BARGAINING SOLUTIONS AND STATIONARY SETS IN n-PERSON GAMES

Robert James Weber

Cornell University

Prepared for:

Office of Naval Research
National Science Foundation
Connecticut University

July 1974

UNCLASSIFIED
Security Classification

AD-787 047

DOCUMENT CONTROL DATA - R&D
(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1. ORIGINATING ACTIVITY (Corporate author)
Cornell University
Department of Operations Research
Ithaca, New York 14850

2a. REPORT SECURITY CLASSIFICATION
Unclassified

2b. GROUP

3. REPORT TITLE
BARGAINING SOLUTIONS AND STATIONARY SETS IN n-PERSON GAMES

4. DESCRIPTIVE NOTES (Type of report and inclusive dates)
Technical Report

5. AUTHOR(S) (Last name, first name, initial)
Weber, Robert J.

6. REPORT DATE
July 1974

7a. TOTAL NO. OF PAGES
92

7b. NO. OF REFS
26

8a. CONTRACT OR GRANT NO.
N00014-67-A-0077-0014

b. PROJECT NO.

c.

d.

9a. ORIGINATOR'S REPORT NUMBER(S)
Technical Report No. 223

9b. OTHER REPORT NO(S) (Any other numbers that may be assigned this report)

10. AVAILABILITY/LIMITATION NOTICES
Approved for public release; distribution unlimited.

11. SUPPLEMENTARY NOTES

12. SPONSORING MILITARY ACTIVITY
Operations Research Program
Office of Naval Research
Arlington, Virginia 22217

13. ABSTRACT

The formal proceedings of an n-person cooperative game can be modelled as a multi-stage negotiation process. At each stage, coalitions may object to the proposal at hand. If a particular objection is given recognition, then the non-objecting players respond to that objection in such a manner that a new proposal results from the objection and response. An equilibrium collection of strategies in this negotiation game may be viewed as a "standard of behavior" to which no nonconformist pressures exist. This paper explores the relationship between the set of proposals to which no objections are made in such an equilibrium collection, and the von Neumann-Morgenstern stable sets of the original game.

DD FORM 1 JAN 64 1473

UNCLASSIFIED
Security Classification

14. KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
game theory characteristic function games bargaining solutions von Neumann-Morgenstern solutions stable sets equilibria						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2*a*. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2*b*. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year; or month, year. If more than one date appears on the report, use date of publication.

7*a*. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7*b*. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8*a*. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8*b*, 8*c*, & 8*d*. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9*a*. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9*b*. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS), (S), (C), or (U)*

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.

DEPARTMENT OF OPERATIONS RESEARCH
COLLEGE OF ENGINEERING
CORNELL UNIVERSITY
ITHACA, NEW YORK

TECHNICAL REPORT NO. 223

July 1974

BARGAINING SOLUTIONS AND STATIONARY SETS IN n-PERSON GAMES

by

Robert James Weber

Cornell University
and
University of Connecticut

This research was supported in part by the Office of Naval Research under Contract Number N00014-67-A-0077, task NR-047-094, and by the National Science Foundation under Grants GK 29838 and GP 32314X.

ABSTRACT

A variety of bargaining models have been recently used to analyze solution concepts for n-person games. In this paper, the cooperative negotiation process in a side-payment game is modelled as a multi-stage game, in which each stage consists of an objection raised by some coalition to a proposal under consideration, and a response made to that objection by the remaining players. A bargaining solution to the negotiations is a collection of objection and response strategies for the players, from which no player is motivated to deviate. Associated with each bargaining solution is a set of stationary proposals, to which no objections will be raised. Thus each stationary proposal corresponds to a stable agreement between the players, in which every threat to the agreement is balanced by a counter-threat which dissuades the threatening coalition from action.

The first part of the paper lays the foundations of the bargaining theory. Motivating our approach from earlier theories, we present first the essential background definitions, and then define a bargaining game in extensive form based on a given characteristic function game. Cooperative equilibria for this game are discussed, and basic notational simplifications are derived.

The remainder of the paper applies this theory to several well-known classes of games. All stationary sets are determined for three-person games. For several types of games with stable cores, it is shown that their cores are also stationary sets. Two games, pathological in their behavior with respect to the classical von Neumann-Morgenstern theory, are

shown to be amenable to our approach. Finally, it is shown that our bargaining theory is only partially successful in treating voting games. Suggestions are made concerning possible changes in our approach to cover these games, and other possible directions of future research are discussed.

ACKNOWLEDGMENTS

I would like to express my gratitude to my advisor, William F. Lucas, for his patience and encouragement throughout the preparation of this paper. Professor Lucas, and his colleague L. J. Billera, have been sources of insight into many aspects of game theory, and I consider my association with them to have been an invaluable experience.

A special note of thanks is due to Professor John C. Harsanyi, of the University of California at Berkeley. The original inspiration for this paper was found in work of his, and although I accept full responsibility for all comments in this paper, I gratefully acknowledge his influence.

Finally, I sincerely express my thanks to Mrs. Deborah Cagey for her prompt and accurate work in the preparation of this manuscript.

TABLE OF CONTENTS

CHAPTER 1

DEVELOPMENT OF A THEORY OF BARGAINING SOLUTIONS

1. Introduction

For thirty years, the von Neumann-Morgenstern theory of stable sets has been studied as a fruitful approach to the understanding of the resolution of conflict in cooperative situations. More recently, attempts have been made to cast the original theory in a new light, by associating stable sets in some natural manner with solutions to formalized bargaining games. In this paper we formulate and investigate a model of bargaining which has some of the features of the von Neumann-Morgenstern theory, yet allows consideration of some of the dynamics of the bargaining process. A major feature of our model is that it explicitly treats not only objections which various coalitions may make to proposals, but also responses which other coalitions make to such objections.

In the first part of this paper, we discuss at length the formulation of a bargaining theory. The discussion is partly rigorous and partly heuristic, and is intended to provide a general context for the investigation of various bargaining models. The formal definition of the theory we derive appears in Section 10, and is dependent only upon a few definitions from other sections. The remaining sections are devoted to motivating this theory from more general principles. Since much of the notation necessary to these sections is unfamiliar and cumbersome, formal definitions have been replaced or supplemented with more expansive ones wherever possible.

Section 2 presents two classical approaches to the "solution" of n-person games. After briefly discussing the work of von Neumann-Morgenstern and of Vickrey, we raise certain issues which the theory of this paper is meant to resolve in part. Section 3 gives an overview of the approach to be taken, and Sections 4 and 5 present definitions which this section shows to be necessary for an analysis of bargaining problems.

Section 6 presents several formal models of bargaining games, and specifies the game with which we shall work. In Section 7, individual strategies are defined for situations in which bargaining games arise. Section 8 gives a set of criteria used to characterize collections of strategies with certain stability properties, and Section 9 uses this characterization to somewhat simplify the model, and to justify the specific formal notation used to present the theory formally in Section 10. The summary in Section 11 completes the first part of the paper.

The second part applies our bargaining theory to several classes of games. A separate introduction to that part of the paper appears as Section 12.

2. The Classical Theory, and Some Observations

Games in characteristic function form* were first considered in 1944 by von Neumann and Morgenstern (vNM) [17]. Their theory of behavior in cooperative situations is predicated on two assumptions. First, it is assumed that each coalition S of players can assure itself of a particular amount $v(S)$ of resource, independently of what the remaining

*All terms not defined in this text appear in the appendix.

players do. Second, it is assumed that any coalition may divide what it receives among its players in a completely arbitrary manner (in other words, there is no restriction on side payments between players).

vNM also proposed the concept of "stable sets" as solutions to a game. The basic feature of this solution concept is the idea of dominance. Each imputation represents an allocation of available resources among the n players. An allocation x dominates another allocation y if there is some coalition S for which every player of S receives more in x than in y, and if furthermore the players in S have the ability to guarantee themselves their amounts in x. In terms of this dominance relation, vNM defined a stable set of a game to be a collection K of imputations with the complementing properties of internal stability (no imputation in K dominates another imputation in K) and external stability (every imputation not in K is dominated by at least one imputation in K). The partly-dynamic rationale for requiring these properties lies in the following arguments, which in turn are based on the assumption that the players of the game are convinced that the imputations in a particular solution K are "sound" while the remaining imputations are "unsound". First, no coalition can use a sound imputation to discredit another sound imputation (internal stability). Also, any unsound imputation can be discredited by a sound imputation (external stability). Further, vNM note [17; pp. 265-266] that any unsound imputation which might be used to discredit a sound imputation is itself subject to discrediting by another sound imputation. This last argument will be discussed in more detail later in this section.

The vNM theory received much attention in the years following its introduction - attention well-deserved since it was the first theory which

attempted to analyze rational multi-person social behavior in cooperative situations. However, as study of stable sets progressed, it became apparent that on strictly mathematical grounds the theory contained some unpleasant results. A bewildering multiplicity of solutions existed for many games, yet the general question of whether every game had a solution was not settled. Then, in a series of results, Shapley [20,22] and Lucas [11,12,13] exhibited a number of games with particularly pathological stable set solutions, and finally Lucas [14] gave an example of a game with no stable sets.

These difficulties alone should not have lessened interest in the vNM theory, for it may be quite reasonable to believe that some social situations are inherently pathological. However, a number of philosophical objections also arose. Several of these objections are presented below.

(1) An argument presented earlier was that if y, an unsound imputation, dominates a sound imputation x with respect to a coalition S, then there is a sound imputation z which in turn dominates, and hence discredits, y. The implication was that S therefore has nothing to gain by trying to get the players of the game to consider y rather than x. However, if S can use y to shift attention from x to z, it is possible for all players of S to gain by this shift, and thus S gains by forcing consideration of the unsound imputation y.

Example 2.1: Consider the 4-person symmetric constant-sum game, with

$$v(S) = \begin{cases} 0 & \text{if} \quad |S| = 1 \\ \frac{1}{2} & \text{if} \quad |S| = 2 \\ 1 & \text{if} \quad |S| = 3,4 \ . \end{cases}$$

Let K be the stable set consisting of the imputation $(\frac{1}{4},\frac{1}{4},\frac{1}{4},\frac{1}{4})$ and the twelve permutations of $(\frac{3}{8},\frac{3}{8},\frac{1}{4},0)$. If

$$x = (0,\tfrac{1}{4},\tfrac{3}{8},\tfrac{3}{8}),$$

$$y = (\tfrac{1}{16},\tfrac{5}{16},\tfrac{9}{16},\tfrac{1}{16}),$$

and

$$z = (\tfrac{1}{4},\tfrac{3}{8},0,\tfrac{3}{8}),$$

then $z \ dom_{124} \ y \ dom_{12} \ x$, and the situation described above holds with $S = \{1,2\}$.

(2) A related objection is to the assumption that <u>immediate</u> "corrective" action (a shift to a sound imputation) will occur when an unsound imputation is proposed. It seems quite possible that attention might shift through a sequence of unsound imputations before a sound imputation is reached.

(3) The vNM theory allows any imputation y, through domination with respect to a coalition S, to be used to discredit another imputation x. However, this implies that the players of S control the allocation of resources not only among themselves, but also among the players of the complementary coalition N-S. It seems unrealistic that this should be so. Rather, it is more natural to assume that S can force a shift of attention from x to <u>some</u> imputation z with $z^S = x^S$, but that

determination of the components of z corresponding to the players in $N-S$ is made by these latter players.

Example 2.1 (continued): It may be observed that an essential feature of this example is that players 1 and 2, when presenting the imputation y, force an unequal division of resources between the remaining players. Subsequently, player 3's temporary riches in y prove to be an embarrassment when z arises, leaving him with nothing.

(4) As vNM admit [17; pp. 44-45], their theory is essentially static. It makes no pretense of considering either the dynamics of coalition formation or the effects of indirect action of the type suggested in (2), in which several successive stages may occur in the bargaining between players. Because of this static nature, the theory must consider all imputations which dominate a particular imputation to be equally credible, and thus degree or strength of dominance cannot be considered.

In an approach to the first of our objections, Vickrey [26] proposed the concept of "self-policing" sets of imputations, and investigated the existence of vNM solutions with the self-policing property. Since we retain some of the flavor of Vickrey's theory in our own approach, a brief outline of Vickrey's work is given here.

Let K be a set of imputations, and x an imputation in K. A heresy to x by the coalition S is any imputation y not in K for which $y \ \mathrm{dom}_S \ x$. A policing action to such a heresy is any imputation z in K which dominates y. If there is some player i in S for whom $z_i < x_i$ in every policing action z to the heresy y, then the heresy is suicidal for i. Finally, if every heresy to each x in K

is suicidal for some member of the heretical coalition then x is strong in K, and if all imputations in K are strong then K is a self-policing set.

Since with no further restrictions both the empty set and the set of all imputations are self-policing sets, Vickrey directed his attention to self-policing sets which are also stable sets. He found all self-policing solutions for three-person games, and characterized the constant-sum simple games for which self-policing solutions exist, showing in the latter case that when such a solution exists it is unique. Vickrey also found a number of games for which self-policing solutions do not exist, and commented that his approach seemed fruitful primarily for constant-sum games.

We note that Vickrey's theory fails to fully respond to our objections (2), (3), and (4), and also that it has a slightly ad hoc flavor imparted by the act of restricting its application solely to vNM solutions. In the next section we discuss an alternative approach which deals with all four of our previously-stated objections.

3. Bargaining Models

In view of the comments in the preceding section, it seems natural to attempt to formulate a model which gives full play to the possibilities of indirect action by a coalition in its hope of attaining an eventual goal. In dealing with indirect action, only two approaches seem reasonable. The first is, like Vickrey, to deal with all possible results of a particular action as equally likely and plausible; the second is to consider the strategic likelihood of particular results occurring. In this paper we take the second approach.

In 1950, Nash [15] proposed a general approach to solution concepts for cooperative n-person games.

> "A...type of application is to the study of cooperative games... One proceeds by constructing a model of the pre-play negotiations so that the steps of negotiation become moves in a larger non-cooperative game describing the total situation...thus the problem of analyzing a cooperative game becomes the problem of obtaining a suitable, and convincing,...model for the negotiation." [16]

The solutions of the game are then described by collections of strategies which are "in equilibrium" for the players in this larger game; that is, collections of strategies for which no player or group of players can gain by unilaterally changing their strategies in the collection. This approach has been used by both Harsanyi [7,8,9] and Selten [19] to define solutions for several forms of games.

In the particular problem to be treated here, we shall formalize the underlying bargaining procedure of a game as a multi-stage process. At the beginning of a stage, an imputation x is given to represent the proposal for final allocation which is presently under consideration. All coalitions which wish to amend this proposal by making an objection to x declare the action they wish to take. In this model, a permissible action for a coalition S is the suggestion of an allocation y^S among the players of S of a total amount not exceeding $v(S)$, where this allocation is strictly preferred by all members of S to their present shares in x. We assume that social factors in some manner determine which of the (possibly more than one) objecting coalitions is actually "given the floor" to make its suggestion. Next, the players in the complementary coalition $N-S$ respond to the action of S by agreeing on an allocation

z^{N-S} of the remaining resources $v(N) - y^S(S)$ among themselves. Thus a new proposed allocation $y^S + z^{N-S}$ is constructed, and the next stage of the bargaining process commences with this new proposal replacing x. When no coalition opposes a proposal at some stage, it is considered to be accepted by all players, and the final division of resources among the players occurs accordingly.

Several comments are in order. It should be realized that the particular final agreed-upon outcome may depend upon the initial imputation from which the first stage begins. Thus the collection of all possible final outcomes forms a set which is, in a sense, stable and which corresponds in principle to the vNM solution concept as a "standard of behavior" to which the players will conform. Also, a formal model will be needed to describe the social choice mechanism which selects at each stage a particular objecting coalition to be "yielded the floor". Finally, since players will be called upon to act in the face of uncertainty with regard to other players' actions, it will be necessary to discuss the preferences of the players over outcomes of a probabilistic nature. These last two points are treated in detail in the next sections.

4. Hierarchies

When the players of a game consider a proposed allocation of resources among themselves, it is possible that several coalitions will wish to raise objections to this proposal. If several coalitions do indeed wish to act, there must be some mechanism of society that decides which coalition is given the floor to state its objection - else the formal proceedings of the game will degenerate into an interminable shouting match. In the

real world, this selection may be governed by a number of factors: random choice, the size of the coalitions, the seniority of players in the coalitions, the relative strength of objections being raised, and so on.

For the purposes of this paper, we make one general assumption concerning the mechanism of social choice. This assumption is that the mechanism depends solely on the proposal being considered and on the collection of objecting coalitions, and therefore the mechanism is independent of time, history, and experience. The mechanism may be of a probabilistic nature, and may include the possibility that no objecting coalition is given the floor (as in, for example, a legislative body with a parliamentarian who is empowered to invoke cloture).

Specifically, we define a coalitional hierarchy H to be a function which assigns to every collection $S = \{S,T,\ldots\}$ of coalitions a "sub-probability distribution" over S. That is, to each S in S, $H(S)$ assigns a probability $P_{H(S)}(S)$, so that $\sum_{S \in S} P_{H(S)}(S) \leq 1$. The probability that the hierarchy selects no coalition from the collection of objecting coalitions S is $H_0(S) = 1 - \sum_{S \in S} P_{H(S)}(S)$.

Example 4.1: A hierarchy H on the two-player set $N = \{1,2\}$ can be described by:

S	$P_{H(S)}(S)$		
S =	1	2	12
{1}	·		
{2}		·	
{12}			·
{1,12}	·		·
{2,12}		·	·
{1,2}	·	·	
{1,2,12}	·	·	·

where the dots in each row represent non-negative numbers, and each row-sum is less than or equal to one.

A hierarchical structure for a game is a mapping H which associates a hierarchy H_x to each imputation x in the imputation space X. Hence a hierarchical structure represents the social choice mechanism which chooses between objecting coalitions at any proposal which arises in the course of a bargaining game, and abstracts all external factors which play a role in such a choice mechanism.

Various forms of measurability, compactness, or continuity requirements may be imposed on the hierarchical structures considered. An interesting (and often simplifying) requirement is that no hierarchy assign positive probability to both a coalition and any of its sub-coalitions when they occur in the same objecting collection. We shall not make any of these requirements, but merely note the possibility.

Several types of hierarchical structures have special intuitive appeal. One type is the uniform structure, which assigns

$$P_{H_x(S)}(S) = \frac{1}{|S|}$$

for all x in X and H_x in H. Another type is the excess structure, in which for each x in X, $H_x(S)$ assigns equal probability to all coalitions S in S which maximize $v(S) - x(S)$ (and zero probability to all other coalitions). A third type is the linear structure, in which a weight w_i is assigned to each player i in the player set N, and each $H_x(S)$ assigns equal probability to all coalitions in S of equal, maximal, total weight.

We shall restrict ourselves in the second half of this paper to considering games with the uniform hierarchical structure. However, it should be noted that solutions derived for a game with respect to a particular hierarchical structure may be subjected to a form of "sensitivity" analysis, in which study is made of the degree to which the hierarchical structure may be varied without changing the solution. Preliminary results, which shall not be given here, seem to indicate that many of our bargaining solutions are quite insensitive to variations from the uniform hierarchical structure - indicating that the solutions are stable over a wide range of social patterns.

5. Preferences

In the model of bargaining previously discussed, it was noted that when a player acts as a member of a coalition, he lacks determinate knowledge of what the ultimate result of this action will be. Instead, the most that the player can do is anticipate the probable result of his actions. Therefore, in trying to analyze the problem of what actions a

player will take, it is necessary to have some knowledge of his preferences over probabilistic outcomes.

The anticipated result of a player's actions can be considered as a probability distribution over the space of possible outcomes. Such a distribution will be called <u>finite</u> (respectively, <u>discrete</u>) if the distribution is concentrated on a finite (respectively, countable) number of these outcomes. In this paper we shall be concerned only with finite or discrete distributions. We shall assume that each player is concerned solely with the amount he personally receives in any imputation, and therefore the discrete probability distributions which represent probabilistic outcomes to a player may be described by sequences of the form $(x_1,p_1; x_2,p_2;\ldots)$, where the x_i are distinct real numbers, the p_i are positive and sum to one, and where the meaning of such a sequence is that the probability of the player receiving x_i in an outcome is p_i. Thus each discrete distribution over the imputation space induces a discrete distribution for each player (on his component of the imputations).

Given distributions $A = (x_1,p_1;\ldots)$ and $B = (y_1,q_1;\ldots)$, for any $0 \leq t \leq 1$ we define the distribution $tA + (1-t)B = (z_1,r_1;\ldots)$ by $r_i = t\cdot P_A(z_i) + (1-t)\cdot P_B(z_i)$ where $\{z_1,z_2,\ldots\} = \{x_1,x_2,\ldots\} \cup \{y_1,y_2,\ldots\}$. A <u>preference ordering</u> for a player is a (non-strict) total ordering '$\succsim$' of the space of all discrete distributions for which the following three axioms hold. Let A, B and C be any discrete distributions.

(P1) If $A \sim C$, then for any $0 \leq r \leq 1$ and B,
$(rA + (1-r)B) \sim (rC + (1-r)B)$.

(P2) If $A \succ C$, then for any $0 < r \leq 1$ and B,
$(rA + (1-r)B) \succ (rC + (1-r)B)$.

(P3) If x and y are real numbers and $x > y$, then $(x,1) \succ (y,1)$.

It is easily shown that if '$\succsim$' satisfies one further axiom,

(P4) If $A \succsim B \succsim C$, then there exists some $0 \leq r \leq 1$ such that $(rA + (1-r)C) \sim B$,

then the player's preferences are induced by a utility function over the space of discrete distributions. However, this last axiom seems to be of a different order than the first three, and we shall not require it.

In the second half of this paper, we shall be particularly concerned with one specific preference ordering. That is the <u>expected value</u> ordering '$\succsim_E$', for which $(x_1,p_1;\ldots) \succ_E (y_1,q_1;\ldots)$ if and only if $\sum x_i p_i > \sum y_i q_i$. It may be verified that the expected value ordering satisfies (P4), and that "expected value" may be viewed as a utility function.

Another ordering which seems of interest is the <u>maximin</u> ordering '$\succsim_M$', which is intended to describe the preferences of a player whose primary concern is to avoid any possibility of a low payoff. In this ordering, $(x_1,p_1;\ldots) \succ_M (y_1,q_1;\ldots)$ if and only if

$$\inf\{x_i : \text{either } x_i \neq y_j \text{ for all } j, \text{ or } x_i = y_j \text{ and } p_i > q_j\} >$$
$$\inf\{y_j : \text{either } y_j \neq x_i \text{ for all } i, \text{ or } y_j = x_i \text{ and } q_j > p_i\}.$$

It should be noted that this extremely conservative preference ordering does not satisfy (P4). However, it has the property that if A and B are two distributions concentrated on sets of real numbers which are

well-ordered by the natural order on the reals, and if $A \sim_M B$, then $A = B$. Thus the maximin ordering is a strict total ordering on the space of finite distributions. This ordering and the associated "maximax' ordering (replacing "inf" by "sup" in the preceding definition) are interesting because they bound, in a sense, all other preference orderings.

When a coalition takes action which results in a distribution over a set of imputations, the probabilistic outcome to each player in the coalition is the distribution induced over his component of the imputations. To facilitate consideration of situations in which a coalition of players is contemplating some action, we present some notation for coalitional preferences.

Assume that to each player i in a coalition S there is an associated preference ordering '$\succsim_i$'. Let A (respectively, B) be a discrete distribution over a set of imputations, and let A_i (respectively, B_i) be the distribution induced for i. Then S <u>prefers</u> A to B, written $A \succ_S B$, if $A_i \succ_i B_i$ for every player i in S. It may be noted that the relation '$\succsim_S$' satisfies (P1), (P2), and (P3), but is generally not a total order.

It is possible to define more than one system of coalitional preferences. The stated definition corresponds to "strong" coalitional preference, while "weak" coalitional preference arises when all players of a coalition non-strictly prefer (prefer or are indifferent to) one distribution over another, and when at least one of the players strictly prefers the first distribution. We mention this possibility because it may appear, on the surface, that Vickrey is dealing with weak coalitional preferences in his theory of self-policing solutions. However, in the

following sections we present a theory, based on strong coalitional preferences, which we feel fully preserves the spirit of Vickrey's work. This point will be further discussed in Section 11.

The possibility of analyzing solutions with regard to their sensitivity to changes in players' preferences exists here as well as with hierarchies, and seems to be an interesting avenue for research.

6. The Bargaining Games

There are several formal models of the type of bargaining process discussed in Section 3. We shall present two such models and mention a third. A primary difficulty in modelling the processes we are considering is encountered in the treatment of "stopping rules". In the real world, we certainly do not anticipate negotiations of infinite duration. Any of a number of factors may contribute to the termination of a bargaining process: an agreement may be reached which is satisfactory to all parties; proceedings may end in irreconcilable differences, in which case some of the players may receive "conflict" payoffs; there may be an external time limit which forces termination of the proceedings; or the participants may reach a point of exhaustion, at which point they concede any further objections they may have to the proposal being considered. In this section we shall define games with termination rules representative of the "time limit" and "exhaustion" stopping rules. In a later section we shall focus our attention on strategies in these games which have the desirable property of leading to agreement between all players on a final outcome. The bargaining games to be presented may be considered as formalizations of a binding arbitration procedure. If so considered, they

should be viewed as depicting the bargaining process which ensues after the players enter into a pact governing the structure within which they will work to resolve their conflicts.

We first define the "time limit" bargaining game. The definition is of a recursive nature. Let (N,v) be the characteristic function game under consideration, and H be the hierarchical structure associated with the imputation space X of this game. Let x be an imputation in X, and c be a real n-vector (which represents conflict payoffs to the players).

The bargaining game $L(N,v,H,c,x,0)$ is the "null" game in which each player i in N receives the payoff x_i.

For T a positive integer, the bargaining game $L(N,v,H,c,x,T)$ is played in the following manner. Each player i in N declares, for each coalition S containing i, a vector $y^{i,S} \in R^S$. All declarations by all players are made simultaneously. Let

$$S = \{S:\ y^{i,S} = y^{j,S} \equiv y^S \text{ for all } i,j \in S, \text{ and } y^S \text{ dom}_S\ x\}$$

be the collection of all coalitions whose players unanimously declare an allocation which dominates the current proposal. The hierarchy H_x is used to select a coalition from S. If the hierarchy fails to select a coalition (as will always be the case if S is empty), the game ends with final payoff vector x. On the other hand, if a coalition W in S is selected by the hierarchy, then the response-bargaining game $\overline{L}(N,v,H,c,W,y^W,T)$ ensues.

Take N, v, H, c, and T as previously defined. Let W be a coalition in N, and $y^W \in R^W(v(W))$ be a vector (in R^W, with

$y^W(W) \leq v(W)$). The response-bargaining game $\bar{L}(N,v,H,c,W,y^W,T)$ is played in the following manner. Each player i in $N-W$ declares a vector $z^{i,N-W} \in R^{N-W}$ for which $y^W + z^{i,N-W} \in X$. The declarations by all players are made simultaneously. If for some pair of players i and j in $N-W$, $z^{i,N-W} \neq z^{j,N-W}$, then the game ends with final payoffs y^W_k for all players k in W and payoffs c_k for all k in $N-W$. On the other hand, if all $z^{i,N-W} \equiv z^{N-W}$, then the bargaining game $L(N,v,H,c,y^W + z^{N-W},T-1)$ ensues.

We next define the bargaining game which terminates upon "exhaustion" of the players. The definition is again recursive. Let N, v, H, and c be as previously defined. Let δ be a real number such that $0 < \delta \leq 1$, and let x be an imputation in X. The bargaining game $E(N,v,H,c,x,\delta)$ is played similarly to the game $L(N,v,H,c,x,1)$, with the following differences. With probability δ a chance event occurs before the players make their declarations, and the game ends with final payoff vector x. Otherwise, if the players in coalition W unanimously declare y^W, and W is selected by the hierarchy H_x, then the response-bargaining game $\bar{E}(N,v,H,c,W,y^W,\delta)$ ensues. This response-bargaining game is similar to the game $\bar{L}(N,v,H,c,W,y^W,1)$ except that if the players in $N-W$ unanimously declare z^{N-W}, then the bargaining game $E(N,v,H,c,y^W + z^{N-W},\delta)$ ensues. Since $\delta > 0$ implies that the bargaining game has probability one of terminating after a finite number of moves, we arbitrarily assign payoffs of zero to all players in the (negligible) case of infinite play.

The stopping probability δ represents the chance that the players will be so exhausted after any stage of the game as to forego their possible objections to the current proposal. It may be noted that δ can be incorporated in a natural manner into the hierarchical structure

H to yield a new structure H' which embodies the stopping rule. In this way the game $E(N,v,H,c,x,\delta)$ represents our intuitive idea of the game $L(N,v,H',c,x,\infty)$. Similarly, if T is a random variable with geometric distribution, so that

$$\text{Prob}(T = k+\ell \,|\, T \geq \ell) = \text{Prob}(T = k) = \delta(1-\delta)^k,$$

then the game $E(N,v,H,c,x,\delta)$ represents our intuitive idea of the game $L(N,v,H,c,x,T)$*.

In both types of bargaining games, the information structure is such that each player has full knowledge of the parameters of the game, and remembers the full history of the game as it progresses, including all players' declarations at each previous stage. In the next section we shall briefly consider all strategies available to the players. Having done so, we will then restrict our considerations to strategies in which the players use only information dependent on the parameters N, v, H, c, and the proposal x under consideration. The reasons for this restriction are primarily those of convenience, and we shall make the restriction in the context of the strategies rather than the games themselves. We simply note here that such a restriction can, if desired, be incorporated into the information structure of the game.

A comment is in order concerning the structure of the response-bargaining games. These games are not intended to provide accurate models of a real-world negotiation process. Rather, each response-bargaining

---

*Indeed, we could have taken the alternative approach of defining a game in terms of a random stopping time T. Both the "time limit" and "exhaustion" games would be special cases (depending on the distribution of T) of such a game.

game represents a "black box" used to stand for the bargaining which takes place after a valid objection is made to a proposal and before a new proposal is accepted for consideration. Ideally, a response-bargaining game should be a detailed negotiation model depicting what we (personally) believe to be the most complex aspect of the bargaining process. Unfortunately, the construction of an accurate model of such situations is very difficult, due at least in part to the lack of empirical data regarding behavior in such situations. Hence, we merely affirm our stance that the "black box" approach is a feasible one for preliminary investigation. Concerning the conflict payoffs in each bargaining game, it should be noted that they will be used to "force" behavior along certain cooperative lines, and are not intended to ever be attained in the play of a game. This will become evident in the next sections.

It was mentioned earlier that a third type of bargaining game could be defined. This type would treat termination difficulties directly, by not imposing a stopping rule but rather assigning payoffs to every infinite-play possibility. We consider this approach to be not very fruitful due to the difficulty in deciding upon meaningful infinite-play payoffs, and also because real-world bargaining processes cannot, generally, continue forever.

Finally, of the first two types of bargaining games discussed, we shall treat only the type based on the "exhaustion" stopping rule (that is, the games E and $\bar{E}$). A primary reason for this specialization will be seen in Section 9, when we derive certain simplifying results which do not hold for the "time limit" games. In view of our eventual intention, the following sections will be presented in terms of these "exhaustion" games.

7. Strategies

As a bargaining game is played, a sequence of declarations (by the players) and chance selections (by the hierarchical structure) is generated. Any play of the game up to a particular moment in time (at which the play of a bargaining or response-bargaining subgame commences) may be described by the initial parameters of the game and such a sequence. We call such a description a history of the bargaining game.

A bargaining situation $B(N,v,H,c_0,\delta_0)$ is the collection of all bargaining and response-bargaining games which are defined in terms of N, v, and H, with conflict payoff vector c satisfying $(c_0)_i \leq c_i < v(i)$ for all i in N, and with stopping probability δ satisfying $0 < \delta \leq \delta_0$. Thus a bargaining situation consists of all games in a "neighborhood" of the "game" with conflict payoffs $c_i = v(i)$ and with stopping probability zero.

We wish to consider systems of behavior for a player which describe how he will act in any game in a given bargaining situation. Therefore, define a global pure strategy for a player i in a bargaining situation as a function which maps each game in the situation, and each possible associated history of that game, into an action by i of the type called for in the resulting subgame (depending on this subgame, such an action is a collection of declarations $\{y^{i,S}\}_{S \ni i}$, a declaration $y^{i,N-W}$, or a "null" action if the subgame is a response-bargaining game in which i has no move). Note that this definition requires a full plan of action for every game in the given bargaining situation, and therefore a global pure strategy is simply a collection of pure strategies of the usual type.

We shall actually work with the more general concept of a <u>global behavioral strategy</u>. Such a strategy for a player in a bargaining situation maps each bargaining game and associated history into a finite (probabilistic) sample space and a function from the sample space into the set of actions which the player may properly take. Each sample space and associated function correspond to a "random experiment" performed by the player to select his action at a particular stage of the game, and for later simplicity we assume that all such experiments are independently repeatable. It should be noted that every global pure strategy corresponds in an obvious way to a global behavioral strategy.

We require that the sample spaces be finite for reasons of both theoretical and notational convenience. As discussed by Aumann [1], to allow an overly-wide class of randomizing actions at each stage of a game is to risk measure-theoretic difficulties in the outcome space. Aumann's avoidance of these difficulties involves defining behavioral strategies as a special class of mixed strategies in which only a single randomization takes place, at the beginning of play. Since we wish to work with the notationally-simpler idea of randomization at each stage, we are forced to limit ourselves to a restricted form of randomization.

A <u>strategy n-tuple</u> $\sigma = (\sigma_1,\ldots,\sigma_n)$ for a bargaining situation $B(N,v,H,c_0,\delta_0)$ is a collection of global behavioral strategies, with σ_i being the strategy of player i. Any such collection of strategies associates to every game in the bargaining situation a discrete distribution over the outcome space of the game. For example, assume $B(N,v,H,c_0,\delta_0)$ is a bargaining situation, $E(N,v,H,c,x,\delta)$ is a specific game in the situation, and σ is a strategy n-tuple for $B(N,v,H,c_0,\delta_0)$.

Then after a move in E by all the players, and chance moves with respect to δ and H, either the game ends with a specific payoff vector or a response-bargaining subgame ensues. The crucial point is that there is only a finite number of alternatives that can arise after the first move of the game, when the players follow their strategies in σ. Similarly, at each following stage of the game only a finite number of alternatives can occur, because each player's strategy at that stage is a randomization over a finite number of actions. Therefore there is only a countable number of histories that may be generated as the players use their strategies in σ, and only a countable number of distinct payoff vectors may finally result. The game is of finite length with probability one, and therefore, by associating to each possible payoff vector the sum of the probabilities of all sequences of play which result in that vector, a discrete distribution results as claimed. (It should be noted that the outcome space referred to here includes vectors of conflict payoffs; thus the outcome space is $X \cup \{y \in R^n$: for some coalition S, $y^S \leq v(S)$ and $y^{N-S} = c^{N-S}\}$).

A class of global behavioral strategies to which we will give particular attention is the class of reactive strategies. A strategy for a player in a bargaining situation $B(N,v,H,c_0,\delta_0)$ is a <u>reactive strategy</u> if it specifies the same action for the player in all bargaining and response-bargaining games which differ only in their conflict payoffs, stopping probabilities, and histories leading up to the games. Thus, when playing a reactive strategy in a given situation, a player "reacts" only to the vector x in a game $E(N,v,H,c,x,\delta)$, and only to the vector y^W in a game $\bar{E}(N,v,H,c,W,y^W,\delta)$ - with no regard for the circumstances

which led to the play of the game, or for the values of c and δ. More specifically, a reactive strategy in a situation $B(N,v,H,c_0,\delta_0)$ associates to each imputation x in X a finite sample space and a function from that sample space into the set of actions which the player may properly take in a bargaining game commencing with the proposal x, and associates to each pair (W,y^W), where W is a coalition and $y^W \in R^W(v(W))$, a finite sample space and function from that sample space into the set of actions which the player may take in any response-bargaining game commencing with an objection y^W made by W.

The idea of reactive strategies in multi-stage bargaining games is originally due to Harsanyi [8]. Although restricting the players of a game to using reactive strategies eliminates the possibility of players making threats contingent upon the actions of others, we shall discuss in a later section the manner in which this restriction does not affect the stability of certain strategy n-tuples. It should be noted in contrast that restriction to consideration only of reactive strategies is extremely severe with regard to the "time limit" bargaining games of the preceding section. It is for this reason that we have excluded such games from our subsequent work.

We earlier referred to the sample spaces we use as "independently repeatable experiments". The reason for this reference may now be apparent. Since a position may occur more than once in the play of a game, a player may have to draw upon the same sample space several times in his randomizations. We wish to allow this, but require that the randomizations in different stages of the game be independent.

There is also a specific reason for defining strategic randomization in terms of sample spaces as "experiments", rather than merely working in

terms of finite distributions over the spaces of available actions. This is to emphasize the possibility that several players may adopt strategies which involve use of a common sample space for their randomizations, thus introducing a correlation into their actions. We allow only a restricted form of such correlation. That is, the actions of any set of players may be correlated only with respect to declarations which they make simultaneously, and furthermore this correlation may only occur between declarations related to a specific coalition to which all these players belong. As an example, assume players 1 and 2 are involved in the play of a bargaining game. Then the declarations $y^{1,\{1,2\}}$ and $y^{2,\{1,2\}}$ may be correlated. The declarations $y^{1,\{1,2,3\}}$ and $y^{2,\{1,2,3\}}$ may also be correlated (possibly with $y^{3,\{1,2,3\}}$ as well), but the declaration $y^{1,\{1,2,3\}}$ must be independent of the declarations $y^{1,\{1,2\}}$ and $y^{2,\{1,2\}}$.

8. Equilibria

From the collection of all strategy n-tuples in a game, it is desirable to be able to single out those which exhibit some form of stability. A general approach to this was first suggested by Nash [15]. Roughly speaking, Nash defined an n-tuple of strategies σ to be in equilibrium if no single player could gain (over his payoff when σ was played) by unilaterally changing his strategy while the remaining players played their strategies in σ. In studying solutions for cooperative games, a common extension of this approach is more generally to require that no coalition of players can improve <u>all</u> of their payoffs by changing their strategies from σ while the remaining players play their strategies in

σ. Recently, a further extension has been to consider strategies for a collection of interrelated games, and ask that these strategies be in equilibrium for all games in the collection. We shall use this approach to analyze bargaining situations, making these ideas more precise in the following paragraphs.

Let $\sigma = (\sigma_1,\ldots,\sigma_n)$ be an n-tuple of (global behavioral) strategies for the situation $B(N,v,H,c_0,\delta_0)$. As discussed in the previous section, to any particular game G in B, σ associates a discrete probability distribution $\psi(\sigma,G)$ over the imputation space of the game, and this in turn induces a collection of distributions $\{\psi_i(\sigma,G)\}_{i\in N}$, one for each player over his outcome space. Let σ_i' be a strategy for player i, and let $\sigma' = (\sigma_1,\ldots,\sigma_i',\ldots,\sigma_n)$. σ_i' is a <u>better response</u> than σ_i in σ if for some game G in B, $\psi_i(\sigma',G) \succ_i \psi_i(\sigma,G)$, and for every G in B, $\psi_i(\sigma',G) \succsim_i \psi_i(\sigma,G)$. Thus a better response for a player with respect to an n-tuple of strategies for a bargaining situation is a strategy change which benefits him in some game of the situation, and which hurts him in no game of the situation. An n-tuple of strategies σ is an <u>individual equilibrium point</u> for a bargaining situation if no player has a better response than his strategy in σ.

It should be noted that this is a relatively weak condition. It is not difficult to show that many bargaining situations have individual equilibrium points in which no coalition takes effective unanimous action at any stage of any game. We therefore consider individual equilibrium points which involve certain forms of coalitional cooperation.

Let S be a fixed coalition, and σ an individual equilibrium point for the bargaining situation $B(N,v,H,c_0,\delta_0)$. A strategy n-tuple

$\tau = (\tau_1,\ldots,\tau_n)$ is <u>related to</u> σ by S if the following conditions are satisfied. First, for every player i not in S, $\sigma_i = \tau_i$. Second, for every player i in S, τ_i is arbitrary with respect to all response-bargaining subgames arising in B, but in any bargaining subgame σ_i and τ_i differ only in the specification of declarations of the form $y^{i,S}$. That is, in the terminology of the previous section, the sample spaces and functions $\{\tau_i\}_{i\in S}$ associated with each particular bargaining subgame are such that the induced distributions and correlations between declarations made by the players, with respect to coalitions other than S, remain unchanged. An individual equilibrium point σ is a <u>coalitional equilibrium point</u> for the bargaining situation $B(N,v,H,c_0,\delta_0)$ if there is no coalition S and n-tuple τ related to σ by S for which $\psi_S(\tau,G) \succ_S \psi_S(\sigma,G)$ for some G in B, and $\psi_S(\tau,G) \succsim_S \psi_S(\sigma,G)$ for all G in B.

This coalitional condition, which examines the effect of "coalitional strategic deviation" in which the players of a coalition only change those of their actions which are naturally related to that coalition, is similar to conditions used by Harsanyi [7], in which the coalitions, as "syndicates", are considered to be "large players" with overlapping interests. The requirement that a global strategy be in equilibrium with respect to all games in a situation is related to Selten's [19] definition of a "perfect" equilibrium.

One further condition remains to be discussed. A primary question of interest at any position (game and associated history) of a bargaining situation is which coalitions will take unanimous action. It seems reasonable to restrict investigation to those equilibria in which, at

every position where several coalitions take effective action, at least one of these coalitions is motivated to take that action. The condition we give here is quite unsophisticated, but suffices for the analysis in this paper. A more involved condition, which reflects "deeper" motivations, would be of interest.

Let σ be a reactive coalitional equilibrium point, let G be a bargaining game (not a response-bargaining game) in the associated bargaining situation, and let x be the proposal associated with G. Further, let S be the collection of all coalitions which take effective unanimous action (raise a valid, enforceable objection) against x. Then σ is motivated at x if for at least one coalition S in S, that coalition prefers its result from its objection at x to the certain outcome of x. If σ is motivated at every x in X, σ is a motivated reactive coalitional equilibrium point.

We restrict ourselves to the consideration of motivated equilibria in order to avoid certain possibilities which are best described heuristically. Assume A and B are disjoint coalitions in a given game, and that a strategy n-tuple σ involves such behavior when an imputation x is proposed, that the players of A and the players of B each make effective objections to x while no other coalitions act. Further assume that the players in A prefer their shares in x to the outcome of their objection and prefer the outcome of their objection to the outcome of the objection by B, and assume that the analogous statement holds for the players in B. Then σ may well be in coalitional equilibrium (since either coalition loses by ceasing its objection unilaterally), but all players in A and B gain by ceasing both objections. It is to avoid

such equilibria that we impose the above "motivational" condition. Further discussion will be given to this topic in Section 17.

It should be noted briefly that, in discussing equilibria with respect to correlated strategies, such strategies are not to be considered binding after randomizations. Furthermore, any players using the same sample space for randomization are assumed to have identical information concerning all outcomes in the space. This is to avoid consideration of complications of the type discussed by Aumann [2].

9. Preliminary Simplifications

If we restrict our attention to n-tuples of reactive strategies which form motivated coalitional equilibrium points, a number of notational simplifications become possible. In this section we derive several results which justify these simplifications.

The first result allows us to restrict our search for equilibria to only those situations in which all coalitions always cooperate fully in all response-bargaining games.

Theorem 9.1. If σ is a reactive equilibrium point, then in every response-bargaining subgame in which the players of a coalition S are to move, with probability one all of these players make the same declaration.

Proof. We first show that for any player i and imputation x with $x_i \geq c_i$ (c_i is the conflict payoff to player i), player i prefers the result of following σ, in the bargaining game beginning at x, to the outcome of receiving c_i with certainty. This follows immediately from the observation that i can refuse to cooperate in objections to

x and also refuse to cooperate in all responses arising from objections to x, and in this manner he can assure himself of at least c_i in all eventualities.

Assume that the players of S fail to cooperate in some response in σ. Then in their non-cooperative response, each player i receives c_i. However, by dividing the amount available in response equally, each player i in S receives at least $v(i)$ immediately, and expects no less than c_i eventually (by the preceding paragraph). Thus the cooperative response is a better response for all players of S than their strategies in σ, and σ cannot be in equilibrium.

From the proof of the theorem, we immediately have a guarantee of individual rationality at all imputations to which no objections are made.

Corollary 9.2. If the hierarchical structure gives positive probability to the recognition of some coalition whenever objections are raised, then every imputation x, to which no objection is made, is individually rational (satisfies $x_i \geq v(i)$, for all players i).

Our next result shows that each coalition, if it has positive probability of raising an effective objection to a proposal x, might as well object to x with probability one.

Theorem 9.3. If σ is a reactive equilibrium point, x is an imputation, and S is a coalition of players who have in σ a positive probability of all making the same objection (as players of S) against x, then there exists an equilibrium point in which the players of S correlate their play so that they are unanimous in their objection with probability one.

Proof. Assume the players of S cooperate with probability p in raising an objection to x. Let q be the probability that they are selected by the hierarchy at x when they make a unanimous objection. Since σ is in equilibrium, we must have the event

A: with probability $p(1-q)$, S objects and some other coalition is given the floor;

with probability pq, S objects and is given the floor;

with probability $(1-p)$, S doesn't raise an objection to x,

preferred (although not necessarily strictly preferred) by each player of S to the event

B: with certainty, S doesn't raise an objection to x.

But then, by assumption (P2) of Section 5, all players in S prefer a certain ($p = 1$) objection by S to the event A. Changing the strategies of the players of S in σ to conform in this manner with the statement of the theorem yields the required new equilibrium point.

These results permit us to consider "coalitional", rather than individual, strategies in our formal definition of a bargaining solution. They also justify establishing the formal model without reference to conflict payoffs, as long as we require that coalitional behavior conform by definition to the results above.

There is one result we would like to give, but have been unable to derive in a general setting. This would be a theorem similar to Lemma 6 of Harsanyi [8], stating that a player has a better reply to an n-tuple of reactive strategies only if he has a better reactive reply. Such a result would provide some justification for restricting our considerations

exclusively to reactive strategies. However, although the result can be shown to hold in all specific cases with which we have worked, a general proof has not yet been found.

10. Bargaining Solutions

In view of the preceding sections, we are now prepared to formally define a theory of bargaining solutions. The definitions will be so given as to characterize motivated coalitional equilibria in reactive strategies.

We take as given an n-person game (N,v), a hierarchical structure H on the imputation space X, and a system of individual preferences $\{\succsim_i\}_{i\in N}$ from which coalitional preferences may be derived. For any coalition S, define

$$R^S = \{x \in R^n: \ x_i = 0 \quad \text{for all} \quad i \notin S\}$$

and

$$R^S(\alpha) = \{x \in R^S: \ x(S) \leq \alpha\}.$$

Recall that $F(A)$, and $D(A)$, are repsectively the set of all finite probability distributions, and discrete probability distributions, on a set A. If ρ is any such distribution then, for any $a \in A$, $P_\rho(a)$ is the probability assigned to a by ρ, and

$$\overline{\rho} = \{a \in A: \ P_\rho(a) > 0\}.$$

A <u>coalitional strategy</u> σ_S for a coalition S is a pair of functions (σ_S^1,σ_S^2), such that

$$\sigma_S^1: \; X \to F(R^S) \quad \text{and} \quad \sigma_S^2: \; R^{N-S}(v(N-S)) \to F(R^S).$$

For each x in X, $\sigma_S^1(x)$ satisfies eit'

(a) $y \in \overline{\sigma_S^1(x)}$ implies $y \; \text{dom}_S \; x$,

or

(b) $\overline{\sigma_S^1(x)} = \{x^S\}$.

For each x in $R^{N-S}(v(N-S))$, $\sigma_S^2(x)$ satisfies

$$y \in \overline{\sigma_S^2(x)} \quad \text{implies} \quad (x+y) \in X.$$

The strategy σ_S^1 is the "objection strategy" of the coalition S, and σ_S^2 is the "response strategy". The conditions on $\sigma_S^1(x)$ are that either (a) S raises a dominating objection to x, or (b) S does not object to x. The condition on $\sigma_S^2(x)$ is simply that, after an objection and response, the resulting proposal must be an imputation.

Let $\sigma = \{\sigma_S\}_{S \subset N}$ be a collection of coalitional strategies. For each x in X, define

$$\eta(x) = \{S: \; \overline{\sigma_S^1(x)} \neq \{x^S\}\}.$$

Thus $\eta(x)$ is the collection of coalitions which, in σ, raise objections to x. σ induces a <u>transition map</u>

$$\theta_\sigma: \; X \to F(X)$$

defined by

$$\overline{\theta_\sigma(x)} = \bigcup_{S\in\eta(x)} \{w:\ w = y+z,\ \text{where}\ y \in \overline{\sigma^1_S(x)}\ \text{and}\ z \in \overline{\sigma^2_{N-S}(y)}\}$$

$$\cup \{w:\ w = x,\ \text{and either}\ \eta(x) = \emptyset\ \text{or}\ (H_x)_0(\eta(x)) > 0\},$$

where for each $w \in \overline{\theta_\sigma(x)}$, either

$$w \neq x,\ \text{and}\ P_{\theta_\sigma(x)}(w) = \sum_{S\in\eta(x)} \sum_{\substack{y+z=w \\ y\in\overline{\sigma^1_S(x)} \\ z\in\overline{\sigma^2_{N-S}(y)}}} P_{\sigma^1_S(x)}(y)\cdot P_{\sigma^2_{N-S}(y)}(z)\cdot P_{H_x(\eta(x))}(S)$$

or

$$w = x,\ \text{and}\ P_{\theta_\sigma(x)}(x) = 1 - \sum_{\substack{w\in\overline{\theta_\sigma(x)} \\ w\neq x}} P_{\theta_\sigma(x)}(w)\ .$$

The set $\overline{\theta_\sigma(x)}$ is the collection of all imputations which might arise in the stage of the bargaining game immediately following the stage in which x is proposed. The last set in the definition serves only to include cases in which the imputation x results from itself (that is, the game ends).

For any $0 < \delta \leq 1$, σ also induces a valuation map

$$\psi_{\sigma,\delta}:\ X \to \mathcal{D}(X)$$

defined by

$$\overline{\psi_{\sigma,\delta}(x)} = \bigcup_{k=0}^{\infty} \overline{(\theta_\sigma)^k(x)},$$

where for each $y \in \overline{\psi_{\sigma,\delta}(x)}$,

$$P_{\psi_{\sigma,\delta}(x)}(y) = \sum_{k=0}^{\infty} P_{(\theta_\sigma)^k(x)}(y) \cdot \delta(1-\delta)^k.$$

In this definition, $(\theta_\sigma)^k(x)$ is the iterated distribution over X which arises k stages after the proposal x is made. Thus $\psi_{\sigma,\delta}$ assigns to each imputation the distribution which arises after the play of a bargaining game, with stopping probability δ, in which all players follow the strategies in σ. Similarly, for each $0 < \delta \leq 1$ and each $x \in X$, σ induces a response-valuation map

$$\phi_{\sigma,\delta,x}: \eta(x) \to D(X),$$

where $\phi_{\sigma,\delta,x}(S)$ is the distribution which arises after the play of a bargaining game, beginning at x with coalition S having the floor (that is, having just been selected by the hierarchy H_x).

σ is a bargaining solution to the game (N,v), with respect to the given hierarchical structure and system of preferences, if the following conditions are satisfied.

(1) There exists $0 < \delta_0 \leq 1$ such that for every $0 < \delta < \delta_0$, $i \in N$, $S \ni i$, $y \in X^{N-S}(v(N-S))$, and $z \in \overline{\sigma_S^2(y)}$,

$$\psi_{\sigma,\delta}(y+z) \succsim_i (1,v(i)).$$

(Recall that $(1,v(i))$ signifies the event "receiving $v(i)$" with probability 1.) This condition is derived from Theorem 9.1, and guarantees that σ is in equilibrium with respect to response strategies.

(2) For any $i \in N$, let τ be a collection of coalitional strategies for which $\tau_S^1 = \sigma_S^1$ for all $S \neq i$, and $\tau_S^2 = \sigma_S^2$ for all S. Further assume that for each $x \in X$ and S containing but not equal to i, $P_{\tau_S^1(x)}(y) \leq P_{\sigma_S^1(x)}(y)$ for all $y \neq x^S$. Then there is no $0 < \delta_0 \leq 1$ for which, for all $x \in X$ and $0 < \delta < \delta_0$,

$$\psi_{\tau,\delta}(x) \succsim_i \psi_{\sigma,\delta}(x),$$

and for which, for some $x \in X$ and for every $0 < \delta_0' < \delta_0$, there is a $0 < \delta < \delta_0'$ such that

$$\psi_{\tau,\delta}(x) \succ_i \psi_{\sigma,\delta}(x).$$

This condition is merely a restatement of the requirement that σ be in equilibrium, with respect to objection strategies, for each individual $i \in N$ and all stopping probabilities "sufficiently close" to zero.

(3) For any coalition S, let τ be a collection of coalitional strategies for which $\tau_W = \sigma_W$ for all $W \neq S$. Then there is no $0 < \delta_0 \leq 1$ for which, for all $x \in X$ and $0 < \delta < \delta_0$,

$$\psi_{\tau,\delta}(x) \succsim_S \psi_{\sigma,\delta}(x),$$

and for which, for some $x \in X$ and for every $0 < \delta_0' < \delta_0$, there is a $0 < \delta < \delta_0'$ such that

$$\psi_{\tau,\delta}(x) \succ_S \psi_{\sigma,\delta}(x).$$

This requires coalitional equilibrium in all bargaining subgames with a sufficiently small stopping probability. Further, we require the analogous condition in all such response-bargaining subgames.

(4) There is a $0 < \delta_0 \leq 1$ for which, for every $x \in X$ with $\eta(x) \neq \emptyset$ and every $0 < \delta < \delta_0$, there exists a coalition $S \in \eta(x)$ with

$$\phi_{\sigma,\delta,x}(S) \succ_S (1,x^S).$$

This condition restates the requirement that some coalition which objects to x be motivated in its objection.

There is a fifth condition, not yet discussed, which we shall use throughout the second half of this paper. It is the requirement that σ be <u>bounded</u>.

(5) There is a positive integer B such that, for all $x \in X$, every sequence of imputations $\{y_k\}_{k=0}^{B}$ with $y_0 = x$, which satisfies

$$y_k \neq y_{k-1} \quad \text{and} \quad y_k \in \overline{\theta_\sigma(y_{k-1})} \quad \text{for all} \quad k = 1,2,\ldots,B,$$

also satisfies

$$y_B = y_{B-1}.$$

Restated, this simply requires that, in σ, every bargaining subgame must end, after at most B stages, with a proposal to which no coalition objects. An advantage of working with this reasonable-sounding requirement for bargaining solutions will be seen after the next definition.

Consider any (bounded) bargaining solution σ to a game. Associated with σ is the set of all imputations to which no coalition objects. Formally, the stationary set S of σ is defined by

$$S = \{x \in X: \overline{\theta_\sigma(x)} = \{x\}\}.$$

Each element of S is a stationary imputation of σ.

Two final comments are in order. First, the requirement that σ be bounded implies the existence of at least one stationary imputation. This is because, regardless of the initial proposal, after at most B stages of the bargaining game all objections must end. And second, a most important observation is that the definition of "stationary set" is what all our work to this point has been directed towards. In the second half of this paper, we shall show that stationary sets have a close relationship to von Neumann stable sets in many games.

11. Summary

In the preceding pages, we have presented a solution theory for n-person cooperative games. The basic approach was to embed a characteristic function game in a formalized bargaining context. In the process of analyzing the bargaining game several conditions, corresponding to our intuitive notion of which collections of strategies seem to be stable as

"standards of behavior" for the players, were used to select certain strategy collections as solutions to the game. The set of imputations left stationary by any such collection of strategies was proposed as an analogue to the classical von Neumann-Morgenstern stable set.

Several comments should be made concerning our bargaining model. A major problem already discussed is to find a convincing representation of response-bargaining games. An appropriate approach may involve the playing of a rather complex subgame each time a response is called for. In such an approach, it may be hoped that equilibria for the full bargaining game will be a composition of solutions for the bargaining subgames with solutions for each response-bargaining subgame, in which case the two types of games could be separately treated.

Minor variation in our theory may be derived by requiring only weak conditional preferences where we have required strong preferences. However, we would expect such a change to have little effect on our results, due to the use we have made of stopping probabilities in a limiting sense. The reason for such use was the appeal of Vickrey's work in which a valid heresy does not require that all heretical players show a strict eventual gain, but rather that all players gain, at least momentarily, when they take heretical action. Our formulation of the bargaining process gives a slight payoff at each such moment, reflecting the intermediate gain of the players.

Applications of the principles of this bargaining theory to other forms of games may be made. As an example, we apply our work to a non-side-payment game in a later section. A most promising line seems to be direct application of our techniques to cooperative games in normal form.

A final note may be made of the potential use of a theory such as this in the field of multi-national relations. In an age of "detente", the principle technique for maintaining international equilibrium is the establishment of "credible" responses to every possible deviation (objection) of opposing forces from the equilibrium position. Naturally, the work we present here is far from such applications. But the example gives some indication of the importance we feel is attached to game-theoretic models which adequately deal with response mechanisms.

CHAPTER 2

EXAMPLES OF BARGAINING SOLUTIONS AND STATIONARY SETS

12. Introduction

The first chapter presented a general solution theory for n-person games. We now illustrate this theory by applying it to several well-known classes of games. In treating these various classes, we sharply restrict the generality of the preceding sections by limiting consideration to games with a particular type of hierarchical structure and a particular system of preferences. In previous terms, we shall work only with uniform hierarchical structures and expected value preferences. For the sake of completeness, we discuss here the specific bargaining model with which we work, and consider the simplifications that our specification permits.

A bargaining solution for a game consists of two parts. The first part is a list, for each imputation, of all coalitions which object to that imputation and what objections (allocations among their players which dominate the given imputation) they make. The second part is a description, for each objection, of the response (allocation of remaining resources which, with the objection, forms a new imputation) made by the coalition of non-objecting players. In this way, objection and response strategies for each coalition are specified, and this collection of strategies is a bargaining solution if it is bounded (there is a uniform upper bound on the number of objection-and-response stages that can occur before an imputation results to which no coalition objects), and if it is in equilibrium (no coalition of players can gain by changing its strategy while the other coalitions hold theirs unchanged).

A partial description of the bargaining game will extend the vocabulary with which we work. At any given imputation, all coalitions declare their objections, if any, to that imputation as a final division of resources. If no objections are made, the game ends with that division. Otherwise, a coalition is chosen (equiprobably) at random from the objecting coalitions. The objection by this coalition, with a response by the complementary coalition, yields a new imputation, from which the game continues in the above manner.

At each stage of the game, there is a small probability $\delta > 0$ that external pressures will force the end of the game. Although we shall rarely refer to a specific stopping probability in the following sections, we will implicitly refer to it by discussing "momentary" outcomes (those which occur as a result of this externally-induced end of the game) in conjunction with "essential" outcomes (those which result when no coalitions object to the proposed imputation at some stage). We shall assume that all players are driven solely by a desire to maximize their expected return in the play of the bargaining game. This assumption is implicit in discussions of preferences of coalitions of players between alternative actions. Since our objective is to solve a "game-situation" for all sufficiently small δ, it should be noted that, under this preference structure, players choose between alternatives primarily in terms of their essential expectations in these alternatives. Only in the case of equal essential expectations do they secondarily refer to their momentary expectations.

With any bargaining solution is associated a stationary set (the set of all imputations to which no coalition objects). The assumption that

each bargaining solution is bounded implies that each stationary set is nonempty. Furthermore, since the core of a game consists of those imputations to which no objection can be made, the core is a subset of every stationary set. Also, the results of a preceding section show that every imputation in a stationary set is individually rational.

Section 13 deals with three-person games. We characterize all stationary sets which arise with respect to uniform hierarchies and expected value preferences, and find that these stationary sets are some of the von Neumann-Morgenstern stable sets which are self-policing. The bargaining solutions exhibited extend over all possible allocations (both individually-rational and non-individually-rational imputations).

In Section 14 we treat games with vNM-stable cores. We show that the core is a stationary set for three well-known classes of such games, and relate the construction of response strategies to a recent characterization of all such games. Section 15 considers two "pathological" games: a five-person game discovered by Lucas, and a seven-person non-side-payment game with no stable set, discovered by Stearns. It is shown that the core is a stationary set for both of these games.

Section 16 contains both positive and negative results. Although the symmetric stable set of any constant-sum symmetric majority game is shown to be stationary set of the game, it is also shown that the symmetric stable set of the four-person simple game, in which all three-player coalitions win, is not a stationary set. Section 17 discusses a negative potentiality of our theory, and briefly presents an extension of our approach which properly handles such a difficulty. Section 18 summarizes the results of these sections, and concludes with several comments concerning future directions of this work.

13. Three-Person Games

A natural approach to the investigation of bargaining solutions for three-person games is to study those sets of imputations which can be stationary sets for solutions. This approach is quite fruitful, for we will be able to characterize all stationary sets of each such game as self-policing stable sets of the game. We derive this result in some detail, omitting the most tedious arguments.

For notational convenience, we shall work with 3-person games in (0,1)-normal form. Thus a game is defined on the player set $N = \{1,2,3\}$ by

$$\begin{aligned} v(i) &= 0 && \text{for all} \quad i \in N, \\ 0 \leq v(i,j) &\leq 1 && \text{for all} \quad i,j \in N, \\ v(N) &= 1, \end{aligned}$$

where the indices i, j, and k will be used throughout to denote distinct players.

An important simplification may be made when analyzing 3-person games. In the simplex of individually-rational imputations, only 2-player coalitions can raise objections, and so only 1-player response strategies must be specified. Furthermore, 1-player responses are trivial (that is, if $\{i,j\}$ objects with an allocation (x_i, x_j), then k must take $1 - x_i - x_j$ in response). Therefore, discussion of 1-player responses will be omitted from many of the arguments of this section.

Let σ be a bargaining solution to a given 3-person game, and let S be the stationary set of σ.

Lemma 13.1. S is internally stable.

Proof. Assume to the contrary that $x,y \in S$ and $x \text{ dom}_T y$ for some coalition $T = \{i,j\}$. Then T gains by making the objection (x_i,x_j) at y, and σ cannot be in equilibrium. On the other hand, if $T = \{i\}$, then $0 = v(i) \geq x_i > y_i$, and this contradicts Corollary 9.2, that non-individually-rational imputations cannot be stationary.

Lemma 13.2. No imputation z can dominate all imputations in S.

Proof. Since the bargaining solution σ is bounded, the continuity of "expected value" preferences implies that no 2-player coalition can be motivated to raise an objection to z. Thus z is stationary and dominates itself. This cannot be.

With the aid of these lemmas, we first consider the 3-person constant-sum game, in which $v(ij) = 1$ for all i and j. Notice that every imputation is dominated by some other imputation, and Lemma 13.2 therefore implies that no stationary set can consist of just one imputation. Also, Lemma 13.1 and the dominance pattern of this game imply that no two imputations x and y can both be stationary unless for some player i, $x_i = y_i$. Only three cases remain to be considered:

(1) The stationary set S consists solely of imputations x for which $x_i = c$, for a specific player i and constant $0 \leq c < 1$. Without loss of generality, assume $i = 3$.

By Lemma 13.2, S is dense in $\{x: x_1 \geq 0, x_2 \geq 0, x_3 = c\}$. (Otherwise, an interval $\{x: a \leq x_1 \leq b, x_3 = c\}$ is disjoint from S,

and the imputation

$$z = (\tfrac{1}{4}(3a+b),\ 1-c-\tfrac{1}{4}(a+3b),\ c+\tfrac{1}{2}(b-a))$$

dominates all imputations in S). Select any imputation w, with $w_1, w_2 > 0$ and $c < w_3 < 1$, for which the expectation of player 3 in the bargaining game beginning at w is at least c (if no such w exists, a simple argument applies in the set of imputations $\{w: w_3 < c\}$). Let $(\alpha, 1-c-\alpha, c)$ be the essential expectations of the players, in the game beginning at w. Either $\alpha > w_1$ or $1-c-\alpha > w_2$; without loss of generality assume the former. Since S is dense, there must be a $y \in S$ for which $w \operatorname{dom}_{\{1,3\}} y$. But then $\{1,3\}$ gains (3 only temporarily) by an objection of $w^{\{1,3\}}$ to y, and c cannot be in equilibrium, a contradiction. Thus no stationary set of this type can exist.

(2) The stationary set S consists of three imputations

$$x = (\alpha, \beta, 1-\alpha-\beta),$$
$$y = (\alpha, \gamma, 1-\alpha-\gamma) \quad \text{with} \quad \beta > \gamma, \quad \text{and}$$
$$z = (\alpha+\beta-\gamma, \gamma, 1-\alpha-\beta).$$

Then $\frac{1}{3}(x+y+z)$ dominates all imputations in S, contradicting Lemma 13.2.

(3) The stationary set S consists of three imputations

$$x = (\alpha, \beta, 1-\alpha-\beta),$$
$$y = (\alpha, \gamma, 1-\alpha-\gamma) \quad \text{with} \quad \beta < \gamma, \quad \text{and}$$
$$z = (\alpha+\beta-\gamma, \gamma, 1-\alpha-\beta).$$

Consider the cases $\alpha+\beta-\gamma > 0$, $\beta \cdot 0$, $1-\alpha-\gamma > 0$; assume without loss of generality that the first holds. Then

$$w = (0,\ \gamma+\tfrac{1}{2}(\alpha+\beta-\gamma),\ 1-\alpha-\beta+\tfrac{1}{2}(\alpha+\beta-\gamma))$$

dominates all three imputations with respect to $\{2,3\}$, contradicting Lemma 13.2. The only remaining possibility is that

$$S = \{(\tfrac{1}{2},\tfrac{1}{2},0),(\tfrac{1}{2},0,\tfrac{1}{2}),(0,\tfrac{1}{2},\tfrac{1}{2})\}.$$

Theorem 13.3. This set S is a stationary set (and hence the only stationary set) of the three-person constant-sum game.

Proof. Let the strategies in σ be defined as follows.

(a) If x is an imputation and

(1) $x_i, x_j < \frac{1}{2}$, then $\{i,j\}$ objects to x with $(\frac{1}{2},\frac{1}{2})$.

(2) $x_i, x_j \geq \frac{1}{2}$ and $x_k < 0$, then $\{k\}$ objects to x with (0).

(b) If (α) is an objection of $\{k\}$, then $\{i,j\}$ responds with $(\frac{1-\alpha}{2}, \frac{1-\alpha}{2})$.

The effect of these strategies is represented in Figure 13.1, and it is easily seen that σ is in equilibrium and that S is the stationary set of the bargaining solution σ.

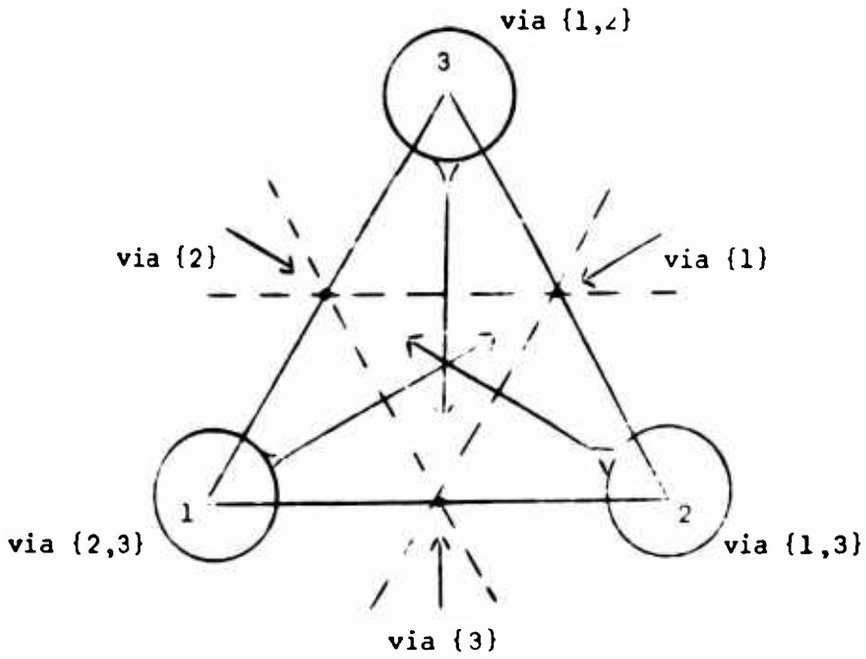

Figure 13.1 A solution of the constant-sum game.

A discussion of other bargaining solutions yielding this same stationary set will be given at the end of this section.

We now turn our attention to the 3-person simple game with one veto player. We assume the veto player to be player 3; thus the <u>veto game</u> we consider satisfies

$$v(i) = 0 \quad \text{for all} \quad i,$$

$$v(12) = 0,$$

$$v(13) = v(23) = v(123) = 1.$$

<u>Lemma 13.4</u>. No stationary set S of this game contains two distinct imputations

$$x = (\alpha, \gamma-\alpha, 1-\gamma)$$

and

$$y = (\beta, \gamma-\beta, 1-\gamma), \quad \text{with} \quad \alpha < \beta.$$

Proof. Assume to the contrary that the two imputations are both in a stationary set S of a bargaining solution σ. Consider the imputation

$$z = (\tfrac{1}{4}(3\alpha+\beta), \gamma - \tfrac{1}{4}(\alpha+3\beta), 1- \gamma + \tfrac{1}{2}(\beta-\alpha)).$$

Clearly, $z \text{ dom}_{\{1,3\}} x$ and $z \text{ dom}_{\{2,3\}} y$. Therefore, by Lemma 13.1, z is not in S. By boundedness of σ, all imputations essentially resulting from objections to z must be in S. But since any imputation w in a chain of dominance beginning at z satisfies $w_3 > z_3 > 1-\gamma$ (by the dominance pattern of the game), any stationary imputation w arising from z must satisfy $w_1 \leq \alpha < z_1$ and $w_2 \leq \gamma-\beta < z_2$ to preserve internal stability of S. Thus neither $\{1,3\}$ nor $\{2,3\}$ is motivated to object to z, and z must be stationary, a contradiction.

Lemma 13.5. If S is a stationary set for the veto game, then for every $0 \leq \alpha \leq 1$ S contains an imputation of the form $(\beta, \alpha-\beta, 1-\alpha)$, for some $0 \leq \beta \leq \alpha$.

Proof. Let $A \subset [0,1]$ be the set of all α for which no such imputation is in S. We first show that A contains no intervals. Assume to the contrary that $I = (c,d)$ is a maximal open interval in A. Omitting the easily-treated special cases $c = 0$ and $d = 1$, let λ be a limit point of S with $\lambda_3 = d$, and η be a limit point of S with $\eta_3 = c$, and consider the (non-stationary) imputation $\mu = \frac{1}{2}(\lambda+\eta)$. Due to the

internal stability of S, λ cannot dominate η and therefore does not dominate μ. A further application of internal stability (near λ) shows that every stationary point essentially arising from μ has both first and second components less than those of μ and hence no coalition is motivated to act against μ. This means that μ is stationary, contradicting the definition of A. Therefore, $[0,1]\backslash A$ is dense in $[0,1]$. A slight modification of the preceding argument now establishes that A is empty.

Lemmas 13.4 and 13.5, and internal stability, imply that S is a monotone curve from the imputation $(0,0,1)$ to an imputation in $\{x: x_3 = 0\}$. We finally show that this curve has no chords parallel to $\{x: x_1 = 0\}$ or to $\{x: x_2 = 0\}$, and is in fact a straight line.

Lemma 13.6. Let $f: [0,1] \to R$ be defined by $f(c) = \alpha$, where $(\alpha, 1-c-\alpha, c)$ is in S. Then f is strictly decreasing, and linear.

Proof. If $f(c_2) > f(c_1)$ for $c_2 > c_1$, then the stationary imputation with third component c_2 dominates the stationary imputation with third component c_1 with respect to the coalition $\{1,3\}$, violating internal stability. Thus to complete the proof we need only show that f is not constant on any interval. Assume to the contrary that $f(c_2) = f(c_1)$ for some $c_2 > c_1$, and let

$$w = (f(c_1) + \tfrac{1}{2}(c_2-c_1),\ 1 - \tfrac{1}{4}(c_1+3c_2) - f(c_1),\ \tfrac{1}{4}(3c_1+c_2)).$$

Then w is not stationary, and some coalition must be motivated against it. Since $w \operatorname{dom}_{\{1,3\}} (f(c_1),\ 1 - c_1 - f(c_1),\ c_1)$ and no one objects to

$(f(c_1), 1 - c_1 - f(c_1), c_1)$, this coalition must be $\{2,3\}$. But every stationary point which dominates w with respect to $\{2,3\}$ has first component $f(c_1)$, and motivation considerations imply that only such stationary points can essentially arise from w. The impact of this is that $\{1,3\}$ gains by objecting to $(f(c_1), 1 - c_1 - f(c_1), c_1)$ with w, contradicting the assumption that this imputation is stationary. Linearity is required to prevent each coalition from being able to randomize objections in such a manner that the resulting expectations dominate a stationary imputation.

Theorem 13.7. Every linear curve S, strictly decreasing for both players 1 and 2 in the sense of Lemma 13.6, is a stationary set of the veto game.

Proof. Fix α and β so that $0 < \alpha,\beta < 1$, and for each $x \notin S$ define

$$L_1(x,\alpha) = \{(x_1 + t\alpha x_2, (1-t)x_2, x_3 + t(1-\alpha)x_2): \ t > 0\}$$

and

$$L_2(x,\beta) = \{((1-t)x_1, x_2 + t\beta x_1, x_3 + t(1-\beta)x_1): \ t > 0\}.$$

Let the strategies in σ be defined as follows.

(a) If $L_1(x,\alpha) \cap S = \{w\}$, then $\{1,3\}$ objects to x with $w^{\{1,3\}}$.

If $L_2(x,\beta) \cap S = \{w\}$, then $\{2,3\}$ objects to x with $w^{\{2,3\}}$.

(b) If $L_1(x,\alpha) \cap S = \emptyset$ and $L_2(x,\beta) \cap S = \emptyset$, and

(1) $x_1 < 0$ and $x_1 + \alpha x_2 \leq 0$, then $\{1\}$ objects to x with (0).

(2) $x_2 < 0$ and $\beta x_1 + x_2 \leq 0$, then $\{2\}$ objects to x with (0).

(3) neither of the above conditions holds, then {3} objects to x with (0).

(c) If (γ) is an objection of {i} $(i \neq 3)$, then {j,3} responds with $(-\frac{\gamma}{2}, 1 - \frac{\gamma}{2})$.

If (γ) is an objection of 3 and $w = (w_1, w_2, 0)$ is the (unique) imputation in S with third component equal to zero, then {1,2} responds with $(w_1 - \frac{\gamma}{2}, w_2 - \frac{\gamma}{2})$.

It is easily shown that σ is indeed in equilibrium, and that S is the stationary set of σ. The effect of these strategies is represented in Figure 13.2.

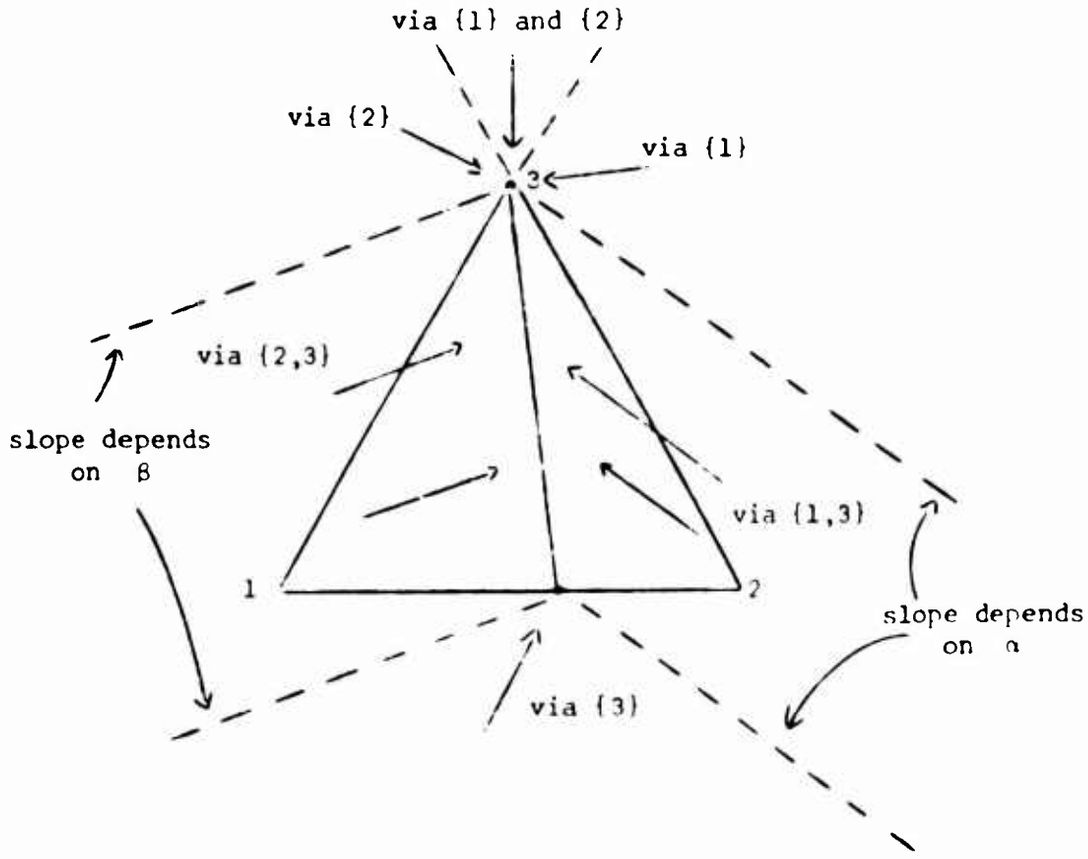

Figure 13.2 A solution of the veto game.

Having solved the constant-sum and veto games, we can now explicitly characterize all stationary sets for every three-person game.

Theorem 13.8. If the game (N,v) has an empty core, then every stationary set for the game consists of the three imputations

$$(1 - v(2,3),\ \tfrac{1}{2}(v(1,2) + v(2,3) - v(1,3)),\ \tfrac{1}{2}(v(1,3) + v(2,3) - v(1,2))),$$

$$(\tfrac{1}{2}(v(1,3) + v(1,3) - v(2,3)),\ 1 - v(1,3),\ \tfrac{1}{2}(v(1,3) + v(2,3) - v(1,2))),\ \text{and}$$

$$(\tfrac{1}{2}(v(1,2) + v(1,3) - v(2,3)),\ \tfrac{1}{2}(v(1,2) + v(2,3) - v(1,3)),\ 1 - v(1,2)),$$

and three linear curves of the type in Lemma 13.6, one from each of these imputations to the corresponding side of the imputation simplex.

If the game has a nonempty core, then every stationary set consists of the core and three linear curves, one from each vertex of the core to the corresponding side of the simplex.

Proof. Duplicating arguments in vNM [17; pp. 403-419], it easily follows that every stationary set of a 3-person game can be obtained by "piecing together" the core with stationary sets of "smaller" constant-sum and veto games. Figures 13.3 and 13.4 represent typical bargaining solutions obtained in this manner.

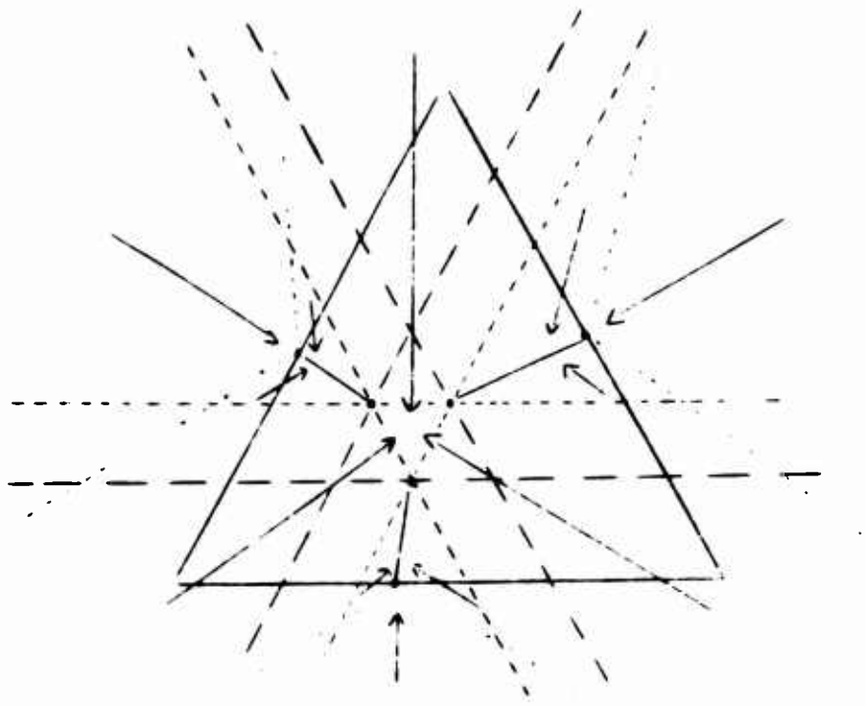

Figure 13.3 A solution of a game with empty core.

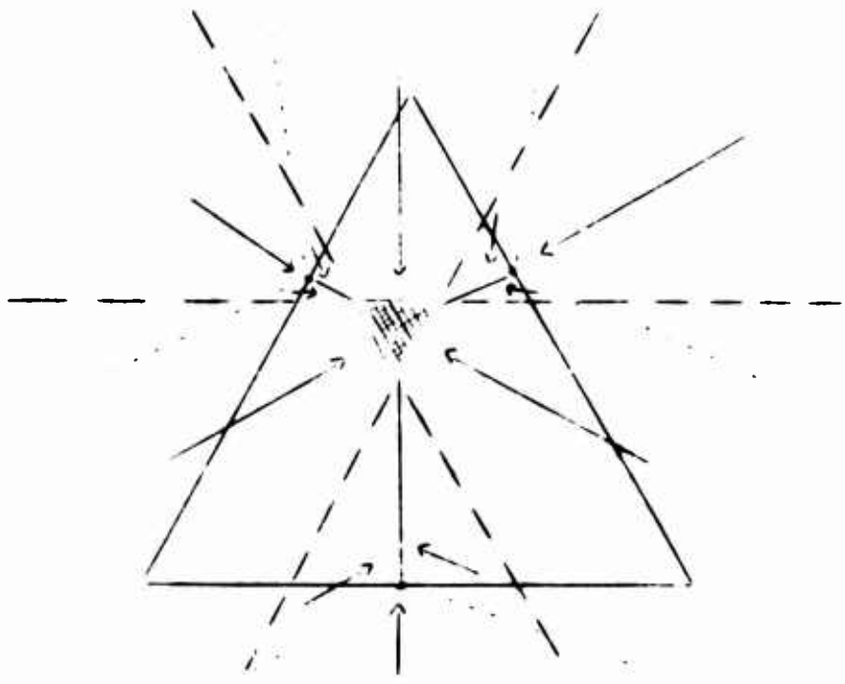

Figure 13.4 A solution of a game with nonempty core.

If we think of a bargaining solution σ as a "standard of behavior" for a game, we see that much more information is available to us now than in the classical vNM approach. Along with a set of imputations which is "stable" or "stationary", we also have a description of the dynamic process leading to agreement upon one of these imputations. Indeed, Theorem 13.7 suggests that many different types of bargaining behavior may be associated with the same stationary set. The constant α in the construction may be considered as a measure of the "bargaining ability" of player 1 against player 3, where values close to one indicate that 1 is the dominant bargainer and values close to zero attribute greater bargaining ability to 3. β may be similarly interpreted. Of course, other bargaining solutions will yield the same stationary set, and thus different standards of behavior may be associated with the same set of stationary outcomes.

The situation is similar for the constant-sum game. For example, the collection of imputations $\{x: x_i < \frac{1}{2} \text{ for all } i\}$ may be arbitrarily partitioned between the three 2-player coalitions so that only one coalition objects to each such imputation in the manner specified in Theorem 13.3. If all other objections and responses are kept as in Theorem 13.3, this new collection of strategies is another bargaining solution, with the same three-imputation stationary set. This result, and the manner in which bargaining solutions can vary outside the simplex of individually-rational imputations by variation of 2-player response strategies, suggest the complexities of bargaining behavior which may stand behind a deceptively simple-looking stationary set.

14. Games with Stable Cores

In this section, we consider games for which the core is a stationary set, and characterize a class of such games for which the core is also the unique vNM-stable set. Included in this class are all convex games, all games with "large" cores (in the sense of Gillies), and all symmetric games with vNM-stable cores. In the course of our investigation we shall use our results to prove the vNM-stability of the cores of the games considered, and shall explicitly construct bargaining solutions for which only imputations in the core are stationary. An open question is whether the class of games with stationary cores contains all games with vNM-stable cores.

We will require the concept of the "cover" of a game. Let M be any coalition of players. A collection $\{\gamma_T\}_{T\subset M}$ of non-negative numbers is balanced on M if for every $i \in M$,

$$\sum_{T\ni i} \gamma_T = 1.$$

Let (N,v) be an arbitrary game. The cover $\bar{v}$ of this game is the characteristic function on N defined for every coalition $M \subset N$ by

$$\bar{v}(M) = \max \sum_{T\subset M} \gamma_T v(T),$$

where the maximization is over all collections $\{\gamma_T\}$ which are balanced on M. The game (N,v) is slightly convex if for every pair of coalitions $A \cup B = N$,

$$v(N) + \bar{v}(A \cap B) \geq v(A) + v(B).$$

Let $M \subsetneq N$ be a coalition, and $0 \leq \alpha \leq v(M)$. The game (M^c, v_α^M) is the game on M^c defined by

$$v_\alpha^M(M^c) = v(N) - \alpha,$$

and for all $\emptyset \neq T \subsetneq M^c$,

$$v_\alpha^M(T) = \max_{W \subset M} (v(T \cup W) - \bar{v}(W)).$$

This game may be thought of as the "residual" game which remains after the amount α has been withdrawn from the game and promised to the players of M. The first theorem will require a simple lemma.

Lemma 14.1. Let M be a coalition, and $x \in R^M$. If $x(T) \geq v(T)$ for all $T \subset M$, then $x(M) \geq \bar{v}(M)$.

Proof. For any collection $\{\gamma_T\}$ balanced on M,

$$x(M) = \sum \gamma_T x(T) \geq \sum \gamma_T v(T).$$

Therefore

$$x(M) \geq \max \sum \gamma_T v(T) = \bar{v}(M).$$

The impact of our first theorem relates to any slightly convex game (N,v) and coalition M for which $(M^c, v^M_{v(M)})$ has a nonempty core. If M objects to an imputation and M^c responds with an allocation in the core of v^M, then any minimal coalition which can make an objection in the next stage of the bargaining game must be contained in M.

Theorem 14.2. Let $x^M \in R^M$ satisfy $x^M(M) \leq v(M)$, and assume v is slightly convex. Assume that $(M^c, v^M_{x^M(M)})$ has a nonempty core. Take $x^{M^c} \in \text{core}(v^M)$, and $x = x^M + x^{M^c}$. If $x(S) < v(S)$, and $x(T) \geq v(T)$ for all $T \subsetneq S$, then $S \subset M$.

Proof. Assume that S satisfies the conditions of the theorem, but that $S \not\subset M$. Then $x^{M^c} \in \text{core}(v^M)$ implies $S \not\subset M^c$, and either

$$\begin{aligned}
\text{(a)}\quad S \not\supset M^c, \text{ and } x(S) &= x(S \cap M) + x(S \cap M^c) \\
&\geq x(S \cap M) + (v(S) - \bar{v}(S \cap M)) \\
&\geq x(S \cap M) + v(S) - x(S \cap M) \\
&= v(S),
\end{aligned}$$

where the second inequality follows from the preceding lemma and the conditions on x and S, or

$$\begin{aligned}
\text{(b)}\quad S \supset M^c, \text{ and } x(S) &= x(S \cap M) + x(M^c) \\
&\geq \bar{v}(S \cap M) + v(N) - v(M) \\
&\geq v(S),
\end{aligned}$$

where the second inequality follows from slight convexity.

Either case contradicts the assumption that $x(S) < v(S)$, and therefore $S \subset M$, as claimed.

Corollary 14.3. If (N,v) is slightly convex, $x^M \in R^M$, $x^M(M) = v(M)$ and $x^M(T) \geq v(T)$ for all $T \subset M$, and if $x^{M^c} \in \text{core}(v^M_{v(M)})$, then $x = x^M + x^{M^c} \in \text{core}(v)$.

Proof. By the preceding theorem, any minimal S for which $x(S) < v(S)$ is contained in M. But $S \subset M$ implies $x(S) \geq v(S)$. There re $x(S) \geq v(S)$ for all $S \subset N$, and $x \in \text{core}(v)$.

Corollary 14.4. Let (N,v) be slightly convex and have the property: for every $y \notin \text{core}(v)$ there is a coalition M and an $x^M \in R^M$ such that $x^M(M) = v(M)$, $x^M \text{ dom}_M y$, $x^M(T) \geq v(T)$ for all $T \subset M$, and $\text{core}(v^M_{v(M)})$ is not empty. Then v has a vNM-stable core.

Proof. For every $y \notin \text{core}(v)$, the previous results and the stated property imply, for a specific M and x^M, the existence of an x^{M^C} for which $x = x^M + x^{M^C} \in \text{core}(v)$ and $x \text{ dom}_M y$.

A game (N,v) is extremely stable if, for every $M \subset N$, the core of $(M^C, v^M_{v(M)})$ is nonempty, and if the game (N,v) is slightly convex.*

Theorem 14.5. If (N,v) is extremely stable, then the core of the game is a stationary set.

Proof. We first define objection and response strategies for all coalitions.

(a) At each $y \notin \text{core}(v)$, select a minimal coalition M for which $y(M) < v(M)$. Let x^M be defined by

$$x^M_i = y_i + (v(M) - y(M))/|M|, \quad \text{for all} \quad i \in M.$$

*In the specific games treated later in this section, slight convexity is easily verified.

Then M objects to y with the allocation x^M, and no other coalition objects to y.

(b) For every coalition M and allocation y^M (with $y^M(M) \leq v(M)$), let the coalition M^c respond to the objection y^M with a $y^{M^c} \in \text{core}(v^M_{y^M(M)})$.

Call the above-described collection of strategies σ. We claim that σ is in equilibrium. If it is not, there is some subgame in which a coalition S gains by a change of strategy. We examine the possibilities by cases.

(1) S changes its objection to an imputation x. Since S is a minimal coalition for which $x(S) < v(S)$, if S changes its objection from y^S to z^S with $z^S(S) = v(S)$ then by the extreme stability of (N,v) and Corollary 14.3 it follows that the response of S^c in σ yields a (stationary) imputation in the core of v. Thus the players in S share the same total amount $(z^S(S))$ as they did in σ $(y^S(S))$, and since $z^S \neq y^S$, some player in S loses in the change of strategy. On the other hand, if the new objection z^S satisfies $z^S(S) < v(S)$, then by Theorem 14.2 the response of S^c leaves S as the <u>only</u> minimal coalition with $z(S) < v(S)$. Thus no other coalition will object in σ in the next stage of the bargaining game, and the players of S share an intermediate amount less than $v(S)$ and in no later stage share more than $v(S)$. Thus some player in S again loses in the change of strategy.

(2) S initiates an objection y^S to an imputation x, to which some other coalition T objects in σ. By reasoning similar to that in (1), if S is a minimal coalition for which $x(S) < v(S)$ then the players of S expect at most $v(S)$ from this strategy change, but expected at least $v(S)$ in the (stationary) core imputation resulting from the objection of T and response of T^c. Thus not all players in S gain in the strategy change. If S is not minimal, then Theorem 14.2 states that only subcoalitions of S might object in the next stage, and hence the players of some such subcoalition T share exactly $v(T)$ in the final outcome. Again, not all players of T and therefore not all players of S gain in the strategy change.

(3) S changes its response to some objection x^{S^c} of S^c. In σ, S must have expected $v(N) - v(S^c)$ in the response-bargaining game beginning at x^{S^c}. Any change in strategy must yield a game in which the players of S^c receive at least $v(S^c)$ (that is, a core imputation), and therefore the players of S share no more at the end of the game than they did in σ. Hence if any player in S gains, another must lose.

This completes the consideration of possible coalitional deviations from σ, and we have shown that σ is in equilibrium.

Having shown that extremely stable games have cores which are both vNM-stable and stationary sets, we shall now show that all games of several well-known types are extremely stable. A game (N,v) is <u>convex</u> if for every pair of coalitions $A,B \subset N$,

$$v(A \cup B) + v(A \cap B) \geq v(A) + v(B).$$

Shapley has considered convex games in several papers [21,23], and has shown that all convex games have vNM-stable cores. After a preliminary lemma, which shows that convexity is characterized by "increasing marginal gains", we adapt Shapley's approach to prove that all convex games are extremely stable.

Lemma 14.6. A game (N,v) is convex if and only if for every $k \in N$ and coalitions $S \supset T$ with $k \notin S$,

$$v(S \cup \{k\}) - v(S) \geq v(T \cup \{k\}) - v(T).$$

Proof. The forward implication follows directly from the definition of convexity, taking $A = S$ and $B = T \cup \{k\}$. The reverse implication is trivial for any pair A and B with $A \subset B$ or $B \subset A$, and follows from a simple inductive proof otherwise.

Theorem 14.7. If the game (N,v) is convex, then it is extremely stable.

Proof. Let M be any coalition. We wish to prove that $\text{core}(v^M_{v(M)})$ is nonempty. Relabel the players in N so that $M^c = \{1,2,\ldots,s\}$, and define $x \in R^{M^c}$ by

$$x_k = v(M \cup \{1,2,\ldots,k\}) - v(M \cup \{1,2,\ldots,k-1\})$$

for all $k \in M^c$. To see that $x \in \text{core}(v^M)$, consider any $\emptyset \neq T \subset M^c$. Then

$$x(T) = \sum_{k \in T} [v(M \cup \{1,2,\ldots,k\}) - v(M \cup \{1,2,\ldots,k-1\})].$$

Let the players of T be $k_1 < k_2 < \cdots < k_t$. Lemma 14.6 implies that for any player k_r,

$$v(M \cup \{1,2,\ldots,k_r\}) - v(M \cup \{1,2,\ldots,k_r-1\}) \geq$$

$$v(M \cup \{k_1,\ldots,k_r\}) - v(M \cup \{k_1,\ldots,k_{r-1}\}).$$

Thus

$$\begin{aligned} x(T) &\geq \sum_{r=1}^{t} [v(M \cup \{k_1,\ldots,k_r\}) - v(M \cup \{k_1,\ldots,k_{r-1}\})] \\ &= v(M \cup T) - v(M) \\ &\geq v(W \cup T) - v(W) \end{aligned}$$

for any $W \subset M$, by convexity. Since $\bar{v}(W) \geq v(W)$ (indeed, equality holds for convex games), this implies $x(T) \geq \max_{W \subset M} [v(W \cup T) - \bar{v}(W)] = v^M(T)$.

We next consider games with "large" cores. Such games were first studied by Gillies [5], who showed that any n-person game with characteristic function satisfying

$$v(S) \leq \frac{v(N)}{n} \quad \text{for all} \quad S \subset N$$

has a VNM-stable core. The following theorem slightly extends this result.

Theorem 14.8. Let the game (N,v) satisfy

$$v(S) \leq v(N)/n-1 \quad \text{for all} \quad S \subset N.$$

Then the game is extremely stable.

Proof. To see that $\text{core}(v^M_{v(M)})$ is nonempty for any M, define $x \in R^{M^c}$ by

$$x_k = (v(N) - v(M))/(n-m)$$

for all $k \in M^c$, where $m = |M|$. For any $\emptyset \neq T \subset_{\neq} M^c$,

$$\begin{aligned} x(T) &= \frac{t}{n-m}(v(N) - v(M)) \\ &\geq \frac{1}{n-m}(v(N) - v(N)/n-1) \\ &= \frac{n-2}{(n-m)(n-1)} v(N) \\ &\geq v(N)/n-1, \text{ since } m \geq 2 \text{ (or } v(M) = v(i) = 0). \end{aligned}$$

Also,

$$\begin{aligned} v^M(T) &= \max_{W \subset M} (v(T \cup W) - \bar{v}(W)) \\ &\leq \max_{W \subset M} v(T \cup W) \\ &\leq v(N)/n-1. \end{aligned}$$

Thus $x(T) \geq v^M(T)$ for all $T \subset M$, and $x \in \text{core}(v^M)$.

Finally, we treat the class of symmetric games with vNM-stable cores. A game is symmetric if $v(S) = v(T)$ wherever S and T are coalitions containing the same number of players. Thus a symmetric game may be described by the collection of numbers $\{v_s\}_{s=1}^n$, where $v(S) = v_s$ for all coalitions S containing exactly s players. Shapley [24] has characterized the symmetric games with vNM-stable cores as those games satisfying

$$\frac{v_n - \bar{v}_k}{n-k} \geq \frac{v_t - \bar{v}_k}{t-k} \quad \text{for all} \quad k < t < n,$$

where $\{\bar{v}_s\}_{s=1}^{n}$ is the collection corresponding to the cover of v.

Corollary 14.4 and the following theorem establish the sufficiency of this condition.

Theorem 14.9. Let (N,v) be a symmetric game satisfying

$$\frac{v_n - \bar{v}_k}{n-k} \geq \frac{v_t - \bar{v}_k}{t-k} \quad \text{for all} \quad k < t < n.$$

Then the game is extremely stable.

Proof. To see that $\text{core}(v^M_{v_m})$ is nonempty for each M, define $x \in R^{M^c}$ by

$$x_k = (v_n - v_m)/(n-m) \quad \text{for all} \quad k \in M^c.$$

For each $T \subset M^c$, we wish to show

$$t \cdot \frac{(v_n - v_m)}{n-m} = x(T) \geq v^M(T) = \max_{w \leq m} (v_{t+w} - \bar{v}_w).$$

It is therefore sufficient to show

$$\frac{v_n - v_m}{n-m} \geq \frac{v_{t+w} - \bar{v}_w}{t} \quad \text{for all} \quad t < n-m,\ w \leq m.$$

If $w = m$, the result is obvious, since $v_m \leq \bar{v}_m$. If $w < m$, then

$$\frac{v_n - \bar{v}_w}{n-w} \geq \frac{v_m - \bar{v}_w}{m-w} \quad \text{implies} \quad \frac{v_n - v_m}{n-m} \geq \frac{v_n - \bar{v}_w}{n-w},$$

and this result combined with

$$\frac{v_n - \bar{v}_w}{n-w} \geq \frac{v_{t+w} - \bar{v}_w}{t}$$

immediately yields the desired inequality. Thus $x \in \text{core}(v^M_{v_m})$, and the game is extremely stable.

Kulakovskaja [10] has recently announced a characterization of games with vNM-stable cores. A comparison of our definition of extreme stability with Kulakovskaja's condition is interesting.

Extreme stability: For every M, $\text{core}(v^M_{v(M)})$ is nonempty.

Kulakovskaja's condition: For every $y \notin \text{core}(v)$, there is some M (minimal with respect to $y(M) < v(M)$) for which $\text{core}(v^M_{v(M)})$ is nonempty.

The sufficiency of Kulakovskaja's condition follows from Corollary 14.4. Furthermore, Kulakovskaja's condition implies extreme stability for every game in which every M can occur as the unique minimal dominating coalition for some y (clearly, this in turn would imply $v = \bar{v}$). This apparently close relationship leads us to pose the following open question:

Is the core a stationary set for every game with vNM-stable core?

The core can be a stationary set for games in which it is not stable. This is the subject of the next section.

15. Two Classical Pathologies

In the early days of the vNM theory, a number of conjectures were made concerning the theory's "regularity". For example, it was conjectured that no game had a unique stable set strictly containing the core, and that every game had at least one stable set. These conjectures fell with most others in the 1960's, when a number of "pathological" games were discovered [11,12,13,14,20,22,25]. All of these pathologies have two common characteristics. They are all games with nonemptv cores of dimension less than that of the imputation simplex, and in each of them this "small" core dominates all except a lower-dimensional set of imputations.

In this section we consider two such games, and show that the core is a stationary set of each. The first example, due to Lucas [11], is a five-person game with a unique stable set strictly larger than the core. The second example is a seven-person non-side-payment game, discovered by Stearns [25], which has no stable set. The fact that we can easily show that the core is a stationary set in each of these games encourages us in our hope that the classical pathologies can all be adequately explained by an objection-response theory, and leads us to shakily conjecture that the other pathologies, including Lucas' [14] example of a 10-person side-payment game with no stable set, similarly have stationary cores.

The Lucas five-person example is defined by

$$v(12345) = 2,$$

$$v(12) = v(34) = v(135) = v(245) = 1,$$

$$v(S) = 0 \quad \text{otherwise.}$$

The core of this game is the line segment with vertices $(1,0,0,1,0)$ and $(0,1,1,0,0)$ and the unique vNM stable set is the square with vertices $(1,0,1,0,0)$, $(0,1,0,1,0)$, and the two preceding. Note that the core is simply a diagonal of this square.

We shall now define a bargaining solution σ, by first presenting the objection strategies in tabular form. Let x be any imputation.

Condition	Objecting coalition	Objection
$1 - x(12) = \varepsilon > 0$	$\{1,2\}$	$(x_1 + \frac{\varepsilon}{2}, x_2 + \frac{\varepsilon}{2})$
$x(12) \geq 1$ $1 - x(34) = \varepsilon > 0$	$\{3,4\}$	$(x_3 + \frac{\varepsilon}{2}, x_4 + \frac{\varepsilon}{2})$
$x(12), x(34) \geq 1$ $1 - x(135) = \varepsilon > 0$	$\{1,3,5\}$	$(x_1 + \frac{\varepsilon}{3}, x_3 + \frac{\varepsilon}{3}, \frac{\varepsilon}{3})$
$x(12), x(34), x(135) \geq 1$ $1 - x(245) = \varepsilon > 0$	$\{2,4,5\}$	$(x_2 + \frac{\varepsilon}{3}, x_4 + \frac{\varepsilon}{3}, \frac{\varepsilon}{3})$

It should be noted that no objection is made to x (and therefore x is stationary) only if x is in the core, and also that the last two of the four cases imply $x_5 = 0$.

Similarly, we present the response strategies of σ in tabular form.

Objection	Responding coalition	Response
(y_1,y_2) by $\{1,2\}$	$\{3,4,5\}$	$(y_2+\frac{\epsilon}{2},\ y_1+\frac{\epsilon}{2},\ 0)$, where $\epsilon = 2-2y(12) \geq 0$.
(y_3,y_4) by $\{3,4\}$	$\{1,2,5\}$	$(y_4+\frac{\epsilon}{2},\ y_3+\frac{\epsilon}{2},\ 0)$, where $\epsilon = 2-2y(34) \geq 0$.
(y_1,y_3,y_5) by $\{1,3,5\}$	$\{2,4\}$	$(y_3+\frac{\epsilon}{2},\ y_1+\frac{\epsilon}{2})$, where $\epsilon = 2-2y(13)-y_5 \geq 0$.
(y_2,y_4,y_5) by $\{2,4,5\}$	$\{1,3\}$	$(y_4+\frac{\epsilon}{2},\ y_2+\frac{\epsilon}{2})$, where $\epsilon = 2-2y(24)-y_5 \geq 0$.

To see that σ is in equilibrium, first consider the objection strategies. If an objection by either of the coalitions $\{1,2\}$ or $\{3,4\}$ is made, the specified response yields a stationary (core) imputation, and therefore the objection is motivated. On the other hand, if $(x_1,\ 1-x_1)$, $x_3,\ 1-x_3,\ 0)$ is an imputation satisfying $\epsilon = 1-x(13) > 0$, an objection by $\{1,3,5\}$ with the specified response yields a non-stationary imputation to which $\{12\}$ objects. Hence, after <u>two</u> objections and responses, the stationary imputation

$$(x_1+\frac{5}{12}\epsilon,\ x_3+\frac{7}{12}\epsilon,\ x_3+\frac{7}{12}\epsilon,\ x_1+\frac{5}{12}\epsilon,\ 0)$$

results. Thus players 1 and 3 gain, and player 5 gains temporarily, from the objection by $\{1,3,5\}$, and hence this objection is also motivated. A similar result holds for objections by $\{2,4,5\}$.

Next, consider a typical stationary imputation $(\alpha,\ 1-\alpha,\ 1-\alpha,\ \alpha,\ 0)$, for any $0 \leq \alpha \leq 1$. Observe that the coalitions $\{1,2\},\{3,4\},\{1,3\} \subset \{1,3,5\}$ and $\{2,4\} \subset \{2,4,5\}$ all receive exactly $\alpha+(1-\alpha) = 1$ in every

such imputation. From this it easily follows that no coalition gains by changing its strategy in σ. Indeed, it is precisely this "balanced" nature of the core imputations, linking the "powerless" pairs 1 and 4, and 2 and 3, to equal payoffs, which allows us to conclude that the core is a stationary set of the game.

One further note on this game is in order. Any attempt to "verbalize" the characteristic function of the game involves designating player 5 as an agent whose sole role is to empower objections involving $\{1,3\}$ or $\{2,4\}$. This role is well-reflected in the bargaining solution σ, in which 5's expectation is limited to his temporary gains when $\{1,3,5\}$ or $\{2,4,5\}$ act.

The Stearns 7-person example is a non-side-payment game. The general definitions for a non-side-payment game may be found in [3]. For our purposes, we need only slightly extend the concepts with which we have been working. Define four vectors

$$p_1 = (1,1,2,0,0,0,0),$$
$$p_2 = (0,0,1,1,2,0,0),$$
$$p_3 = (2,0,0,0,1,1,0),$$
$$\text{and} \quad c = (2,0,2,0,2,0,1).$$

An imputation will be any convex combination of these vectors, and an imputation x will be said to dominate an imputation y if $x >_S y$, where S is any of the coalitions $\{1,2,7\}$, $\{3,4,7\}$, $\{5,6,7\}$, or $\{1,3,5\}$. With respect to this dominance relation the only imputation which is undominated is c, and hence the core of the game consists of this single imputation.

Let L_i $(i = 1,2,3)$ be the half-open line segment $\{x: x = \alpha p_i + (1-\alpha)c, 0 < \alpha \leq 1\}$. It may be noted that c dominates every imputation (other than itself) not on any of the L_i. We define a bargaining solution σ, first specifying the objection strategies. Let x be any imputation.

Condition	Objecting coalition	Objection
$x \neq c$, $x \notin L_1 \cup L_2 \cup L_3$	$\{1,3,5\}$	$(2,2,2)$
$x \in L_1$	$\{5,6,7\}$	$(\frac{1}{2} - \frac{x_7}{2})(1,1,0) + (\frac{1}{2} + \frac{x_7}{2})(2,0,1)$
$x \in L_2$	$\{1,2,7\}$	$(\frac{1}{2} - \frac{x_7}{2})(1,1,0) + (\frac{1}{2} + \frac{x_7}{2})(2,0,1)$
$x \in L_3$	$\{3,4,7\}$	$(\frac{1}{2} - \frac{x_7}{2})(1,1,0) + (\frac{1}{2} + \frac{x_7}{2})(2,0,1)$

Next, we define the response strategies in σ, keeping in mind the requirement that every objection-and-response must yield an imputation.

Objection	Responding coalition	Response
(x_1,x_3,x_5) by $\{135\}$	$\{2,4,6,7\}$	$(0,0,0,1)$
(x_5,x_6,x_7) by $\{567\}$	$\{1,2,3,4\}$	$x_6(2,0,0,0) + x_7(2,0,2,0) + (\frac{x_5}{2} - x_6 - 2x_7)(0,0,1,1)$
(x_1,x_2,x_7) by $\{127\}$	$\{3,4,5,6\}$	$x_2(2,0,0,0) + x_7(2,0,2,0) + (\frac{x_1}{2} - x_2 - 2x_7)(0,0,1,1)$
(x_3,x_4,x_7) by $\{347\}$	$\{1,2,5,6\}$	$x_4(2,0,0,0) + x_7(2,0,2,0) + (\frac{x_3}{2} - x_4 - 2x_7)(0,0,1,1)$

It is perhaps simplest to visualize the action induced by σ by noting that the objection-and-response to any imputation other than c, not in $L_1 \cup L_2 \cup L_3$, results in the imputation c, while the objection-and-response to $x \in L_i$ results in a convex combination of c, p_j, and p_k ($\{i,j,k\} = \{1,2,3\}$), which after the next objection-and-response leads to c. Since σ has a unique stationary imputation, namely c, it is particularly easy to see that σ is in equilibrium, and that the core of the game is therefore a (one-element) stationary set.

16. Majority Games

Two classes of simple games* which have received much attention are the symmetric simple games and the homogeneous games. A symmetric simple (n,k)-game is an n-player game, with $n < 2k$, in which

$$v(S) = \begin{cases} 0 & \text{if} \quad |S| < k \\ 1 & \text{if} \quad |S| \geq k. \end{cases}$$

Each (n,k)-game has a unique symmetric stable set, described by Bott [4]. A homogeneous game is an n-player game, with an associated set of positive numbers $\{\gamma_i\}_{i=1}^{n}$, in which

*A game is simple if for every coalition S, $v(S) = 0$ or 1. If $v(S) = 0$, S is losing; if $v(S) = 1$, S is winning.

$$v(S) = \begin{cases} 0 & \text{if} \quad \sum_{i \in S} \gamma_i < 1 \\ 1 & \text{if} \quad \sum_{i \in S} \gamma_i \geq 1, \end{cases}$$

$$v(S) + v(S^c) = 1 \quad \text{for every coalition} \quad S,$$

and in which

$$\sum_{i \in S} \gamma_i = 1 \quad \text{for every minimal winning coalition} \quad S.$$

Each homogeneous game has a "simple" stable set, described by Gurk and Isbell [6].

These two classes intersect in a particular type of game, the majority game. Such a game is a (2m-1,m)-game, in which the winning coalitions are all coalitions containing at least half of the (odd number of) players. For a (2m-1,m)-game the symmetric and simple stable sets coincide, and consist of the imputation $(\frac{1}{m},\ldots,\frac{1}{m},0,\ldots,0)$ and its component permutations. When we refer to the symmetric stable set of a majority game, we shall mean this (finite) collection of imputations.

Consider a specific (2m-1,m)-game. We define a bargaining solution σ for this game, for which the stationary set of σ is the symmetric stable set of the game. The objection strategies in σ are such that only minimal (m-player) winning coalitions ever make objections. The m-player coalition S objects to every imputation x for which $x_i < \frac{1}{m}$ for all players in S, the objection of S being the symmetric allocation $(\frac{1}{m},\ldots,\frac{1}{m})$.

Response strategies must be specified for every coalition S with $|S| = k < m$, and for every objection x^{S^c} which S might face. If $1 - x(S^c) < \frac{k}{m}$, then S responds with the allocation $(\frac{1 - x(S^c)}{k}, \ldots, \frac{1 - x(S^c)}{k})$. However, if $1 - x(S^c) \geq \frac{k}{m}$, then S responds by selecting any player j of S at random (with probability $\frac{1}{k}$) and responding with the allocation x^S, where

$$x_i^S = \begin{cases} 0 & \text{if } i \neq j \\ 1 - x(S^c) & \text{if } i = j. \end{cases}$$

Any change of objection strategy by a minimal winning coalition cannot be to the advantage of all members of the coalition, since they already share (in their objection in σ) the entire available quantity 1. Furthermore, the response strategies in σ are arranged specifically so that no non-minimal coalition can benefit all of its members by initiating an objection. Also, the specified response strategies maximize, for each player in a responding coalition, his probability of belonging to the objecting coalition which gains the floor in the next stage of the game. These observations combine to prove that σ is in equilibrium. Three notes are in order. First, when the players follow the strategies in σ, each imputation is replaced by a stationary imputation after at most one stage of the bargaining game, and each objection that actually occurs requires only a trivial response (allocation of 0). And second, the bargaining solution σ is precisely the solution given for the 3-person

constant-sum game (the (3,2)-game) in Section 13. Finally, some careful consideration will convince one that the more obvious types of symmetric response strategies, such as an always-equal division among responding players, do not yield an equilibrium collection of strategies when combined with the stated objection strategies, since non-minimal winning coalitions can then gain in some subgames by initiating objections.

A natural question is whether our result (that the symmetric, simple stable set is a stationary set for majority games) extends in some way either to all (n,k)-games or to all homogeneous games. Although we shall not formally prove these statements, it appears as if the result does not generalize at all to symmetric solutions of (n,k)-games, but does generalize to simple stable sets of homogeneous games. We believe that the simple stable sets described in [6] can be shown to be stationary sets by a careful construction of response strategies similar to that given above.

However, consider a non-homogeneous symmetric simple game, such as the (4,3)-game. The symmetric stable set of this game consists of the set of imputations $\{(\alpha, \alpha, \frac{1}{2}-\alpha, \frac{1}{2}-\alpha): 0 \leq \alpha \leq \frac{1}{4}\}$, and all component permutations of these imputations. If σ is a symmetric bargaining solution for which this set is the stationary set, then $(1,0,0,0)$ is not stationary, and will be objected to by $\{2,3,4\}$ in σ. Since all components of every imputation in the stationary set, other than $(\frac{1}{2},\frac{1}{2},0,0)$ and its permutations, are positive, player 1 expects a positive essential return, say ε, from the bargaining game beginning at $(1,0,0,0)$. Thus each of players 2, 3 and 4 expects an essential return of $\frac{1}{3} - \frac{\varepsilon}{3}$ in σ. Now consider a coalitional deviation from σ by $\{2,3,4\}$ in the game beginning at $(1,0,0,0)$. With equal probability, let the coalition object

to $(1,0,0,0)$ with one of the three permutations of $(\frac{1}{2}-\frac{\varepsilon}{2}, \frac{1}{2}-\frac{\varepsilon}{2}, \frac{\varepsilon}{2})$. Since response strategies are trivial on the (4,3)-game, a stationary imputation will result from such an objection, and therefore each player of $\{2,3,4\}$ will expect an essential return of

$$(\frac{1}{2}-\frac{\varepsilon}{2}) + \frac{1}{3} \cdot \frac{\varepsilon}{2} = \frac{1}{3} - \frac{\varepsilon}{6}$$

from this coalitional deviation to σ. Hence, the indicated stable set cannot be a stationary set of the (4,3)-game.

At the present, we have not found <u>any</u> bargaining solutions to the (4,3)-game. We return to this point briefly in Section 18.

17. An Extension of the Bargaining Model

In a bargaining game, we require that all objections in any stage be simultaneously declared. This requirement, while simplifying much of the notation and discussions of this paper, can be considerably weakened without damaging our results.

Consider a hypothetical situation, in which two disjoint coalitions S and T are the only coalitions capable of objecting to a particular imputation. Further assume that behavior of all players in the bargaining game is so specified apart from this imputation, that the coalitional payoffs to S and T depend on their action in the manner indicated in Figure 17.1.

		T: objects	T: doesn't object
S	objects	(2,2)	(5,0)
	doesn't object	(0,5)	(4,4)

Figure 17.1

The return to coalition S (coalition T) is indicated by the first (second) component of each ordered pair. This situation, a form of the "Prisoner's Dilemma" [18], is not handled well by our model, since the only equilibrium which could exist would involve both coalitions making (simultaneous) objections.

However, consider a modification of the rules of the bargaining game, so that each objection-stage of the game consists of several steps. At each step, a coalition may make an objection to the current proposal, having already heard all objections raised to that proposal in earlier steps. After such a sequence of steps, the hierarchy selects an objecting coalition from all those which have declared objections, and the game continues as before.

In this manner threats become credible, and are explicitly included in reactive strategies. A typical combination of strategies for the two coalitions in the example above would be for neither to raise an objection in the first step, but for each to object in the second step if the other objects in the first step. The result of such strategies, involving as they do a form of "contingency planning", is to permit equilibria in which neither coalition raises an objection in the given situation.

A final note is in order. Permitting contingent objections may allow a more varied class of bargaining solutions for some games. However, every bargaining solution arising from our original model remains a solution in this extended model. Therefore the extension is an enrichment of our original theory in which all positive results of the preceding sections hold.

18. Summary

Having shown how the theory of bargaining solutions and stationary sets applies to several classes of games, we make here a few comments indicating several areas in which much work remains to be done.

With regard to three-person games, a characterization of all possible stationary sets which can arise in the context of any given hierarchical structure and system of preferences would be of interest. Preliminary results indicate that the bargaining solutions we have given are relatively insensitive to such changes, but that new classes of solutions can also arise.

A major question concerning games with vNM-stable cores is whether the core of every such game is a stationary set. A wider problem is to characterize all games with stationary cores, and particularly to determine whether all of the classical "pathologies" are of this type.

We expect that the simple solution of each simple homogeneous game is a stationary set of the game. However, discovery of stationary sets for non-constant-sum symmetric simple games seems difficult. A natural starting place is to ask whether the (4,3)-game has any stationary sets. If it does not, and if the non-existence seems attributable to arguments

similar to that given in Section 16, then a modification of our theory seems desirable. Such a modification, to invalidate arguments of the type referred to, would be of interest since we do not find ourselves wholly convinced that these arguments should be given much weight in determining bargaining solutions.

It may have been noted that all stationary sets determined in the preceding sections are self-policing, in the sense of Vickrey. This seems reasonable for the uncomplicated games investigated here, but there is no reason to expect the coincidence to endure in the analysis of more complex games. Still, we find the similarity of the results of the two classical theories and our own, at least concerning the simplest and most-studied games, quite encouraging.

And finally, we repeat that the derivation of an objection-and-response theory for cooperative normal-form games, carried out in a manner similar to the work done here, seems to be a natural and important direction for continuing investigation.

APPENDIX

DEFINITIONS AND NOTATION

An n-person characteristic function game is a pair (N,v), where $N = \{1,2,\ldots,n\}$ is the set of players. A coalition is a nonempty subset of players. The characteristic function v associates to each coalition S a real number $v(S)$, and satisfies

$$\sum_{P \in \mathcal{P}} v(P) \leq v(N)$$

for every partition $\mathcal{P}$ of N. We shall generally work with 0-normalized games, which satisfy $v(i) = 0$ for all one-player coalitions $\{i\}$.

An imputation is any vector $x \in R^n$ which satisfies $x(N) = v(N)$, where for any x and coalition S we write $x(S) = \sum_{i \in S} x_i$. The set of all imputations is

$$X = \{x \in R^n: \ x(N) = v(N)\}.$$

(At times we shall restrict our attention to individually-rational imputations. We shall then, without fear of confusion, write

$$X = \{x \in R^n: \ x(N) = v(N) \quad \text{and} \quad x_i \geq 0 \quad \text{for all} \quad i \in N\}).$$

For any coalition S and any real number α,

$$R^S = \{x \in R^n: \ x_i = 0 \quad \text{for all} \quad i \notin S\}$$

$$\text{and} \quad R^S(\alpha) = \{x \in R^S: \ x(S) \leq \alpha\},$$

and for any $x \in R^n$, x^S is the projection of x onto R^S, so that

$$x_i^S = \begin{cases} x_i & \text{if} \quad i \in S \\ 0 & \text{if} \quad i \notin S. \end{cases}$$

On occasion we shall "compose" vectors $x^S \in R^S$ and $x^{S^c} \in R^{S^c}$ to yield the vector $x \in R^n$, where $x = x^S + x^{S^c}$.

If x and y are imputations, then x <u>dominates</u> y <u>with respect to</u> a coalition S, written $x \ \text{dom}_S \ y$, if

$$x_i > y_i \quad \text{for all} \quad i \in S,$$

$$\text{and} \qquad x(S) \leq v(S).$$

Since this dominance depends only on x^S, we sometimes write $x^S \ \text{dom}_S \ y$. If $x \ \text{dom}_S \ y$ for some S, then x <u>dominates</u> y, written $x \ \text{dom} \ y$.

Given a set K of imputations, write

$$\text{Dom } K = \{x \in X: \ y \ \text{dom} \ x \quad \text{for some} \quad y \in K\}.$$

A von Neumann-Morgenstern <u>stable set</u> for a game [17] is a set K which satisfies

$$K \cap \text{Dom } K = \emptyset \quad \text{(internal stability)}$$

$$K \cup \text{Dom } K = X \quad \text{(external stability).}$$

The core of a game is defined by

$$\text{core}(v) = \{x \in X: \; x(S) \geq v(S) \quad \text{for all coalitions} \quad S \subset N\}$$

To insure external stability, the core must be contained in every stable set.

REFERENCES

[1] Aumann, Robert J., "Mixed and behavior strategies in infinite extensive games", Advances in Game Theory, Princeton University Press, Princeton, 1964, pp. 627-650.

[2] Aumann, Robert J., "Subjectivity and correlation in randomized strategies", RPGTME 84, Department of Mathematics, The Hebrew University of Jerusalem, 1973.

[3] Aumann, Robert J. and Bezalel Peleg, "von Neumann-Morgenstern solutions to cooperative games without side payments", Bull. Amer. Math. Soc. 66, 1960, pp. 173-179.

[4] Bott, Raoul, "Symmetric solutions to majority games", Contributions to the Theory of Games, Vol. II, Princeton University Press, Princeton, 1953, pp. 319-323.

[5] Gillies, Donald B., "Solutions to general non-zero-sum games", Contributions to the Theory of Games, Vol. IV, Princeton University Press, Princeton, 1959, pp. 47-85.

[6] Gurk, H. M. and J. R. Isbell, "Simple solutions", Contributions to the Theory of Games, Vol. IV, Princeton University Press, Princeton, 19[illegible], pp. 247-265.

[7] Harsanyi, John C., "A bargaining model for the cooperative n-person game", Contributions to the Theory of Games, Vol. IV, Princeton University Press, Princeton, 1959, pp. 325-355.

[8] Harsanyi, John C., "An equilibrium-point interpretation of stable sets and a proposed alternative definition", Management Science 20, 1974, pp. 1472-1495.

[9] Harsanyi, John C. and Reinhard Selten, "A generalized Nash solution for two-person bargaining games with incomplete information", Management Science 18, 1972, pp. 80-106.

[10] Kulakovskaja, T. E., "Necessary and sufficient conditions for the coincidence of core and solution in a classical cooperative game" (translated by R. F. Rinehart), Soviet Math. Dokl. 12, 1971, pp. 1231-1234.

[11] Lucas, William F., "A counterexample in game theory", Management Science 13, 1967, pp. 766-767.

[12] Lucas, William F., "Games with unique solutions that are nonconvex", Pacific J. Math. 28, 1969, pp. 599-602.

[13] Lucas, William F., "A game in partition function form with no solution", SIAM J. Appl. Math. 16, 1968, pp. 582-585.

[14] Lucas, William F., "A game with no solution", Bull. Amer. Math. Soc. 74, 1968, pp. 237-239.

[15] Nash, John F., Jr., "Equilibrium points in n-person games", Proc. Nat. Acad. Sci. USA 36, 1950, pp. 48-49.

[16] Nash, John F., Jr., "Noncooperative games", Annals of Math. 54, 1951, pp. 286-295.

[17] von Neumann, John and Oskar Morgenstern, Theory of Games and Economic Behavior, Princeton Univ. Press, Princeton, 1953, 3rd ed.

[18] Rapoport, Anatol and A. Chammah, Prisoners' Dilemma, University of Michigan Press, Ann Arbor, 1966.

[19] Selten, Reinhard, "A simple model of imperfect competition, where 4 are few and 6 are many", Int. J. Game Theory 2, 1973, pp. 141-201.

[20] Shapley, Lloyd S., "A solution containing an arbitrary closed component", Contributions to the Theory of Games, Vol. IV, Princeton University Press, Princeton, 1959, pp. 87-93.

[21] Shapley, Lloyd S., "Notes on n-person games, VII: cores of convex games", RM-4571-PR, The RAND Corp., Santa Monica, 1965 (see also [23]).

[22] Shapley, Lloyd S., "Notes on n-person games, VIII: a game with infinitely 'flaky' solutions", unpublished paper.

[23] Shapley, Lloyd S., "Cores of convex games", Int. J. Game Theory 1, 1971, pp. 11-26.

[24] Shapley, Lloyd S., personal communication, 1973.

[25] Stearns, Richard E., "A game without side payments that has no solution", Report of the Fifth Conference on Game Theory (mimeographed), Princeton University, Princeton, 1965.

[26] Vickrey, William, "Self-policing properties of certain imputation sets", Contributions to the Theory of Games, Vol. IV, Princeton University Press, Princeton, 1959.