T_i in $\mathcal{C}$ such that $|(S-T_i)\bigcup(T_i-S)| \leq k$. Put another way, S is in $V_k(\mathcal{C})$ if and only if there exists a T_i in $\mathcal{C}$ such that a subset H of T_i and a subset G of $-T_i$

can be found with the properties

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

In words, S is a V_k-critical coalition of γ if there is a coalition T_i of γ which can be modified into S by the addition of players and by the removal of players, so long as the number added plus the number expelled does not exceed k.

Our motivation for this definition is concerned really only with the cases $k = 1$ or 2, and it is based on the ordinary observation that most changes in coalition structures in both the economy and among individuals occur as a sequence of changes, each one of which involves the addition or expulsion of only one or two individuals at a time.

It can easily be argued that V_k omits certain important coalitions from consideration. For example, suppose $k = 1$, then in V_1 we consider only those coalitions which are formed either by the addition or the removal of one player from the coalitions of γ, but in general such simple coalitions as $\{i,j\}$ are not under consideration as possible changes. For i and j to consider bolting their respective coalitions to form the coalition $\{i,j\}$, if it is profitable to do so, seems a very plausible event. We are thus led to define W_k: coalition S is in $W_k(\gamma)$ if and only if either

i. S is in $V_k(\gamma)$,

or ii. $|S| \leqslant k + 1$.

A third special and important case of ψ is the one which includes all possible coalitions for every possible γ; this we shall denote by $E(\gamma)$.

Observe that for any γ, $\psi(\gamma)$ is a set (of coalitions) and

so we may speak of one ψ being included in another. It is easy to see that the following relations are true

$$V_k(\Upsilon) \subset W_k(\Upsilon) \subset E(\Upsilon)$$

$$V_{n-2}(\Upsilon) = W_{n-2}(\Upsilon) = E(\Upsilon)$$

$$V_k(\Upsilon) \subset V_{k'}(\Upsilon) \quad \text{if} \quad k \leq k'$$

$$W_k(\Upsilon) \subset W_{k'}(\Upsilon) \quad \text{if} \quad k \leq k'$$

It is not hard to show that if $\psi(\Upsilon) \subset \psi'(\Upsilon)$ for every Υ, then the fact that a game is ψ'-stable implies that it is ψ-stable, and if it is ψ-unstable, then it is ψ'-unstable. Thus, V_1-instability is, in a sense, absolute instability, for no matter how limited we make the allowable changes in such a game - provided we allow some in each case - there are no stable pairs.

It can be shown that any n-person essential constant-sum game is E-unstable, and so the constant-sum 3-person game is V_1-unstable, as was suggested earlier in this section. If, however, we drop the constant-sum requirement, an example can be given of an E-stable game.

The essential constant-sum 4-person games in -1,0 reduced form have the following characteristic functions

$$v(T) = \begin{cases} 0 \\ -1 \\ 1 \\ 0 \end{cases} \quad \text{when T has} \quad \begin{cases} 0 \\ 1 \\ 3 \\ 4 \end{cases} \quad \text{elements, and}$$

$$v(\{1,4\}) = 2y_1 = -v(\{2,3\})$$

$$v(\{2,4\}) = 2y_2 = -v(\{1,3\})$$

$$v(\{3,4\}) = 2y_3 = -v(\{1,2\})$$

where the numbers y_1 may assume any values in the interval from -1 to +1. Now, it will be recalled that in section III.5.4 we mentioned that all 4-person constant-sum games are quota games, and one can easily see that the quota is

$$\| y_1-y_2-y_3,\ y_2-y_1-y_3,\ y_3-y_1-y_2,\ y_1+y_2+y_3 \|$$

It can be shown that a 4-person constant-sum game is V_k-stable if and only if it is W_k-stable. We may immediately dispose of the case $k = 2$, for from our general result about constant-sum games we know that a 4-person constant-sum game is V_2-stable. For $k = 1$, it can be shown that a 4-person constant-sum game is V_1-stable if and only if the quota is an imputation. For these V_1-stable games, the imputation X of any V_1-stable pair $(X, \mathcal{C})$ is always the quota, and one can explicitly state those $\mathcal{C}$'s for which the pairs are stable. We need not do this here.

It will be recalled that a game is called <u>simple</u> if $m(T) = 0$ or 1 for every T, where m is the 0,1 reduced form. Those coalitions T for which $m(T) = 1$ are called <u>winning</u> and those for which $m(T) = 0$ are called <u>losing</u> coalitions. It can be shown that a simple game is V_k-stable if and only if it is W_k-stable, and this stability may be characterized as follows: A simple game is V_k-unstable if and only if the intersection of all winning coalitions having $k + 1$ members is the empty set; or, stated positively, a simple game is V_k-stable if and only if either

i. there is no winning coalition which has $k + 1$ members,

or ii. there is at least one player who is a member of every winning coalition which has $k + 1$ members. For the case $k = 1$, a full description of the V_1-stable pairs $(X, \mathcal{C})$ in both cases i and ii can be given; the reader is referred to [11].

For the case $k = 1$, a more detailed result about simple games is possible, which is of some interest for it indicates how few simple games are V_1-unstable. If a game consists of two independent games on complementary sets of players, it is said to be decomposable into the two games. More precisely, if one can find a set of players T such that

$$v(S) = v(S \cap T) + v(S-T)$$

for every possible coalition S, then the game is decomposable into games on T and -T. In essence, the original game is not truly what one intuitively calls a game; it is rather a formal conjunction of two disjoint and non-interacting games. The notion is probably not of practical interest, but it must be introduced for there is nothing in the definition of a game which excludes the possibility. It can be shown that any V_1-unstable simple game is decomposable into the 3-person constant-sum game and the (n-3)-person inessential game. But since the inessential game is trivial in a theory of coalition formation, the theory of V_1-unstable simple games is identical to the theory of the 3-person constant-sum game. In effect, then, we know that aside from the 3-person game, there are no other "absolutely unstable" simple games.

We may now consider the uniqueness of ψ-stable pairs. First, it is clear from the above that there are some games which for a particular choice of the function ψ are ψ-unstable, i.e., no stable pair exists. The theory predicts no equilibrium behavior for such situations, e.g., the V_1-instability of the 3-person constant-sum game. For ψ-stable games there is in general more than one equilibrium point. With ψ restricted to either V_1 or W_1, a constant-sum 4-person game is either unstable or it has a unique imputation (the quota) which occurs in all stable pairs. But for

some of the stable 4-person games there is more than one system of coalitions which, combined with the quota, are stable. The theory does not decide which will occur in practice. This situation is analogous to certain physical problems in which there are several points of equilibrium. There it is found that the one which will occur depends on the initial point of the full dynamic system, and that to predict it a full dynamic theory, not just an equilibrium theory, is required. The analogy seems so close that, at least for the present, we shall assign some of the failure of this stability theory to predict a unique outcome to a lack of a full dynamic theory of coalition formation. However, it appears that there may be a further ambiguity which will not be removed by a dynamic theory. Consider a simple game in which there is only one 2-element winning coalition, say $\{1,2\}$. Then it is not difficult to show that the pair

$$\left(\| p, 1-p, 0, \ldots, 0 \| , \left[\{1,2\}, \{3\}, \ldots, \{n\} \right] \right)$$

is V_1-stable, where $0 < p < 1$. The theory does not decide on the value of p, which presumably rests on the bargaining abilities of the two players.

Certain summary comments are in order. Mathematically, the concept of ψ-stability for $\psi = V_k$ and W_k is comparatively easy to work with, much easier, say, than the von Neumann-Morgenstern solution. Evidence of this is the fact that we were able to state certain complete stability results for all constant-sum 4-person games and for all simple game; it will be recalled that only for a limited number of these games has it been possible to obtain complete sets of solutions. It is also of interest that these definitions led to results closely tied into other concepts of game theory - quota games, decomposition of games, etc. Thus from the mathematical point

of view one feels that the definition is justified. From the point of view of social science, more is needed; the definition must have some intuitive merit and, possibly, some empirical merit. It would appear, to a biased author, that the stability notion does have some merit conceptually, because it deals simultaneously with changes in imputations and coalitions, and empirically, because it is often easier to determine the coalition structure of an existing situaition than the payments in that situation. A comparison with experimental data will be discussed in section III.7.1.

Nonetheless, at least one important criticism can be levelled at it. The introduction of the peculiar function ψ, a function which is not explicit in most real situations, is hard to defend adequately. Where there are "standards of behavior" which are implicit, or at least vague, and which are not rigidly enforced, it may be possible to estimate ψ, but there is no assurance - as the theory assumes there is - that someone will not violate it. A possible remedy comes to mind which has not yet been examined. Suppose that instead of assuming the dichotomy, i.e., that a coalition is either ψ-critical or not, we assign to each possible coalition S a probability $p(S, \mathcal{C})$ for each $\mathcal{C}$, which is to be interpreted as follows: $p(S, \mathcal{C})$ is the probability that a change to S will be considered when the players are in the coalition system $\mathcal{C}$. With these given, the theory can be constructed as before, except for assertions of the form "$(X, \mathcal{C})$ is ψ-stable," which will be replaced by "$(X, \mathcal{C})$ is stable with probability p."

Aside from the above proposal, two problems for future research come to mind. First, it is at least mathematically interesting to know under what conditions an imputation X of a V_k-stable or a W_k-stable pair $(X, \mathcal{C})$ is in a von Neumann-Morgenstern solution; this is a real problem,

for in section III.6.3 we shall present an example of such an imputation which is not contained in a solution.

Second, an attempt should be made to devise a dynamic theory which describes the movement from one unstable pair to another pair until a stable one is finally reached, all changes being made within the limitations prescribed by a function ψ. The mechanism of a change should, of course, be the existence of a positive gain for the players participating in the change. The difficulty in giving such a theory seems to stem primarily from the fact that from any given $(X, \mathcal{C})$ there may be several different and incompatible changes in the coalition structure which are all admissible and all profitable; how will it be decided which will occur?

6.2 Value

The next topic is not concerned with the outcome of the game, but rather with an a priori valuation of the game for each of the players. Shapley writes, "In attempting to apply the theory [of games] to any field, one would normally expect to be permitted to include, in the class of 'prospects,' the prospect of having to play a game. The possibility of evaluating games is, therefore, of critical importance. So long as the theory is unable to assign values to the games typically found in application, only relatively simple situations - where games do not depend on other games - will be susceptible to analysis and solution." [22, p. 307]

The solution to this problem for 2-person games is taken to be the minimax value, but certainly this is not suitable in n-person games where coalitions are allowed, for the whole point of joining coalitions in essential games is to do better than $v(\{i\})$. Presumably the "value" for any i will

depend on the values of v(T) for each coalition T having i as one of its members. Just what function would be reasonable to select is not, on the face of it, obvious, and certainly an ad hoc definition would be questioned and countered by other suggestions. Rather than doing this, Shapley employed the more elegant procedure of stating certain requirements as intuitively necessary properties of any notion of numerical value; he listed three apparently weak ones and then, surprisingly, he was able to show that these uniquely determine a value - that there can be only one function satisfying the three conditions, and that there is one.

Suppose a game is given by the characteristic function v. From this game we may generate others by permuting the labelling of the players, but abstractly all of the games are the same one. Shapley's first condition is:

i. Value shall be a property of the abstract game, or more formally, if π is a permutation of the players resulting in a game which we may denote πv, and if $\phi_i(v)$ denotes the value of the game v for player i,

$$\phi_{\pi i}(\pi v) = \phi_i(v).$$

His next condition is:

ii. The individual values of the game form an additive partition of the value of the whole game, i.e.,

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

Now suppose v is a game on the set of players R and w a game on S, where R and S may or may not overlap. We may extend v and w both to the set $R \cup S$ by defining

$$v(T) = v(R \cap T) \quad \text{and} \quad w(T) = w(S \cap T), \quad \text{where } T \subset R \cup S.$$

Suppose we think of these two games as being played by the players $R \cup S$ but played completely independently of one another. This composite game (which includes the notion of decomposable games defined in section III.6.1) may be treated as a single game, called the sum of v and w, with the characteristic function $v(T) + w(T)$. Shapley's last condition is:

iii. For two such games v and w,

$$\phi_i(v + w) = \phi_i(v) + \phi_i(w),$$

or in words, the value of a game composed of two independent games is the sum of the values.

One could hardly ask less of a numerical value; what is surprising is that one need not - dare not - demand more, for these three conditions are sufficient to determine ϕ_i uniquely, and indeed, one can obtain an explicit formula for it, namely,

$$\phi_i(v) = \sum_{S \subset I_n} \gamma_n(s) \left[v(S) - v(S - \{i\}) \right]$$

where $s = |S|$ and $\gamma_n(s) = (s-1)!(n-s)!/n!$

As pointed out by Kuhn and Tucker, Shapley's "result can be interpreted by imagining the random formation of a coalition of all of the players, starting with a single member and adding one player at a time. Each player is then assigned the advantage accruing to the coalition at the time of his admission. In this process of computing the expected value for an individual player all coalition formations are considered as equally likely." [10, p. 303]

6.3 Reasonable Outcomes

Milnor has published a paper [16] in which he takes up the problem of the outcome of a game, and though his definitions are different, the viewpoint is similar to that of Vickrey, Shapley, and Luce above. The attempt is to impose reasonable conditions which isolate a subset of the set G (section III.5.6) of "outcomes" subject to "...the point of view that it is better to have the set too large rather than too small. Thus it is not asserted that all points within one of our sets are plausible as outcomes; but only that points outside these sets are implausible." [16, p. 2] Examples of such subsets in order of decreasing size are the set of outcomes G, the set of efficient outcomes E, the set of imputations I, and the set of imputations which are in at least one von Neumann-Morgenstern solution. Milnor introduces three more conditions, each having a certain degree of reasonableness, and he examines some of their properties.

First, for any player i one may examine the largest contribution he makes to any coalition, i.e.,

$$b(i) = \max_{S} \left[v(S) - v(S - \{i\}) \right]$$

Milnor defined the set B to be those outcomes X of G such that for every i, $x_i \leq b(i)$. He argues that "In any play of the game, player i will wind up in some coalition S. The players of $S - \{i\}$ would be foolish to keep i in their coalition if he tries to get so much that they could do better without him." [16, p. 3] This argument seems questionable if not irrelevant, for one can have a game with the following property: For player i in coalition S there is no temptation to move from coalition S to coalition T but if j in S moves to T then there is a profit for i to move from $S - \{j\}$ to $T \cup \{j\}$. In that case, if i is important to S, it may behoove the coalition to pay j more

than his incremental contribution in order to keep both i and j. As an example, it is easy to give simple games with coalitions S and $T = -S$ and players $i, j \in S$ such that S, $S - \{j\}$, $T \cup \{i\} \cup \{j\}$ are winning and $S - \{i\}$, T, $T \cup \{i\}$, and $T \cup \{j\}$ are losing. It is reasonable for S, if it gets more than 0, to pay j more than $v(S) - v(S - \{j\}) = 0$ in order to keep j and therefore i. Nonetheless, it may be unreasonable to pay him more than b(j) but Milnor's argument is not really directed to this point.

For the 3-person constant-sum game, B contains the set I of imputations. For the 4-person constant-sum games, B does not contain all of I, but judging by one example, it does include a sizable portion of it. In general, it can be shown that B includes both the Shapley value (III.6.2) and all von Neumann-Morgenstern solutions (III.5.1). It is not difficult to show that the imputations of the W_k-stable pairs of any simple game and of any 4-person constant-sum game are in B, and it would not be surprising were this generally true, but it is not. For example, suppose $n \geq 4$ and

$$m(\{i\}) = m(\{i,j\}) = 0,$$
$$m(T) = |T|/n \quad \text{for} \quad |T| \geq 3.$$

The pair $(\| 0,0,\ldots,0,1 \| , [\{1\},\{2\},\ldots,\{n\}])$ is W_1-stable since

$$m(\{i,j\}) = 0 \leq x_i + x_j, \quad i \neq j.$$

Observe that for this game $b(i) = 3/n$ and so for $n \geq 4$,

$$x_n = 1 > 3/n = b(n),$$

that is, $\| 0,0,\ldots,0,1 \|$ is not in B. Here, as so often in mathematics, we find the intuitions of various people in conflict, for both Milnor's conditions and those of W_k-stability have a certain intuitive reasonableness, and yet there are cases - admittedly slightly pathological ones - in which one

or the other must go.

From the above example and from the stated result that any imputation of a von Neumann-Morgenstern solution is in B, it follows immediately that there are imputations of W_k-stable pairs which do not belong to any solution.

Next, Milnor introduces a lower bound for payoffs to coalitions.

Let $$l(S) = \min_{S' \subset S} \left[v(S') + v(S - S')\right]$$

and let L be the set of outcomes (subset of G) such that

$$\sum_{i \in S} x_i = l(S) \quad \text{for all } S \subset I_n.$$

In words, l(S) is the worst that could happen to the players of S if they split into two warring factions. If one assumes that the bargaining of the game will result in two opposing coalitions, and that in order for a coalition T to form it must distribute its payoff in such a fashion that every subset T' of T is given at least v(T'), then the outcome will fall in L.

It can be shown that for 3- and 4-person constant-sum games L is exactly the intersection of B with I. This cannot be generally true, for we know that the intersection of B and I includes the von Neumann-Morgenstern solutions and Shapley's value, and an example can be given both of a game with a von Neumann-Morgenstern solution not wholly in L and of one with the Shapley value not in L. It is not known if L is always non-empty, though Milnor gives a wide class of games for which L is not the empty set.

The final concept is, at least conceptually, somewhat related to that of ψ-stability. A total payment ξ to a coalition S is called an <u>unreasonable demand</u> if there is an outcome X such that

i. X is feasible with respect to the opposing coalition, i.e.,

$$\sum_{i \in -S} x_i \leqslant v(-S),$$

and ii. no subset of $-S$ can be induced to join S in such a way that S receives δ, i.e., for every $T \subset -S$,

$$\sum_{i \in T} x_i > v(S \cup T) - \delta$$

If we define

$$d(S) = \min_{\sum_{i \in -S} x_i = v(-S)} \quad \max_{S' \supset S} \left[v(S') - \sum_{S' - S} x_i \right]$$

then it is not difficult to show that δ is unreasonable if and only if $\delta > d(S)$. D is defined to be the set of outcomes X such that for each subset S, $\sum_{i \in S} x_i \leqslant d(S)$. Relatively little is known about the set D, but an example can be given where Shapley's value is not a member of D, and for the 3-person constant-sum game the intersection of D and E is closely related to the symmetric von Neumann-Morgenstern solution F (it is the simplex spanned by the three points F_1, F_2, F_3).

The principal interest in these definitions resides in the experimental work which was performed in conjunction with them, and which will be discussed in section III.7.1. Mathematically it is not easy to judge them, for as we have seen, relatively few results are known, and while the intuitive considerations which led to the definitions are of vital importance, it is not until the consequences of the definitions are known that one can critically evaluate these intuitions.

6.4 Summary

In this section three theories different from the solution construct, but each based on the characteristic function of a game, were given. The first supposes that the end product of coalition formation, after all the

changes in alliances and threats of such changes have been concluded, will be a pair $(X, \mathcal{C})$, where X is an imputation describing the payments agreed upon and where $\mathcal{C}$ describes the partition of the players into coalitions. The theory attempts to characterize these equilibrium pairs. A function $\psi(\mathcal{C})$ is assumed to be given which states, for each partition $\mathcal{C}$ of the players into coalitions, to which coalitions the players may consider changing from $\mathcal{C}$. The idea behind ψ is that changes in alliances are gradually effected and that only coalitions which are "near" a coalition of $\mathcal{C}$ are acceptable as possible changes from $\mathcal{C}$. For a given ψ, a pair $(X, \mathcal{C})$ is called ψ-stable only if none of the admissible coalitions (according to ψ) can guarantee a profit to the players in the coalition, i.e., if

i. for every $S \in \psi(\mathcal{C})$, $v(S) \leq \sum_{i \in S} x_i$;

and if each of the players in a non-trivial coalition of $\mathcal{C}$ is guaranteed more than he could expect to receive were he to play alone, i.e., if

ii. for every $i \in T$ where $T \in \mathcal{C}$ and $|T| > 1$, $x_i > v(\{i\})$.

Using this definition for two specific classes of functions ψ, certain theorems about all 4-person constant-sum games and all simple games were stated.

Objections were raised to introducing the function ψ since it is not generally part of the rules of a game and, at least in its present non-probabilistic form, it is not to be expected that it can be observed empirically. One can look at it, however, as an explicit ad hoc assumption which replaces the a posteriori verbal discussions necessary with the solution theory.

In the second subsection the problem was raised as to an a priori

valuation by each of the players of a game in characteristic function form. The problem was approached by stating three conditions which one feels intuitively should be met by such a value, namely: it should be a property of abstract games and independent of their particular representation; it should be an additive partition of the total value of the game, $v(I_n)$; and the value to each player of a game which is the "sum" of two games should be the sum of his values in the two separate games. These three conditions determine a unique value in terms of the characteristic function. An interpretation was given by supposing that the coalition of all players is formed by randomly choosing one player and (with equal likelihood) randomly adding one player at a time. If each player is assigned the increment he adds to the coalition at the time he is selected, then the value to each player is the expected value of his increment.

In the final subsection three different and intuitively plausible restrictions were placed on imputations to isolate classes which are "reasonable outcomes," at least in the sense that any imputation not in the class is considered unreasonable. Some questions were raised as to the arguments supporting these definitions, but no final decision as to their merit seems possible at the moment since so few mathematical results are known involving the conditions.

7. Empirical Study of Games

7.1 An Experiment

Notably lacking in all of our discussion so far have been data, or even the mention of data. In part this may be attributed to the realization

that the theory of games is inadequate as a descriptive theory, for human beings simply do not have the perception assumed by any of the theories. But two other reasons are actually more important. Assuming that we wish to check a coalition theory, it is necessary that we know the characteristic function; but we have already pointed out the great difficulty of determining the normal form of any existing game situation and, even assuming that known, the extensive calculations required to obtain the characteristic function (III.3.1 and III.4.1). Without it, we cannot know what any of the theories predict. In addition, suppose the characteristic function is known, then what does the principal theory - von Neumann and Morgenstern's solutions - predict? In discussing the outcome of an experiment run at RAND, the authors remark "It is extremely difficult to tell whether or not the observed results corroborate the von Neumann-Morgenstern theory. This is partly because it is not quite clear what the theory asserts. According to one interpretation a 'solution' represents a stable social structure of the players. In order to test this theory adequately, it would probably be necessary to keep repeating a game, with a fixed set of players, until there seemed to be some stability in the set of outcomes which occurred. One could then see to what extent the outcomes of this final set dominate each other and to what extent other possible imputations are not dominated by them." [6, p. 23]

It appears to us that the most important problem of empirical verification is to develop a method to determine the characteristic function. Of relevance here is our earlier remark that nothing less than a characteristic function could represent a numerical evaluation of coalition strength, and

that the characteristic function governing the behavior of individuals may very well not be the theoretical one of the game they are playing. Probably one of the most important immediate contributions social scientists can make in this area is practical empirical methods of determining the characteristic function approximately. In section III.7.2 we shall propose a method for doing this, one, however, which appears to raise as many problems as it solves. Also, it may not be necessary to determine the entire characteristic function if some such theory as the stability one applies. For suppose there are restrictions on coalition change and we wish to determine whether the present state of affairs is in equilibrium, then we need only determine the characteristic function for the admissible coalitions.

In the laboratory this problem can be by-passed, at least in part, by describing the game in terms of the characteristic function. This is exactly what Kalisch, Milnor, Nash, and Nering have done at RAND [6]. We shall report only the main portion of their experiment, which was concerned with two 4-person constant-sum games. Each game was presented to the subjects in what amounted to a 0,1 reduced form and in an S-equivalent form. For each coalition the subjects were told what the coalition would receive. They were then given 10 minutes to form coalitions and to agree upon payments, which were to be told to an umpire. He reported the agreements back to the group and if there was no dissension he held the players rigidly to the formal agreements at the end of the bargaining. The authors point out that there were in addition numerous informal agreements which were not processed through the umpire and which were kept in good faith.

We feel that the general qualitative impressions of the authors,

while not very surprising, are of sufficient importance to be quoted at length:

"There was a proclivity for members of a coalition to split evenly, particularly among the first members of a coalition. Once a nucleus of a coalition had formed, it felt some security and tried to exact a larger share from subsequent members of a coalition. The tendency for an even split among the first members of a coalition was in part due to a feeling that it was more urgent to get a coalition formed than to argue much about the exact terms.

"Another feature of the bargaining was a tendency to look upon the coalitions with large positive values as the only ones worth considering, often overlooking the fact that some players could gain (sic) a coalition with a negative value to their mutual benefit...

"Coalitions of more than two persons seldom formed except by being built up from smaller coalitions. Further coalition forming was usually also a matter of bargaining between two rather than more groups.

"A result of these tendencies was that the coalition most likely to form was the two-person coalition with the largest value, even though this coalition did not always represent the greatest net advantage for the participants; and in the interest of speed, this coalition usually split evenly. Thus it frequently happened that the player with apparently the second highest initial advantage got the most of the bargaining. The player with the apparently highest initial advantage was most likely to get into a coalition, but he usually did not get the larger share of the proceeds of the coalition.

"Initially the players were more inclined to bargain and wait or

invite competing offers. And this remained true to some extent in those games where the situation did not appear to be symmetric. However, later and in those games which were obviously symmetric, the basic motive was to avoid being left out of a coalition. Hence there was little bargaining, and the tendency was to try to speak as quickly as possible after the umpire said 'go,' and to conclude some sort of deal immediately. Even in a game which was strategically equivalent to a symmetric game, the players did not feel so rushed. We would guess that this was because some players felt they were better off than the others whether or not they got into coalitions, while others felt that they were worse off whether or not they got into coalitions. They seemed to pay little attention to the fact that the net gain of the coalition was the same to all." [6, p. 15-16]

"Personality differences between the players were everywhere in evidence. The tendency of a player to get into coalitions seemed to have a high correlation with talkativeness. Frequently, when a coalition formed, its most aggressive member took charge of future bargaining for the coalition. In many cases, aggressiveness played a role even in the first formation of a coalition; and who yelled first and loudest after the umpire said 'go' made a difference in the outcome.

"In the four-person games, it seemed that the geometrical arrangement of the players around the table had no effect on the result; but in the five-person game, and especially in the seven-person game, it became quite important. Thus in the five-person game, two players facing each other across the table were quite likely to form a coalition; and in the seven-person game, all coalitions were between adjacent players or groups of players. In general as the number of players increased, the atmosphere became more con-

fused, more hectic, and less pleasant to the subjects. The plays of the seven-person game were simply explosions of coalitions formation.

"In spite of an effort to instill a completely selfish and competitive attitude in the players, they frequently took a fairly cooperative attitude. Of course, this was quite functional in that it heightened their chances of getting into coalitions. Informal agreements were always honored. Thus it was frequently understood that two players would stick together even though no commitment was made. The two-person commitments which were made were nearly always agreements to form a coalition with a specified split of the profits, unless a third player could be attracted, in which case the payoff was not specified. This left open the possibility of argument after a third party was attracted, but such argument never developed. In fact, the split-the-difference principle was always applied in such cases." [6, p. 16-17]

We have quoted at such length for three reasons. First, it is important when evaluating the results that the reader have some flavor of the procedure and of the performance. Second, it is interesting that the coalition changes were effected, in the early stages, one person at a time, and in the later stages by one small coalition joining with another. Third, certain aspects of the experimental procedure seem undesirable and could easily be eliminated. The geometrical effects, though possibly interesting in some applications, are not desirable in a study of human response to characteristic function. To eliminate this one might employ telephone communication or a variant on the Bavelas partitioned table for small group studies [3] . The latter would require the use of written messages, which incidentally, would give a permanent record of the bargaining. It would have the slowing effect

that any written communication has, but it is not clear that this would be a disadvantage in this case. Further, in the small group work it was observed that a high degree of anonymity was preserved, and this might allow more ruthless competition than was obtained at RAND.

Each 4-person game was played eight times, a total of eight subjects being employed. Changes in the players were made for each play so that permanent coalitions would not tend to form. The data presented in Figs. 6 and 7 are adapted from [6] in the sense that we have added what W_1-stability theory would predict the payments should be. Probably the most striking fact in the data is the difference between S-equivalent games; it is clear that the subjects had not gotten to the logical base of the matter. As far as prediction goes, the Shapley value and the quota (which is the imputation of the W_1-stable pair) are identical in the symmetric game (Fig. 7) and nearest to the symmetric (reduced form) presentation of that game. For the non-symmetric game, the value and quota differ. The latter is reasonably near the reduced form presentation, but not at all near the S-equivalent form; the reverse is true of the value.

They also present data on the coalitions which actually formed in each play of the game. By W_1-stability theory one expects $\{A,B,C\}$ to form in the non-symmetric game; this actually occurred on only two out of eight trials. It is predicted for the symmetric case that no non-trivial coalitions will form, or one grand one. This occurred only once in eight trials, but in three other trials two opposing two-element coalitions formed and zero-payments for everyone were agreed upon. Four times a three-element coalition formed, and the isolated player was given $v(\{i\})$ and the others divided $-v(\{i\})$ with considerable discrimination against the third addition to the coalition,

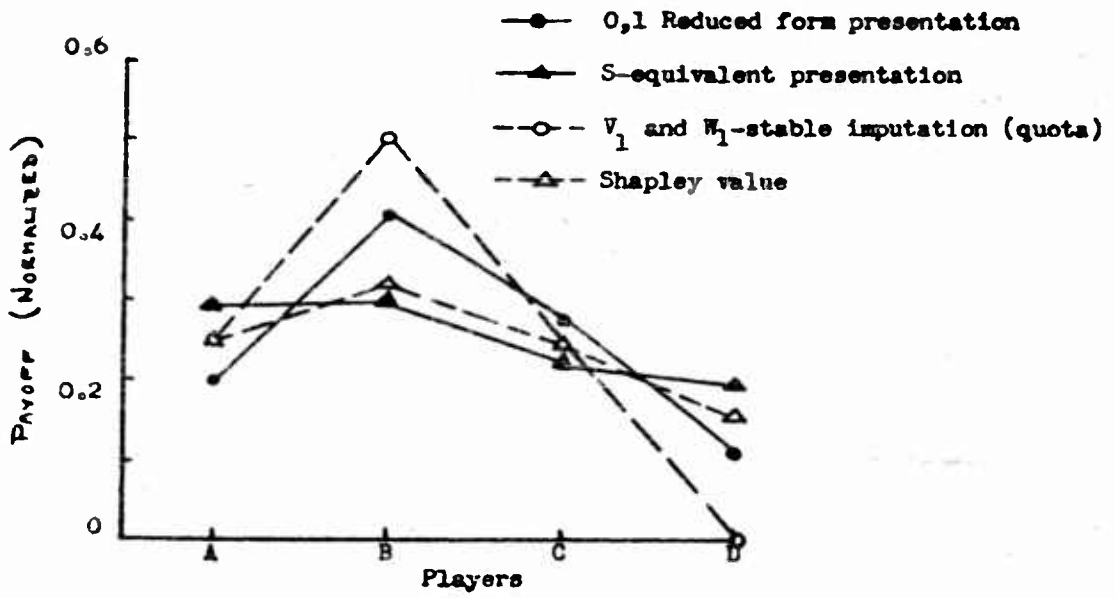

Fig. 6. 4-Person Constant-Sum Game:

$$y_1 = y_3 = \tfrac{1}{2}, \quad y_2 = 0$$

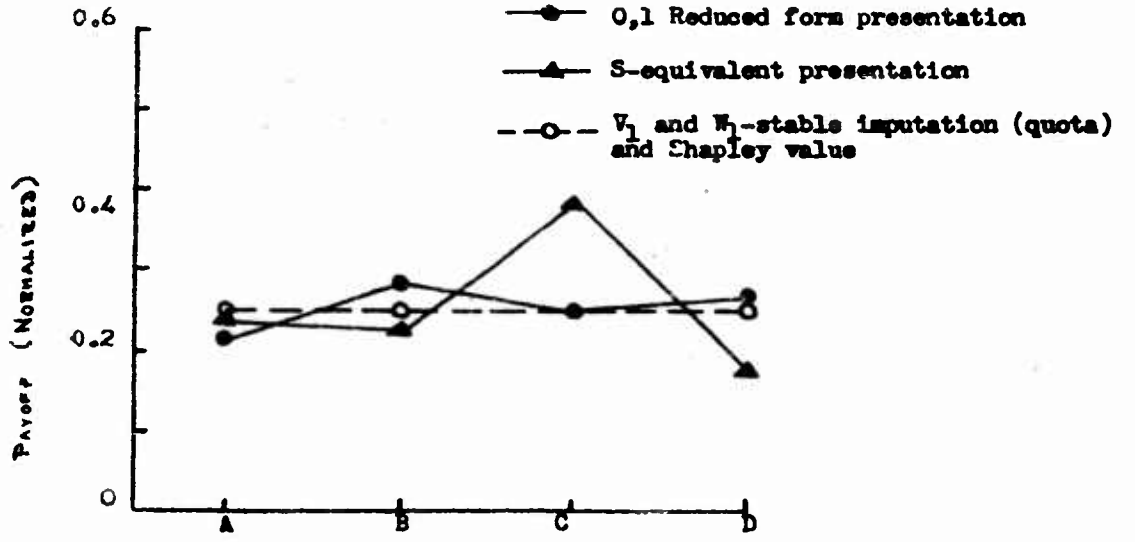

Fig. 7. 4-Person Constant-Sum Game:

$$y_1 = y_2 = y_3 = 0$$

These figures have been adapted from RAND Corporation Memorandum RM-948.

except in one case when the three-element coalition was formed without a two-element intermediate stage.

As we suggested by the quotation presented at the beginning of this section, these authors did not know what the von Neumann-Morgenstern theory asserts in such an experiment, and so essentially no comparison was possible.

It will be recalled that Milnor (III.6.3) defined reasonable bounds for the payoffs, and the data were compared with these. Only once in the 4-person games did one player get as much as or more than the bound b(i), but in most of the plays at least one set S received more than the upper bound d(S). It was concluded that "...the function d(S) seems to have no relation with the way the game was actually played." [6, p. 27] Comparison with the lower bound l(S) was not made except in the 7-person game (which was seriously influenced by the experimental conditions). This game was constructed so that there was a von Neumann-Morgenstern solution giving a set S less than l(S), but it was found that in both plays of the game each set S actually got at least l(S).

Certainly this experiment cannot be considered to be crucial. It is clear that the results do not coincide exactly with any present theory, but it is questionable how much the outcome was influenced by the experimental technique. One senses from the report that the time pressure was high, which seems to be opposed to the assumption in the theory of almost all-knowing players. Furthermore, the geometrical obstacles to coalition formation are certainly not a part of the theory, though this remark may not apply to the 4-person games. More significant, and probably generally true, is t

observation that the subjects do not always respond to the strategic consequences of the characteristic function alone, but sometimes to its mode of presentation as well. We shall return to this point in the next section.

These last comments raise the whole problem of what experiments and experimental procedures will be considered acceptable tests of the theories. While it seems impossible to give an exact prescription of a good experiment in this or any other science, it is often possible to assert that one feels a certain procedure is not the best possible, and this is what we have done. Any deeper comments will lead us into the knotty problem of the relationship of theory and experiment and this is not the place for such a discussion.

7.2 A Method for Empirically Determining Characteristic Functions

We have pointed out earlier that the two major deterrents to applying n-person theory to real situations have been the lack of an adequate descriptive theory based on the characteristic function and the practical impossibility of determining the characteristic function of an existing situation. The latter difficulty stems from the fact that the only way known to obtain the characteristic function is to ascertain the normal form of the game and then to make elaborate calculations involving the minimax theorem. Not only is it next to impossible to find the normal form of a game in an existing situation, but considering the billions of strategies available in any reasonably complex situation the minimax theorem calculations would be completely impractical.

One wonders, therefore, whether there are empirical techniques which can be used to obtain an approximation to the characteristic function directly. One suggestion, offered by Adams and one of the present authors [1],

is based on the following simple idea: A player is required to report his preferences between pairs of possible coalitions of players, these preferences to be based on his conception of their relative strengths. (The definition of preference will be discussed more fully below.) No assumption is made that he knows the underlying normal form of the game or the game theory analysis of it, but rather, his response is based on his subjective evaluations of coalition strength - on the evaluations which presumably govern his behavior. If these evaluations satisfy the von Neumann-Morgenstern utility axioms [21] and one other plausible axiom, then there is a family of set functions which are closely related to the utility functions determined by the von Neumann-Morgenstern axioms and which satisfy the two conditions of a characteristic function.

Before presenting the details of this proposal, let us briefly summarize the von Neumann-Morgenstern utility axioms. Let A denote a set of alternatives. If $R, S \in A$ and if $0 < \alpha < 1$, let $\langle \alpha R, (1-\alpha)S \rangle$ denote the prospect "alternative R with probability α and alternative S with probability $1-\alpha$." Starting with A generate all the possible risk alternatives of the form $\langle \alpha R, (1-\alpha)S \rangle$ and call the resulting set K. K is closed in the sense that if $R, S \in K$ then $\langle \alpha R, (1-\alpha)S \rangle \in K$. We suppose that an abstract binary relation $\preceq$ is defined over K (which we shall ultimately treat as a preference-or-indifference relation, so that if $R \preceq S$, where $R, S \in K$, then the person imposing $\preceq$ on K either prefers S to R or is indifferent between S and R.) If both $R \preceq S$ and $S \preceq R$, then we write $R \sim S$ (and we say R is indifferent to S). If $R \preceq S$ and not $R \sim S$, then we write $R \prec S$ (and we say S is strictly preferred to R).

The relation $\prec$ is said to satisfy the von Neumann-Morgenstern utility axioms if it is a simple ordering of K and

1. if $R \prec S$, then $R \prec \langle \alpha R, (1-\alpha) S \rangle$
2. if $R \succ S$, then $R \succ \langle \alpha R, (1-\alpha) S \rangle$
3. if $R \prec T \prec S$, then there exists an α such that $\langle \alpha R, (1-\alpha) S \rangle \prec T$,
4. if $R \succ T \succ S$, then there exists an α such that $\langle \alpha R, (1-\alpha) S \rangle \succ T$,
5. $\langle \alpha R, (1-\alpha) S \rangle \sim \langle (1-\alpha) S, \alpha R \rangle$
6. $\langle \beta \langle \alpha R, (1-\alpha) S \rangle, (1-\beta) S \rangle \sim \langle \alpha\beta R, (1-\alpha\beta) S \rangle$
7. if $R \sim S$, then for any α and for any T,
$\langle \alpha R, (1-\alpha) T \rangle \sim \langle \alpha S, (1-\alpha) T \rangle$

If $\prec$ satisfies these axioms, then it can be shown that there exists a family $U(\prec)$ of functions from K into the real numbers, called utility functions, such that each $u \in U(\prec)$ satisfies the following two conditions for every R and $S \in K$ and $0 < \alpha < 1$,

i. $R \prec S$ if and only if $u(R) < u(S)$,

and ii. $u(\langle \alpha R, (1-\alpha) S \rangle) = \alpha u(R) + (1-\alpha) u(S)$.

Furthermore, it can be shown that any two members of $U(\prec)$ are linearly related.

Now suppose we have a game situation involving n players and take A to be the set of all subsets of I_n, i.e., all possible coalitions. K is then the set of risk alternatives generated from A, a typical one being "coalition R with probability α and coalition S with probability $1-\alpha$." An observer, possibly one of the players of the game, is to report his preferences for all possible pairs of risk situations (R,S), where $R,S \in K$, under the following assumptions:

i. if he chooses "coalition T with probability α " then with probability α he will receive the total payment that T obtains from the situation in which -T forms a coalition and the game is played between T and -T; if he chooses ϕ, the empty set, he will neither win nor lose.

ii. if he chooses $\langle \alpha R, (1-\alpha)S \rangle$ then he has chosen R with probability α and S with probability $1-\alpha$, where these expressions are defined in i.

Let $\prec$ denote the preference relation so induced on K.

Intuitively, it does not seem unreasonable to suppose that a consistent evaluation of coalition strength should cause $\prec$ to satisfy each of the von Neumann-Morgenstern axioms. While it is unreasonable to expect that people will be so consistent, one may hope that in some cases they will be approximately consistent, in other words, that our model of a player's subjective evaluation of coalition strength is approximately correct.

Furthermore, if R and S are two non-overlapping coalitions in A, then $R \bigcup S$ is at least as strong as R and S separately, hence the alternative of receiving the proceeds of $R \bigcup S$ with a probability of $\frac{1}{2}$ and not participating with a probability of $\frac{1}{2}$ should be no less appealing then the alternative of receiving the proceeds of coalition R with a probability $\frac{1}{2}$ and receiving those of S with a probability of $\frac{1}{2}$. If this intuition is correct, then we may assume the further axiom

8. If $R,S \in A$ and $R \bigcap S = \phi$, then

$$\langle \tfrac{1}{2}R, \tfrac{1}{2}S \rangle \eqsim \langle \tfrac{1}{2}(R \bigcup S), \tfrac{1}{2}\phi \rangle .$$

The assumption that $\prec$ satisfies axioms 1-7 implies the existence of the set $U(\prec)$ of utility functions. If $u \in U(\prec)$, then define C(u) to

be the set of all set functions of the form

$$v(S) = c\left[u(S)-u(\phi)\right] + \sum_{i\in S} a_i$$

where c is a positive constant and the a_i's are constants. It can be argued that the function with $c = 1$ and $a_i = 0$ is a numerical representation of the strength of coalitions as described by the player through the preference relation $\dashv$. More than that, the following theorems* can easily be proved:

i. if $v\in C(u)$, then v is a characteristic function;

ii. if $v\in C(u)$, then $v'\in C(u)$ if and only if v and v' are S-equivalent;

iii. if $u, u'\in U(\dashv)$, then $C(u) = C(u')$.

In words, it does not matter which utility function we use from $U(\dashv)$ for they all generate the same set of functions, C(u), which set consists exactly of one of the equivalence classes of S-equivalent characteristic functions.

In addition to the intuitive argument that a member of C(u) represents the observer's evaluation of coalition strength, one can show that if he knows the game structure of the situation and if he bases his evaluation on that knowledge, then the resulting characteristic functions are S-equivalent to that of the game. Specifically, suppose the game is known in normal form and that the characteristic function v is determined by the method given by von Neumann and Morgenster [21]. Let v be extended from A to K by the following definition

$$v(\langle \alpha R,(1-\alpha)S \rangle) = \alpha v(R) + (1-\alpha)v(S).$$

---

* It is not difficult to show that the same theorems obtain if the person imposing the relation assumes he will receive the average value of the payments to players in the coalition of his choice provided axiom 8 is replaced by

8'. For any $R,S\in A$ such that $R\cap S = \phi$, $R\cup S \vdash \left\langle \frac{|R|}{|R\cup S|}R,\ \frac{|S|}{|R\cup S|}S \right\rangle$

($|R|$ = number of elements in R), and provided the class of characteristic functions is defined by

$$v(R) = c|R|\left[u(R)-u(\phi)\right] + \sum_{i\in R} a_i .$$

Let the observer determine $\prec$ according to the rule

$$R \prec S \text{ if and only if } v(R) < v(S).$$

It is easily shown that $\prec$ satisfies axioms 1-8 and so a set C(u) of characteristic functions is determined; v is one of the elements of that set.

This procedure, therefore, gives a possible method of determining v where either the normal form of a game is not known or it is far too difficult to determine it and to carry out the minimax theorem calculations. The amount of labor required is exactly the same as that needed to determine the utility function approximately in an empirical case having a comparable number of alternatives; see, for example, Mosteller and Nogee [17].

There need be no relation between the actual characteristic function of a game determined from the normal form and the subjective one determined by the above procedure using human subjects, for it is certainly not obvious, even for a person aware of the utility functions over the possible outcomes, that he will react to deductions based on them. He may react to his evaluations of the coalition alternatives more or less independently of his evaluations of the outcomes in normal form. But if this is the case, then it is a player's subjective characteristic function, and not the objective one of the game, which actually determines his behavior, and so it will be needed for predictions of his behavior. It may well be that this would account for the different results obtained for S-equivalent games at RAND (see III.7.1).

A second and more profound problem is that there is little reason to suppose that two different players of the same game will yield S-equivalent characteristic functions. If this is the case then none of the present theories is applicable. It is an open problem to devise theories for the seemingly

more realistic assumption that each player acts upon his own subjective characteristic function. It seems plausible that a theory similar to ψ-stability can be developed describing the payments and coalitions which may be expected when there is a different characteristic function associated with each player. In addition, it would be desirable to modify the notion of the normal form of a game in such a way that one can derive from it a distinct characteristic function for each of the players. Such a theory should include as a special case the von Neumann and Morgenstern reduction of the normal form to a single characteristic function. One possibility is to assume that each of the players has his own utility function over the possible outcomes and that he has beliefs as to the utility functions of the other players, beliefs which in general will be in error. This assumption results in an objective normalized game and for each of the players a fictional game which is the one he believes to exist. Assuming each player responds only to his beliefs, there is associated with each player the characteristic function of the fictional game. If each fictional game is identical to the actual game, then the theory reduces to the von Neumann and Morgenstern one.

In addition to the above theoretical developments, the proposal suggests at least two empirical studies. First, a modified version of the RAND experiment (see III.7.1) should be executed in which the subjective characteristic functions of the players are determined. Not only would this be interesting in and of itself, but it would provide inexpensive experience with the technique of determining such functions. Second, if the first study

goes well and adequate experience is gained, then the method should be applied to some existing conflict-of-interest situation and an attempt should be made to predict the equilibrium behavior using the various n-person theories. Presumably, the situation chosen should be comparatively simple and self-contained.

8. Concluding Remarks

8.1 Summary

Probably the most significant feature of general games is the possibility of communication and collusion among the players, and it is the attempt to deal with this problem in n-person game theory which makes it such a rich possibility, and certainly rich didactically, for the social sciences. The most significant features of the theory - or rather of the several theories we have presented - are (1) the equilibrium (and not dynamic) character, (2) the simple formalization of coalition strength, (3) the inadequate formalization of coalition formation, and (4) the assumption of so-called "rational" players.

Essentially three different approaches to coalition formation have been presented. Nash has treated the problem in which no communication - and so no coalitions - can occur, by extending the notion of an equilibrium point from 2-person theory. He has argued, but not completely convincingly, that introducing communication and bargaining as formal moves of the game allows cooperative games to be included within this framework. It is not clear how this should be done, but even if it were clear, it might not be adequate for social science if no explicit theory of coalition formation

resulted. Nonetheless, an adequate theory of non-cooperative games may be of importance in developing a comprehensive theory of coalition formation (see problem no. 9 of III.8.2).

The other two approaches were both based on the characteristic function form of a game which, as we pointed out, is really much more general than the normal form. In these theories the strength of coalitions, as given by the characteristic function, is used to delimit the set of imputations which may be expected to occur. The topic which has received the most extensive study is the von Neumann-Morgenstern solution, which is a set of imputations "stable" with respect to the dominance relation. It is characteristic of this theory, and of the work of Milnor, Shapley, and Vickrey along the same lines, that only the payments are prescribed; no explicit indication of the resulting coalitions is given. Indeed, the stability of a solution rests not so much on the existence of coalitions as on the potential of forming any needed coalition if an attempt is made to achieve an imputation not in the solution.

The third approach deals explicitly with the payments and coalitions which together are in equilibrium, but a significant theory seems possible only if restrictions on coalition change are made. Essentially, the dynamic model underlying the equilibrium theory assumes that only limited changes in the coalition structure can be made at any one time, and it compares the payments already agreed upon with those which can be guaranteed by the coalition if the change is made. While we feel this general type of equilibrium may prove most useful, serious objections were raised to the present definition of admissible coalition changes.

In each of these three approaches an equilibrium notion is examined

and, as always, it is an equilibrium based on contemplated but unexecuted changes - of strategies in Nash's work, of imputations in von Neumann and Morgenstern's, and of coalition structures in Luce's.

It is frequently said that the theory of games assumes rational men as players, but the term "rational" is not further explicated. One is led to imagine a fantastic calculator who will without emotion examine all possibilities and always choose the best. Certainly each theory does assume players with a considerable overview of the structure of the game and an ability to examine all possible cases; and given a definition of "best" then they do choose the best course of action. But in each theory we have - possibly implicitly - made assumptions about the exact overview the player has and exactly what he shall term the best action. The fact that these assumptions cannot be translated one into another indicates that the word 'rational' has a different meaning in each. Nash's player knows all of the strategies and the payoff function, but when he contemplates a change from one strategy to another he assumes that only he will change. In the von Neumann-Morgenstern theory he knows the characteristic function and has before him at all times the dominance relations. In the stability theory he knows the characteristic function, but presumably cannot see, effect, or contemplate certain coalition changes; but within the allowable changes the player will always act on an assured positive profit, no matter how small.

For general game theory to play a vital role in social science, two modifications appear to be necessary. First, a theory should be developed which is not unlike those presented but with somewhat more realistic assumptions - for example, the players should be assumed to have more limited perceptions. It is not easy here to meet the demands of intuition and of

mathematics. First, because the intuition, even when supported by psychological studies, is not too precise, and second, because a mathematical theory imposes certain considerations of simplicity. The latter should not be misconstrued to mean that no subtle effects can be considered in a mathematical theory (the discussion of contemporary game theory should have dispelled that view), but rather it is a demand that definitions be so chosen that there does not result a large number of special cases which must be dealt with individually. It is well to keep in mind that game theory has, so far, been almost exclusively a mathematical subject, which has received appreciation, but few contributions, from other than mathematical circles.

Second, a modification of game theory to include dynamic as well as equilibrium theories should be of wide interest and importance. It may very well be that most economic situations are not in equilibrium but in a process of dynamic change. There are dim indications of such a theory in both Vickrey's notions of strong and weak solutions and in Luce's work on stability.

8.2 Open Problems

In the course of our discussion we have raised or suggested a number of problems in n-person game theory which at the present time are unsolved or, in many cases, not even adequately formulated. It may be appropriate, if redundant, to summarize these and to add several new ones to the list. It is hardly necessary to point out that the nature of this list is markedly influenced by the research interests and activities of the authors; had we had an active interest in the theory of extensive games or in the study of von Neumann-Morgenstern solutions, for example, there is little doubt that more problems in these areas would be included in the list.

Other authors have presented lists of problems which more adequately cover these areas; see the preface of reference 9, chapter 18 of reference 13, and reference 15.

The problems are put in three classes - mathematical, conceptual, and empirical - but they are numbered consecutively. After each problem the sections of this report which appear to be most relevant are cited.

Mathematical

1. Prove the existence of a solution for every n-person game or give a counter example. (III.5)

2. For some wide class of games, characterize directly the strong solutions without attempting to determine all solutions of the games. (III.5.5)

3. Characterize those games which are V_1-unstable (also those which are W_1-unstable). (III.6.1)

4. For certain "interesting" functions $\psi(C)$, characterize those games which have at least one ψ-stable pair for which the players are partitioned into two opposing coalitions (see 9 below). (III.6.1)

5. For certain "interesting" functions $\psi(C)$, state conditions under which an imputation of a ψ-stable pair is a member of a von Neumann-Morgenstern solution. (III.5, III.6.1, III.6.3)

Conceptual

6. Present an "extensive" theory in which the temporal ordering of moves is not specified in advance and which is a suitable description of many economic situations, just as the present extensive form is a description of parlor games. Such a theory should have a natural notion of "strategy" which allows a reduction to the conventional normal form of a game, and there

should be a natural special case in which timing reduces to a specified temporal ordering and the general extensive form reduces to the present extensive game. (III.2, III.3.1, III.3.2)

7. Devise a suitable reduction of any cooperative game to a non-cooperative one along the lines suggested by Nash. It may well be that a solution to either 6 or 7 will be a solution to the other. (III.3.2)

8. Devise a theory of non-cooperative games which is more adequate than that based on the notion of an equilibrium point. There is a need for one in which the equilibrium states are more in accord with human behavior than seems to be the case for the Nash equilibrium points. (III.3.2)

9. In many cases ψ-stability theory predicts equilibrium states in which the players are divided into three or more coalitions. The theory, however, is based on the characteristic function of a game which, it will be recalled, was derived assuming that a coalition will always be opposed by the coalition of all the remaining players, and so the estimate of coalition strength is conservative when the opposition actually consists of two or more coalitions. This suggests that a "characteristic function" should be derived which depends both on the coalition S and on the arrangement of the remaining players into coalitions, i.e., a function of the form $v(S,\Lambda)$ where Λ is a partition of the remaining players, $-S$. It seems plausible to expect some form of superadditivity to hold again, certainly in the obvious generalization

$$v(R \cup S,\ [T_1,T_2,\ldots,T_t]) \geq v(R,[T_1,T_2,\ldots,T_t,S]) + v(S,[T_1,T_2,\ldots,T_t,R]).$$

In addition, it seems reasonable to suppose that if two of the coalitions which oppose R coalesce, then R will certainly not be better off, i.e.,

$$v(R, [T_1, T_2, \ldots, T_t]) \geq v(R, [T_1, \ldots, T_{i-1}, T_{i+1}, \ldots, T_{j-1}, T_{j+1}, \ldots, T_t, T_i \cup T_j]).$$

To derive the properties of such a function from the normal form of the game, it appears necessary to have a solution to problem 8, for the case of three or more opposing coalitions is a non-cooperative game with more than two players (= coalitions). On the basis of such a modified characteristic function, reconstruct ψ-stability theory. (III.4, III.6.1)

10. Devise a dynamic theory of coalition and imputation change which is in the spirit of ψ-stability theory and which has as its equilibrium points the ψ-stable pairs. (III.6.1, III.8.1)

11. Develop an equilibrium theory which predicts both imputations and coalitions, but instead of having the sharp dichotomy of ψ-stability theory as given by the function ψ, assume that each coalition S has a certain probability of being considered as a possible change when the players are (tentatively) arranged according to $\mathfrak{C}$. Presumably for most applications one would assume the probability is smaller the more different the coalition S is from the coalitions of $\mathfrak{C}$. (III.6.1)

12. Modify the assumptions about the normal form of a game so that each player has imperfect information about the utility functions of the other players, and devise a reduction process analogous to the von Neumann and Morgenstern reduction of the normal form to the characteristic function form. A possible aim in such a generalization would be a natural reduction process leading to each player having his own characteristic function on the

basis of which he evaluates the relative strengths of different coali[tions] (III.7.2)

13. Assuming that each player has his own characteristic [function], devise equilibrium theories which predict imputations and coalition [patterns] and which reduce to known theories when the characteristic functions are assumed to be the same. If the program suggested in 12 does not result in characteristic functions, devise a theory of payments and coalitions on whatever does result. (III.7.2)

Empirical

14. Devise experimental techniques to estimate the subjective characteristic functions of the different players, among other things determining whether the suggestion of section III.7.2 is suitable. (III.7.2)

15. Check the predictions of the various theories based on the characteristic function for experimental situations similar to the RAND experiment. This program should be very closely tied in with the solution of problems 13 and 14. (III.5, III.6, III.7.1)

16. Attempt to make predictions regarding the equilibrium behavior in existing but limited and isolated economic situations. Problems 13, 14, and 15 probably should be carried out first. (III.5, III.6, III.7)