THE CORE OF A COOPERATIVE GAME WITHOUT SIDE PAYMENTS[1]

BY

ROBERT J. AUMANN

The *core* of an n-person game, though used already by von Neumann and Morgenstern [15], was first explicitly defined by Gillies [5]. Gillies's definition is restricted to cooperative games with side payments and unrestrictedly transferable utilities[2], but the basic idea is very simple and natural, and appears in many approaches to game theory. We consider a certain set of "outcomes" to a game, and define a relation of "dominance" (usually not transitive) on this set. The core is then defined to be the subset of outcomes maximal with respect to the dominance relation; in other words, the subset of outcomes from which there is no tendency to move away—the equilibrium states.

To turn this intuitive description of the core notion into a mathematical definition, we need precise characterizations of

(a) the kind of game-theoretic situation to which we are referring (cooperative game, noncooperative game, etc.);

(b) what we mean by "outcome"; and

(c) what we mean by "dominance."

Different ways of interpreting these three elements yield different applications of the generalized "core" notion, many of them well-known in game theory. Gillies's core, Luce's ψ-stability [10], Nash's equilibrium points [12], Nash's solution to the bargaining problem [13][3], and the idea of Pareto optimality—to mention only some of the applications—can all be obtained in this way.

Here we shall be concerned exclusively with cooperative games without side payments[4]. Our procedure will be to generalize von Neumann's fundamental notion of characteristic function to this case, and on the basis of this generalization to define the core in a way that generalizes and parallels the core in the classical theory—i.e., Gillies's core. The generalization of the characteristic function is of interest for its own sake also; for example, a theory of "solutions" has been developed that generalizes and parallels the classical theory of solutions and is based on the characteristic function [3; 16].

[2] Such games will be called *classical* games in the sequel, and the theory described in [5; 15] will be called the *classical* theory.

[3] Cf. [6].

[4] Classical games are known to be special cases of these games.

As in the classical theory, our "outcomes" will all be payoff vectors. We leave aside for the moment the question as to which particular set of payoff vectors we wish formally to consider as our set of outcomes. This brings us to the question of how to characterize the notion of "dominance."

Although formally it is simpler to define the characteristic function first and then to base on it the definition of dominance, the more intuitive procedure is the reverse: We must first state what we require from the dominance relation, and this will enable us to motivate our definition of characteristic function. Following the classical theory, then, we will say that a payoff vector x dominates another one y if

(i) there is a coalition S that prefers x to y, and

(ii) this preference is "not idle," i.e. S can actually achieve at least its portion of x.

What is meant by condition (i) is clear; each member of S must get more in x than in y. As for condition (ii), its exact meaning depends on how we wish to interpret the words "can actually achieve"; or to say the same thing in more technical language, it depends on when we wish to consider the coalition S "effective" for the payoff vector x.

In the sequel we will give a number of different definitions of effectiveness, each one leading to a different notion of dominance and hence to a different core. An alternative procedure is to assume that we already know for each coalition S which are the payoff vectors x for which S is effective; on the basis of this information we can then determine the core, without having to know the normal form of the game or the definition of effectiveness. A game presented in this form is said to be in *characteristic function* form. The characteristic function form of a game can always be calculated from its normal form and a particular definition of effectiveness. Note the similarity with the characteristic function of the classical theory; there there is associated with each coalition S a number $v(S)$, and the vectors x for which S is effective are precisely those for which[5] $\sum_{i \in S} x^i \leqq v(S)$. Here the set of x for which S is effective need not have such a simple form, and cannot be characterized by a single number; we therefore define $v(S)$ to be the set itself, rather than a number that characterizes the set. A considerable part of the theory can be developed on the basis of the characteristic function, without referring to the original game or to the particular notion of effectiveness we are using. As in the classical theory, some assumptions must be made about $v(S)$ to justify this development; these assumptions are natural ones, and we will establish that they hold for the particular definitions of effectiveness that we will wish to use.

We now return to the question of which payoff vectors we wish to consider as "outcomes." One possibility is the set H of all those payoff vectors that can be obtained by means of some correlated mixed strategy of the set

---

[5] The coordinates of the payoff vectors are indexed with superscripts.

N of all players. It is also possible to impose various more or less natural restrictions on the set of outcomes. There are two such restrictions that have received special attention in the literature, namely "individual rationality" and "group rationality." The former restricts the outcomes to payoff vectors in which each individual player gets at least what he can guarantee himself without any aid from the other players; under the latter restriction, a payoff vector is not called an "outcome" if there is another payoff vector in H which yields more to each player. These two restrictions can be imposed on the "outcome" concept in various combinations, so that we obtain four possibilities for this concept. In the classical theory it is easily established that all four lead to the same core; in the present theory this is also true, but the proof is no longer trivial. An interesting sidelight on this theorem is that its proof depends essentially on the assumption that H is a polyhedron (this assumption always holds if we start out with a finite game). If we replace H by a non-polyhedral convex set, the theorem becomes false; such a situation can actually be realized in the case of games with infinite strategy sets.

The paper is divided into two parts: the first part (§§1–7) deals with the theory of games in characteristic function form; the second part (§§8–10) deals with applications to games in normal form. §1 is devoted to a review of notation. In §2 we give the formal definition of a game in characteristic function form. §3 is devoted to the definition of various basic concepts such as domination, individual and group rationality, and core. §§4 and 5 are devoted to the statement and proof of the theorem that all the sets of outcomes discussed above lead to the same core. In §6 we give the counter-example to this theorem when H is not polyhedral. In §7 we discuss the composition of two games, and remark that the core of the composition is the cartesian product of the cores of the components. In §8 we pass to the normal form. We define two kinds of effectiveness, both generalizations of the classical definition, and show that they are different. In §9 we show that both these definitions lead to characteristic functions that satisfy the conditions of §2. In §10 we discuss the connection between the supergame[6] of a game and its various cores; in particular we shall show that the set of acceptable payoff vectors of a game $[1; 2]$ coincides with the core for one of the two definitions alluded to above.

1. **Notation.** N will denote a fixed finite set with n members, who will be called *players*. E^N will denote euclidean space of n dimensions, the coordinates of the points being indexed by the members of N; formally, E^N may be considered the set of functions from N to the reals. The points of E^N will be called payoff vectors. If $x \in E^N$, the coordinates of x will be denoted by x^i, where $i \in N$. For fixed $x \in E^N$ and $S \subset N$, we will call the S-tuple $\{x^i\}_{i \in S}$ an *S-vector* and denote it by x^S.[7] Note that $x = x^N$. If x^S and y^S are S-vectors, then any

---

[6] The game each play of which consists of an infinite sequence of the plays of the original game.

[7] x^S is the projection of x on E^S; if x is considered a function, then x^S is x restricted to S.

relation between x^S and y^S is to be understood coordinate-wise; e.g., $x^S \geqq y^S$ means $x^i \geqq y^i$ for all $i \in S$. If $S \subset N$ then (x^S, y^{N-S}) denotes the payoff vector z such that $z^S = x^S$ and $z^{N-S} = y^{N-S}$.

Subsets of N will be called coalitions, and will be denoted by S and T. Lower case latin letters towards the end of the alphabet will denote payoff vectors. $\varnothing$ denotes the empty set. In addition to its usual meaning, 0 will sometimes denote a vector all of whose components are 0; no confusion will result. The letter i always denotes a player. Unless the contrary is specifically indicated, summation, the taking of maxima or minima, etc., will be over i; for instance, $\sum_S$ means $\sum_{i \in S}$. The symbol $\times$ denotes the cartesian product.

We shall need a norm on E^N. Any norm with reasonable properties would serve our purposes; we shall use the maximum, defined by $\|x\| = \max_N |x^i|$. In addition to the usual norm properties, we note

(1)   if $x > 0$ and $y > 0$, then $\|x + y\| > \max(\|x\|, \|y\|)$.

Similar to the definition of norm on E^N, we define a norm on E^S by $\|x^S\| = \max_S |x^i|$.

The numbering of formulas, theorems, etc., starts from the beginning in each section; references from one section to another specify the section number as well as the formula number.

2. **The definition of a game in characteristic function form.**

DEFINITION. A *characteristic function*([8]) is a pair (N, v), where N is a finite set and v is a function that carries each subset S of N into a subset $v(S)$ of E^N so that

(1) $v(S)$ *is convex*;
(2) $v(S)$ *is closed*;
(3) $v(\varnothing) = E^N$;
(4) *if* $x \in v(S)$ *and* $y^S \leqq x^S$, *then* $y \in v(S)$;
(5) *if* $S \cap T = \varnothing$, *then* $v(S \cup T) \supset v(S) \cap v(T)$.

A *game in characteristic function form*, or simply a *game*, is a triple (N, v, H), where (N, v) is a characteristic function and

(6) *H is a convex compact polyhedral subset of* E^N.

Condition (5) is the natural generalization of the classical notion of super-additivity: it says that if a certain outcome can be achieved by the disjoint coalitions S and T when acting separately, then it can also be achieved by them when acting in concert.

We shall say that (N, v, H) is an *ordinary* game if

(7) $x \in v(N)$ *if and only if there is a* $y \in H$ *such that* $x \leqq y$.

This condition is easily justified intuitively, if we consider the interpretations of H and $v(N)$: H is the set of all payoff vectors that can be achieved by a joint strategy of all of N, whereas $v(N)$ is the set of payoff vectors x such that

---

([8]) Note the similarity with the "end games" used by Isbell [8] in a somewhat different context. (This work is independent of Isbell's.)

N can jointly achieve at least x. The notion of game as originally defined (without (7)) provides a generalization of von Neumann and Morgenstern's "extended" game [15][9]; this is why (7) was not included in the original definition[10].

3. **Domination, core, rationality.** Fix a game (N, v, H). A payoff vector x is said to *dominate* a payoff vector y *via* S (notation: $x \succ_S y$) if $x \in v(S)$ and $x^S > y^S$; x is said to *dominate* y (notation: $x \succ y$) if there is an S such that $x \succ_S y$. If R is an arbitrary set of payoff vectors, we define the R-*core* $\mathfrak{C}(R)$ to be the set of all members of R not dominated by any other member of R.

It is easy to show that for each $i \in N$, there is an extended real number[11] v^i such that $v(\{i\}) = \{x : x^i < v^i\}$. A payoff vector x is called *individually rational* if $x \geq v^N$. x is called *group rational* if there is no $y \in H$ such that $y > x$. We will denote by E the set of group rational payoff vectors in H, and by $\overline{A}$ the set of individually rational payoff vectors in H; also, we set[12] $A = E \cap \overline{A}$ and $\overline{E} = H$.
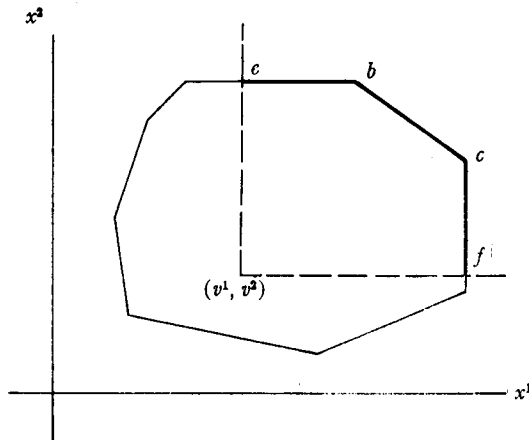
FIG. 1

We will consider the R-cores for $R = E$, $\overline{E}$, A, and $\overline{A}$. For two-person games, all these cores turn out to be equal to A. This is a set which is related to what has been called the "negotiation set" [11, p. 118], but is *not* always the same thing. (See Figure 1, in which the negotiation set is the line bc, whereas the set A is the broken line $ebcf$. Note that eb is horizontal and cf is vertical.)

---

[9] See [3, §7].

[10] Half of (7)—the "only if" half—was included in the definition of "game" as given in [3, §2]. What we call "game" here is called "extended game" in [3], and what we call "ordinary game" here is slightly stronger than either of the definitions in [3].

[11] A real number or $+\infty$ or $-\infty$.

[12] Following the notation of [5, p. 58].

4. A lemma on polyhedral sets. If $B \subset E^N$, denote by $I(B)$ the set[13] of all members x of B for which there is no y such that $y > x$. For example, $A = I(\overline{A})$ and $E = I(\overline{E})$. The lemma we shall establish in this section may be described as follows:

If B is a bounded polyhedron in E^N, then for each x in B but not in $I(B)$, there is a payoff vector x'' in $I(B)$ such that if we move along the ray connecting x to x'' at a constant speed (which is the same for all $x \in B$), then the rate of increase of each of the coordinates is uniformly bounded away from zero (for all $x \in B$).

If B is not a polyhedron this need not be true, as we shall see in an example.

For $x > 0$, define $f(x) = \max_{i, j \in N} (x^i / x^j)$. We have

$$(1) \qquad\qquad f(x + y) \leqq \max(f(x), f(y)).$$

LEMMA 2. *For every closed polyhedron B in E^N, there is a positive number K such that for all $x \in B - I(B)$, there is an $x' \in B$ such that*

$$(3) \qquad\qquad x' > x \quad and \quad f(x' - x) \leqq K.$$

Proof. Suppose B to be defined by the set of linear inequalities $L_1(x) \geqq b_1, \cdots, L_m(x) \geqq b_m$; we denote this set by M. Each subset Q of M defines a subset B_Q of B, namely the set of those elements of B which satisfy the inequalities in Q strictly, and the inequalities in $M - Q$ as equalities. Some of the B_Q may be empty; but those that are not are distinct, and we have $B = \bigcup_{Q \subset M} B_Q$. (Geometrically, the B_Q are the interiors of the faces of B.)

For each Q such that $B_Q - I(B) \neq \varnothing$, choose a payoff vector x_Q in $B_Q - I(B)$. Then there is a payoff vector $y_Q \in B$ such that $y_Q > x_Q$. Now let x be an arbitrary element of B_Q; define $y_\delta = x + \delta(y_Q - x_Q)$. For sufficiently small positive δ, y_δ satisfies the inequalities in Q strictly; the inequalities in $M - Q$ are satisfied by y_δ for *all* positive δ. Hence for positive δ sufficiently small, $y_\delta \in B$ and $y_\delta > x$; we define $x' = y_\delta$ for this δ. Then $x' - x = \delta(y_Q - x_Q)$, so that $f(x' - x) = f(y_Q - x_Q)$. Setting $K = \max_{Q \subset M} f(y_Q - x_Q)$, we obtain $f(x' - x) \leqq K$ for all those x that are located in some B_Q for which $B_Q - I(B) \neq \varnothing$. But since every $x \in B - I(B)$ is located in some such B_Q, our proof is complete.

LEMMA 4. *For every compact polyhedron B in E^N, there is a positive number K such that for all $x \in B - I(B)$, there is an $x'' \in I(B)$ such that $x'' > x$ and $f(x'' - x) \leqq K$.*

Proof. For each $x \in B - I(B)$, let F_x be the union of the single point x with the set of all $x' \in B$ satisfying (3). F_x is compact, and therefore the function $\|y - x\|$, considered as a function of y, attains its maximum in F_x, say

---

at the point x''. Suppose $x'' \notin I(B)$. Then by Lemma 2 there is a point $y \in B$ such that $y > x''$ and $f(y - x'') \leq K$. From (1) it then follows that $f(y - x) \leq K$ and from (1.1) that $\|y - x\| = \|(y - x'') + (x'' - x)\| > \|x'' - x\|$; hence x'' does not have the maximum property by which it was defined, which is a contradiction. This completes the proof.

COROLLARY 5. *If B is a compact polyhedron in E^N, then there is a positive number K, such that for each $x \in B - I(B)$, there is an $x'' \in I(B)$ such that $x'' > x$ and for each $i \in N$, $x''^i - x^i \geq \|x'' - x\|/K$.*

5. **Relations between the R-cores for $R = E$, $\overline{E}$, A, and $\overline{A}$.** Fix a game (N, v, H). The crux of this section is the following theorem:

THEOREM 1. *Let B be a compact polyhedron in E^N, and let $y \in I(B)$. If there is a $z \in B$ which dominates y, then there is also a $w \in I(B)$ which dominates y.*

Proof. We may assume without loss of generality that $y = 0$. Let V denote the closed positive orthant $\{x : x \geq 0\}$. Since $0 = y \in I(B)$, B cannot intersect the interior of V, and there is therefore a hyperplane $g(x) = \sum_N c^i x^i = 0$ which separates B from V. W. l. o. g. $g(x) \leq 0$ for all $x \in B$, and $g(x) \geq 0$ for all $x \in V$; from the latter fact it follows that $c^N \geq 0$. Note that if $x \in B$ and $g(x) \geq 0$, then $x \in I(B)$; otherwise we would have an $x_1 \in B$ such that $x_1 > x$, and since not all the c^i vanish, it would follow that $g(x) < g(x_1) \leq 0$.

Let the effective set for the domination of z over 0 be S; set $h(x) = \sum_{N-S} c^i x^i$. Suppose there is an $x \in B$ such that $x >_S 0$ and $h(x) \geq 0$. Then since $x^S > 0$ and $c^S \geq 0$ it follows that $\sum_S c^i x^i \geq 0$. Hence $g(x) \geq 0$, and therefore $x \in I(B)$; but then we are finished (set $w = x$). Therefore we may assume without loss of generality that

(2)   if $x \in B$ and $x >_S 0$, then $h(x) < 0$.

Let $k = (\min_S z^i)/2$; note that $k > 0$. Let C be the set of those x in B for which $x^S \geq 0$ and $\|x^S\| = k$. C is compact, and therefore $h(x)$ attains its maximum in C at a point x_1 in C. If $x_1 \notin I(B)$, then there is an $x_2 \in B$ for which $x_2 > x_1$; hence $\|x_2^S\| > \|x_1^S\| = k$, and

(3) $h(x_2) \geq h(x_1)$.

Set $x_3 = (k/\|x_2^S\|)x_2$. Then $\|x_3^S\| = k$, $x_3^S > 0$, and since B is convex, $x_3 \in B$. Hence $x_3 \in C$, and therefore

(4) $h(x_1) \geq h(x_3)$.

Since $\|x_3^S\| = k$, it follows that $x_3^S < z^S$; but since $z \in v(S)$, it follows from (2.4) that $x_3 \in v(S)$. Hence from $x_3^S > 0$, it follows that $x_3 >_S 0$. Hence by (2), $h(x_3) < 0$. But $h(x_3) = (k/\|x_2^S\|)h(x_2)$, and $k/\|x_2^S\| < 1$; hence $h(x_2) < h(x_3)$, which contradicts (3) and (4). We conclude that $x_1 \in I(B)$.

Now since $\|x_1^S\| = k$, it follows that $x_1^S < z^S$; therefore $x_1 \in v(S)$. If $x_1^S > 0$ it then follows that $x_1 \succ_S 0$, and since $x_1 \in I(B)$, we are finished ($w = x_1$). It therefore remains only to deal with the case in which one of the coordinates of x_1^S vanishes. In this case, set $\delta = k/(K+1)$, the K being that of Corollary 4.5. Let $x_4 \in B$ be such that $x_4^S > 0$ and $\|x_4 - x_1\| \leq \delta$; such an x_4 can be constructed by choosing a point sufficiently close to x_1 on the line segment joining z to x_1. Define x_4'' in accordance with Corollary 4.5. If $\|x_4'' - x_4\| > K\delta$, then $x_4''^i - x_4^i > \delta$ for each $i \in N$; therefore for $i \in N$ we have

$$x_4''^i - x_1^i = (x_4''^i - x_4^i) + (x_4^i - x_1^i) > \delta - |x_4^i - x_1^i| \geq \delta - \|x_4 - x_1\| \geq 0.$$

Hence $x_4'' > x_1$ and $x_4'' \in I(B)$, contradicting $x_1 \in I(B)$. Hence $\|x_4'' - x_4\| \leq K\delta$. But then

$$\|x_4'' - x_1\| \leq \|x_4'' - x_4\| + \|x_4 - x_1\| \leq K\delta + \delta = k;$$

hence

$$\|x_4''^S\| \leq \|x_4'' - x_1\| + \|x_1^S\| \leq 2k = \min_S z^i.$$

Hence $x_4''^S \leq z^S$ and from (2.4) we deduce $x_4''^S \in v(S)$. Since $x_4''^S > x_1^S \geq 0$, we obtain $x_4'' \succ_S 0$ and $x_4'' \in I(B)$. The proof of Theorem 1 is now complete ($w = x_4''$).

COROLLARY 5. *If B is a compact polyhedron, then $\mathcal{C}(I(B)) = \mathcal{C}(B) \cap I(B)$.*

Proof. If $y \in \mathcal{C}(I(B))$, then surely $y \in I(B)$. If y were not in $\mathcal{C}(B)$, then it would be dominated by a member z of B, and hence by Theorem 1, by a member w of $I(B)$; but then it would not be in $\mathcal{C}(I(B))$. Hence $y \in \mathcal{C}(B)$ also.

Conversely, if $y \in I(B)$ and is not dominated by any member of B, then a fortiori it is not dominated by any member of $I(B)$. Hence $y \in \mathcal{C}(I(B))$.

COROLLARY 6. $\mathcal{C}(E) = \mathcal{C}(\overline{E}) \cap E$; $\mathcal{C}(A) = \mathcal{C}(\overline{A}) \cap A$.

COROLLARY 7. *If (N, v, H) is an ordinary game, then $\mathcal{C}(E) = \mathcal{C}(\overline{E})$ and $\mathcal{C}(A) = \mathcal{C}(\overline{A})$.*

Proof. We need only remark that if $B \subset v(N)$ then $\mathcal{C}(B) \subset I(B)$; for any payoff vector not in $I(B)$ is dominated via N by some other payoff vector. Our result now follows by applying Corollary 5 with $B = \overline{E}$ or $\overline{A}$.

THEOREM 8. *If (N, v, H) is an ordinary game, then $\mathcal{C}(E) = \mathcal{C}(\overline{E}) = \mathcal{C}(A) = \mathcal{C}(\overline{A})$.*

Proof. It is sufficient to prove that $\mathcal{C}(\overline{E}) = \mathcal{C}(\overline{A})$. Clearly $v^N \in \bigcap_N v(\{i\})$; hence by (2.5), $v^N \in v(N)$. Hence by (2.7) there is a $y \in H$ such that $y \geq v^N$. Now let $x \in \mathcal{C}(\overline{E})$. If $x \notin \overline{A}$, then for some i, $x^i < v^i \leq y^i$. Let y_1 be on the line segment connecting y to x, but so close to x so that $v^i \geq y_1^i > x^i$. Since $\overline{E}$ is convex and both x and y are in $\overline{E}$ (which is the same as H), so is y_1; but then

$y_1 \succ_{\{ij\}} x$, and therefore $x \notin \mathcal{C}(\overline{E})$, a contradiction. Therefore $x \in \overline{A}$. Therefore if $x \notin \mathcal{C}(\overline{A})$ then there is a $z \in \overline{A}$ such that $z \succ x$; but since $\overline{A} \subset \overline{E}$, it follows that $z \in \overline{E}$, and therefore $x \notin \mathcal{C}(\overline{E})$, again a contradiction. We have shown that $\mathcal{C}(\overline{E}) \subset \mathcal{C}(\overline{A})$.

Conversely if $x \in \mathcal{C}(\overline{A})$, then surely $x \in \overline{E}$, since $\overline{E} \supset \overline{A} \supset \mathcal{C}(\overline{A})$. Hence if $x \notin \mathcal{C}(\overline{E})$, there must be a $y \in \overline{E}$ and an $S \subset N$ such that $y \succ_S x$. In particular, $y^S > x^S$ and $y \in v(S)$. Set $z = (y^S, v^{N-S})$; then by (2.4), $z \in v(S)$, and therefore $z \in v(S) \cap \bigcap_{N-S} v(\{i\})$. Hence by (2.5), $z \in v(N)$. Hence by (2.7) there is a $w \in H$ such that $w \geq z$. In particular, $w^S \geq z^S = y^S > x^S \geq v^S$ (since $x \in \overline{A}$), and $w^{N-S} \geq z^{N-S} = v^{N-S}$; hence $w \in \overline{A}$. Now let w_1 be on the line segment connecting w to x, but so close to x so that $y^S > w_1^S > x^S$. Since $\overline{A}$ is convex and both x and w are in $\overline{A}$, so is w_1; but since $y \in v(S)$, it follows from (2.4) that $w_1 \in v(S)$. Since $w_1^S > x^S$ and $w_1 \in v(S)$, it follows that $w_1 \succ_S x$; therefore since $w_1 \in \overline{A}$, it follows that $x \notin \mathcal{C}(\overline{A})$, a contradiction. Hence $x \in \mathcal{C}(\overline{E})$, and the proof is complete.

If G is an ordinary game, we shall call the common value of $\mathcal{C}(\overline{E})$, $\mathcal{C}(E)$, $\mathcal{C}(\overline{A})$, and $\mathcal{C}(A)$ the *core* of G.

6. A counter-example. The results of §§4 and 5 may fail if B (or H) is not polyhedral. In the case of Lemma 4.2 a circle in two dimensions is a counter-example. In the case of Theorem 5.1, let $N = \{1, 2, 3\}$ and let B be the convex hull of the sets C and D, where

$$C = \{x: x^1 \geq 0, \, x^2 \geq 0, \, x^3 = 0, \, (x^1)^2 + (x^2)^2 \leq 1\},$$

$$D = \{x: x^1 \geq 0, \, x^2 \geq 0, \, x^3 = 1, \, (x^1)^2 + (x^2 + 1)^2 \leq 4\}.$$

Then

$$I(B) = D \cup \{x: x^1 = 0, \, x^2 = 1, \, 0 \leq x^3 \leq 1\}.$$

Define the characteristic function v by

$$v^N = 0, \qquad v(\{ij\}) = \{x: x^i \leq 1/2 \text{ and } x^j \leq 1/2\},$$

$$v(N) = \{x: \text{there is a } y \in H \text{ such that } y \geq x\}.$$

We have $(1/2, 1/2, 1/2) \in B$ and $(1/2, 1/2, 1/2) \succ_{\{13\}} (0, 1, 0)$, but there is no member of $I(B)$ that dominates $(0, 1, 0)$. If we set $H = B$ we obtain counter-examples to the other results of §5.

7. Composition. Let $G_1 = (N_1, v_1, H_1)$ and $G_2 = (N_2, v_2, H_2)$ be games whose player sets N_1 and N_2 are disjoint. Intuitively, the *composition* G of G_1 and G_2 is the game each play of which consists of a play of G_1 and a play of G_2, played without any interconnection. Formally, we define $G = (N, v, H)$, where $N = N_1 \cup N_2$, $H = H_1 \times H_2$, and for each $S \subset N$, $v(S) = v_1(S \cap N_1) \times v_2(S \cap N_2)$.

Let $R_1 \subset E^{N_1}$ and $R_2 \subset E^{N_2}$, and set $R = R_1 \times R_2$. Then it is easily seen that $\mathcal{C}(R) = \mathcal{C}_1(R_1) \times \mathcal{C}_2(R_2)$. Furthermore, if G_1 and G_2 are ordinary, then so is G.

It follows that in this case, *the core of G is the cartesian product of the cores of G_1 and G_2.*

8. α-effectiveness and β-effectiveness. Up to now we have been treating games in characteristic function form only; we now turn to games given in normal form, and ask how we may obtain the characteristic function form from the normal form. As we remarked in the introduction, this may be done in a number of ways, depending on our definition of effectiveness. Here we shall give two such definitions.

A (finite) game Γ in normal form consists of a finite set N, called the set of *players*, a finite set P^i for each $i \in N$, called the set of *pure strategies* for player i, and a function F from the cartesian product P of all the P^i to E^N; F is called the *payoff* function([14]), and its ith coordinate F^i is the *payoff to i*. If $S \subset N$, we write $P^S = \prod_S P^i$, the cartesian product being meant. A probability measure on P^S will be called a *c-strategy S-vector* (c for correlated); the set of all c-strategy S-vectors will be denoted C^S. Note that a c-strategy $\{i\}$-vector is the same as a mixed strategy for player i. If $c^N \in C^N$, then $F(c^N)$ will denote the expected payoff if the c-strategy N-vector c^N is played. If S, $T \subset N$, $S \cap T = \varnothing$, then $(c^S \times c^T)$ denotes the product measure([15]) on $P^{S \cup T} = P^S \times P^T$ induced by c^S and c^T. Occasionally we shall have cause to consider a topological and a convex structure on C^S; in this case C^S will be considered a subset of E^{P^S}.

DEFINITION. (1) A coalition S is said to be *α-effective* for the payoff vector x if there is a $c^S \in C^S$ such that for each $c^{N-S} \in C^{N-S}$, we have $F^S(c^S \times c^{N-S}) \geqq x^S$.

(2) S is said to be *β-effective* for x if for each $c^{N-S} \in C^{N-S}$ there is a $c^S \in C^S$ such that $F^S(c^S \times c^{N-S}) \geqq x^S$.

Intuitively, α-effectiveness means that S can assure itself, independently of the actions of $N-S$, that each of its members i will receive at least his coordinate x^i of x. β-effectiveness means that S can always act so that each of its members i receives at least x^i, but the strategy that it must use to achieve this end may depend on the strategy used by $N-S$; in other words, $N-S$ cannot effectively prevent S from obtaining at least([16]) x^S. Although α-effectiveness seems at first to be the intuitively more straightforward concept, technically speaking β-effectiveness possesses certain interesting properties not shared by α-effectiveness (see §10) which lead one to think that it may eventually turn out to be the more significant concept.

To construct a game in which α-effectiveness and β-effectiveness are not the same, let $N = \{1, 2, 3\}$ and $S = \{1, 2\}$. Let P^S have two members p_1^S and p_2^S, and P^3 two members p_1^3 and p_2^3. Define F^S by the matrix

---

([14]) Denoted by H in $[1; 2]$.

([15]) Denoted (c^S, c^T) in $[1]$.

([16]) The difference between the two kinds of effectiveness may be formulated as the difference between a maxmin and a minmax; for 2-person games it follows from the von Neumann theorem that the two concepts coincide, but this does not generalize to more players.

	p_1^3	p_2^3
p_1^S	1, −1	0, 0
p_2^S	0, 0	−1, 1

;

the values of F^3 need not concern us. Then for $(0, 0, 0)$, S is β-effective but not α-effective.

9. **Passage from the normal form to the characteristic function form.** Fix a game Γ in normal form, and for each $S \subset N$, let $v_\alpha(S)$ be the set of payoff vectors for which S is α-effective. Define $v_\beta(S)$ similarly, using β-effectiveness instead of α-effectiveness. Define $H = F(C^N)$; H is the convex hull of all the payoff vectors of the form $F(p)$, where $p \in P$. Both (N, v_α, H) and (N, v_β, H) are ordinary games (though they may be different, as we saw in the previous section); except for Condition 2.5 in the case of β-effectiveness, all the conditions of §2 are easily verified for both these games. To establish Condition 2.5 for (N, v_β, H), let $x \in v_\beta(S) \cap v_\beta(T)$ and $c^{N-S-T} \in C^{N-S-T}$. Define subsets U and V of $C^S \times C^T$ as follows:

$$U = \left\{ (c^S, c^T) : F^T(c^S \times c^T \times c^{N-S-T}) \geq x^T \right\},$$
$$V = \left\{ (c^S, c^T) : F^S(c^S \times c^T \times c^{N-S-T}) \geq x^S \right\}.$$

Applying the von Neumann-Kakutani fixed point theorem[17], we obtain the existence of a point (c_0^S, c_0^T) in $U \cap V$. Setting $c_0^{S \cup T} = c_0^S \times c_0^T$, we obtain $F^{S \cup T}(c_0^{S \cup T} \times c^{N-S-T}) \geq x^{S \cup T}$, and it follows that $x \in v_\beta(S \cup T)$.

From (2.6) and (2.7) it follows that in an ordinary game, $v(N)$ must be polyhedral. The reader may suspect that in the characteristic function form of a finite game in normal form, $v(S)$ must be polyhedral for all S. This is true for $v_\alpha(S)$, but not for $v_\beta(S)$. The example is the same as in the previous section, except that $F^1(p_1^S, p_1^3) = F^2(p_2^S, p_2^3) = 0$ rather than 1. (See Figure 2; $v_\beta(S)$ is the cylinder whose cross-section is the shaded area.)

Note that we always have $v_\alpha(N) = v_\beta(N)$ and $v_\alpha^N = v_\beta^N$ (the former is trivial, the latter follows from the minimax theorem for 2-person zero-sum games [15]). In particular, α- and β-effectiveness are equivalent for all 2-person games.

If Γ_1 and Γ_2 are games in normal form with disjoint player sets, we may define their *composition* Γ by $N = N_1 \cup N_2$, $P = P_1 \times P_2$, $F(p_1, p_2) = (F_1(p_1), F_2(p_2))$. It is easily established that either definition of effectiveness yields a characteristic function form for Γ that is the composition of the corresponding characteristic function forms of Γ_1 and Γ_2 in the sense of §7.

10. **The supergame.** Nash's notion of equilibrium point for noncooperative games [12] is an example of the core notion as described in the introduc-

_____

[17] [9, Theorem 2]; see also [14].

tion. The "outcomes" are strategy n-tuples; one strategy n-tuple f "dominates" another one g if they coincide for all but one of the players, and that one player prefers([18]) f to g. Thus an equilibrium point is a strategy n-tuple f with the property that if all the players have reason to believe that f will be played, then no player will be tempted to deviate from f. In the context of cooperative games, it is natural to broaden the definition of dominance so that f dominates g whenever they coincide for all players not in a certain coalition S, and the players in S each prefer f to g. When dominance between strategy n-tuples is defined in this way, members of the core are called *strong* equilibrium points([19]).

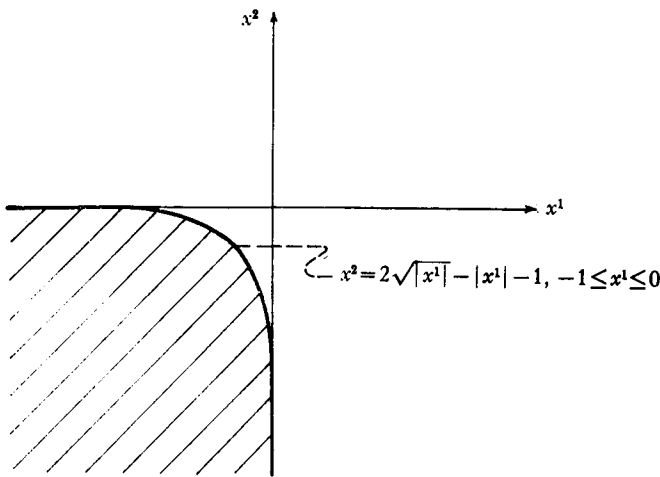$$x^2 = 2\sqrt{|x^1|} - |x^1| - 1, \quad -1 \leq x^1 \leq 0$$

Fig. 2

This definition of dominance seems well justified for a single play of a game which is not to be repeated. If the game is to be played repeatedly, though, then a player or group of players may be unwilling to deviate even if the deviation *will* yield a temporary advantage, for fear of future retaliation. If future retaliation is to be ruled out, then S must be able to *maintain* its payoff at the level of f; that is, S must be effective for f. We are thus led to the conclusion that a strategy n-tuple for one of a long sequence of plays of a game Γ should be considered in equilibrium if its payoff is in the core of the characteristic function form of Γ.

The question now arises, is it the α-core or the β-core that is appropriate for use in this context, or possibly we should use an altogether different notion of effectiveness? To answer this question, we consider a long sequence of plays of Γ as a single play of a game Γ^*, which we call the *supergame* of Γ [1; 11]. It stands to reason that equilibrium behavior for Γ, knowing that

([18]) I.e., receives a higher payoff when g is played.
([19]) A related definition is given in [4].

there will be more plays of Γ in the future, should correspond to equilibrium behavior in Γ^*, provided that Γ^* is not repeated. But for games that are not repeated, we have a perfectly well-defined equilibrium notion, namely that of strong equilibrium point. And it turns out that it is precisely the β-core of the characteristic function form of Γ that corresponds to the strong equilibrium points in Γ^*.

Formally[20], the supergame Γ^* is the game each play of which is an infinite sequence of plays of Γ. As in finite games, an n-tuple f of supergame strategies is said to *dominate* another n-tuple g if they coincide for all players not in a certain coalition S, and the players in S each prefer f to g. To define the word "prefer" in this context, we consider a sequence $\Gamma_1, \Gamma_2, \cdots$ of plays of Γ, and look at the average payoff for all the plays up to the kth. A number of definitions of preference are now possible, of which the following are the two "extreme" possibilities:

(a) S prefers f to g if the probability is positive that infinitely often the average payoff to each member of S will be uniformly[21] larger if f is used than if g is used.

(b) S prefers f to g if it is certain[22] that from a certain play Γ_k onwards, the average payoff to each member of S will always be uniformly larger if f is used than if g is used.

An n-tuple f of supergame strategies is said to have the payoff x if with probability 1 the average payoffs[23] tend to x. f is said to *correspond*[24] to a c-strategy vector c in Γ if the payoff to f exists and is the same as the payoff to c in Γ. An n-tuple of supergame strategies is said to be a *strong equilibrium point* if it is undominated and if it possesses a payoff. Actually we get two sets of strong equilibrium points, one for each of the two (inequivalent) notions of preference defined above. However, it turns out that both these sets correspond to the same set of c-strategy vectors, called *acceptable* points [1, §4]. By making use of Lemma 9.1 of [1], it is not difficult to show that the set of payoff vectors to acceptable points—the set of acceptable payoff vectors—coincides with the β-core[25]. Hence the β-core of a finite

---

[20] For a more detailed treatment of the supergame, see [1].

[21] The difference must be larger than a fixed (independent of k) positive S-vector.

[22] The probability is 1.

[23] Care should be taken to differentiate between the average payoff for the first k plays, and the expected payoff (for the latter, probability statements would of course be meaningless). Because of the law of large numbers, the existence of a payoff to f is quite plausible. For example, if a "steady state" in strategy choices on the individual Γ_k is ever reached, then f certainly has a payoff. As we have shown in [1, §12], considerations of expected payoff are inappropriate for Γ^*.

[24] The correspondence must be defined via the payoffs because there is no direct method for comparing strategies for individual games with supergame strategies.

[25] Definitions "between" (a) and (b) yield the same set of payoff vectors. For example, in either of the definitions we could substitute "with probability at least $1/2$" for the respective probability statements.

game is the set of payoff vectors to strong equilibrium points in its super-game.

REFERENCES

1. R. J. Aumann, *Acceptable points in general cooperative n-person games*, in [17, pp. 287–324].

2. ———, *Acceptable points in games of perfect information*, Pacific J. Math. vol. 10 (1960) pp. 381–417.

3. R. J. Aumann and B. Peleg, *Von Neumann—Morgenstern solutions to cooperative games without side payments*, Bull. Amer. Math. Soc. vol. 66 (1960) pp. 173–179.

4. R. Farquharson, *Sur une généralisation de la notion d'equilibrium*, C. R. Acad. Sci. Paris vol. 240 (1955) pp. 46–48.

5. D. B. Gillies, *Solutions to general non-zero-sum games*, in [17], pp. 47–83.

6. J. C. Harsanyi, *Approaches to the bargaining problem before and after the theory of games: a critical discussion of Zeuthen's, Hicks', and Nash's theories*, Econometrica vol. 24 (1956) pp. 144–157.

7. J. R. Isbell, *Absolute games*, in [17], pp. 357–396.

8. ———, *A modification of Harsanyi's bargaining model*, Bull. Amer. Math. Soc. vol. 66 (1960) pp. 70–73.

9. S. Kakutani, *A generalization of Brouwer's fixed point theorem*, Duke Math. J. vol. 8 (1941) pp. 457–459.

10. R. D. Luce, *A definition of stability for n-person games*, Ann. of Math. vol. 59 (1954) pp. 357–366.

11. R. D. Luce and H. Raiffa, *Games and decisions*, New York, John Wiley, 1957.

12. J. F. Nash, *Non-cooperative games*, Ann. of Math. vol. 54 (1951) pp. 286–295.

13. ———, *The bargaining problem*, Econometrica vol. 18 (1950) pp. 155–162.

14. J. von Neumann, *Über ein ökonomisches Gleichungssystem und eine Verallgemeinerung des Brouwerschen Fixpunktsatzes*, Ergebnisse eines Math. Kolloquiums vol. 8 (1937) pp. 73–83.

15. J. von Neumann and O. Morgenstern, *Theory of games and economic behavior*, Princeton, N. J., Princeton University Press, 1944, 2nd ed., 1947.

16. B. Peleg, *Solutions to cooperative games without side payments*, to be published.

17. A. W. Tucker and R. D. Luce, editors, *Contributions to the theory of games*. IV, Annals of Mathematics Studies, no. 40, Princeton, N. J., Princeton University Press, 1959.

THE HEBREW UNIVERSITY,
    JERUSALEM, ISRAEL