J. Soc. Indust. Appl. Math.
Vol. 9, No. 4, December, 1961
Printed in U.S.A.

ALMOST STRICTLY COMPETITIVE GAMES[1]

ROBERT J. AUMANN[2]

Introduction. Strictly competitive games (i.e., games equivalent to 2-person 0-sum games) can always be solved; it is always clear what each player should do, and the outcome is strictly determined. As is well known, this ceases to be true when we drop the assumption of strict competition. However, there are certain two-person games which still retain this (somewhat vague) property, although they are not strictly competitive. Examples are games that are obtained from strictly competitive games by the addition of strictly dominated strategies (e.g., the prisoner's dilemma), and certain games of perfect information, which we can "solve" by working our way backward from the final move. We wish to give a characterization of such games.

Roughly, we want to examine a class of games—we will call them *almost strictly competitive* (a.s.c.)—in which it is possible to define a unique value and a set of "good" strategies for each player, so that a pair of good strategies yields the value. We also want a structural theorem of the following kind: Suppose a game G in extensive form decomposes at a move X (cf. [1]) and that the subgame G_X is a.s.c.; suppose further that we define the difference game G_D by stipulating that the payoff at the terminal X of G_D is the value of G_X, and that we assume G_D to be a.s.c., as well. Then it should follow that the original game G is a.s.c., and that the composition of "good" strategies in G_D and G_X yields a "good" strategy in G. Such a theorem is very important in applications which involve complicated games in extensive form; it is also not unreasonable as a theoretical demand.

1. The definition. The definition of strict competition is that for each player, helping himself and hurting his opponent are equivalent. Our basic idea is to "weaken" this condition while retaining its spirit.

Recall the definition of an *equilibrium point* [3]: it is a pair of strategies at which neither player can increase his payoff by a unilateral change in strategy.[3] Let us now define a *twisted equilibrium point* to be a pair of strategies at which neither player can decrease the other player's payoff by a unilateral change in strategy. Twisted equilibrium points are the

[1] Received by the editors January 7, 1961, and in revised form March 15, 1961. The research described in this paper was supported partially by Carnegie Corporation of New York, and partially by Mathematica, Inc.

[2] The Hebrew University, Jerusalem, Israel, and Princeton University.

[3] Unless otherwise specified, the word "strategy" will be used throughout the paper in the sense of mixed strategy (which may in particular be pure).

same as equilibrium points, except that now the object of each player is to hurt his opponent, rather than to help himself. We will call a game *almost strictly competitive* if (i) these two notions are equivalent from the point of view of their outcomes, i.e., if the set of payoff vectors to equilibrium points is equal to the set of payoff vectors to twisted equilibrium points; and (ii) if the set of equilibrium points and the set of twisted equilibrium points intersect.

An alternative way of looking at a.s.c. games is the following: For a given game G, define the twisted game G^* to be the same as G except that the object of each player is to minimize his opponent's payoff rather than to maximize his own; formally, the strategy spaces in G^* are the same as those in G, and if the payoff to a given strategy pair in G is (h_1, h_2), then the payoff to the same pair in G^* is $(-h_2, -h_1)$. Then the twisted equilibrium points of G are precisely the equilibrium points of G^*, and G is a.s.c. if and only if (i) the equilibrium payoffs of G^* can be obtained from those of G by "twisting" them (i.e., substituting $(-h_2, -h_1)$ for (h_1, h_2)); and (ii) G and G^* have some equilibrium points in common.

The prisoner's dilemma

4, 4	0, 5
5, 0	1, 1

is an example of an a.s.c. game.

2. Theorems.

THEOREM A. *In an a.s.c. game, there is a unique equilibrium payoff* (v_1, v_2); *a fortiori this is also the unique twisted equilibrium payoff.*

The unique equilibrium payoff will be called the *value*. We will also refer to a particular player's value or the value of the game to him; this is simply his component of the value.

THEOREM B. *In an a.s.c. game, each player has a strategy which simultaneously guarantees that* (a) *he will obtain at least his own value*; *and* (b) *the other player will obtain at most the other player's value.*

Such a strategy will be called *good.*

THEOREM C. *In an a.s.c. game, a pair of good strategies is both an ordinary and a twisted equilibrium point*; *and conversely, any point which is both an ordinary and a twisted equilibrium point is a pair of good strategies.*

This is the *interchangeability* property for points that are both ordinary and twisted equilibrium points.

THEOREM D. *Let G be a 2-person extensive game which decomposes at a move X, and let G_X be a.s.c. Let G_D be the difference game, where the payoff to G_D at X is the value of G_X ; assume that G_D is a.s.c. as well. Then G is a.s.c.,*

$v(G) = v(G_D)$, and the composition of good strategies in G_X and G_D yields a good strategy in G.

This is the theorem that enables us to build up complicated a.s.c. games from simple ones, and to "solve" the complicated games.

We mention that the question of deciding whether a given game is a.s.c. or not can be answered by using the algorithm of Vorobyev [6], which gives a method for calculating all the equilibrium points of a given two-person game.

3. Discussion.

(a). Theorem B is the analogue of the minimax theorem; it may be restated in saddle-point form as follows: There is a pair of strategies (s_1^0, s_2^0), such that for all s_1, s_2, we have the two sets of inequalities,

$$\begin{aligned} h_1(s_1, s_2^0) &\leqq h_1(s_1^0, s_2^0) \leqq h_1(s_1^0, s_2), \\ h_2(s_1, s_2^0) &\geqq h_2(s_1^0, s_2^0) \geqq h_2(s_1^0, s_2), \end{aligned} \tag{3.1}$$

where h_1 and h_2 are the payoff functions. A consequence is the following minimax statement:

$$\min_{s_2} \max_{s_1} h_1(s_1, s_2) = \max_{s_1} \min_{s_2} h_1(s_1, s_2) = v_1,$$

and

$$\min_{s_1} \max_{s_2} h_2(s_1, s_2) = \max_{s_2} \min_{s_1} h_1(s_1, s_2) = v_2.$$

The minimax statement is not as strong as the saddle-point statement, because it does not provide for the existence of a strategy for (say) player 1 that *simultaneously* guarantees that he will obtain v_1 and that player 2 will not obtain more than v_2. Neither statement assures almost strict competitiveness, as is demonstrated by the example

1, 1	0, 0
0, 0	0, 0

.

(b). The definition of almost strict competitiveness may be applied without change to games with infinite pure strategy sets. Theorems A, B, and C also remain true in this context; the proofs go through without change. As far as Theorem D is concerned, in principle there is nothing to prevent this, too, from applying to the broader context of infinite games. However, both the statement and the proof of Theorem D depend on extensive game theory, and the elements of this theory of which we make use are available only for the finite case. Presumably an appropriate theory for at least some infinite games can be developed without too much difficulty, but we do not propose to do this here.

(c). Although the theory of a.s.c. games is basically noncooperative in its viewpoint, it has some applications to bargaining models for two-person cooperative games. The chief work in this area is that of Nash [2, 4] and Raiffa [5]. We will not attempt here to summarize it, but merely recall some of the terminology. Given a two-person game C to be treated cooperatively, Nash [4] breaks it up into a *threat game* and a *demand game*. The pure strategies of the threat game are threats to use mixed strategies of the original game C; the payoffs for the threat game depend on how the demand game is treated. Nash shows that if the demand game is treated by the method of [2], then the threat game has a saddle point in *pure strategies* (in the sense of (3.1)). Raiffa [5, p. 372] generalizes this by showing that the same result holds if any one of a large class of different schemes—including Nash's—is used for solving the demand game (we shall call these *demand schemes*). The solution of the threat game will depend on which demand scheme is used, but it always yields a payoff that is Pareto optimal [8] in the prospect space of C.

Our first remark is that the threat game is always a.s.c., no matter which demand scheme is used. This is a special case of the following theorem: Let G be a game which is "strictly competitive in pure strategies," i.e., if h and h^* are the payoffs to two pairs of pure strategies, and if $h_1 > h_1^*$, then $h_2 < h_2^*$. Assume further that one of the payoff functions (and hence also the other) has a pure strategy saddle-point. Then G is a.s.c. The proof is easily given. To apply this theorem to the threat game we need only verify that the threat game is "strictly competitive in pure strategies." This follows from the fact that the outcomes are all Pareto optimal.

Our second remark is of a completely different nature. Optimal behavior in the threat game usually depends strongly on the demand scheme being used. It is of great interest to know under what conditions this optimal behavior will be independent of the demand scheme, since cooperative games must often be played without a clear and fixed formal notion of which demand scheme is being used. A sufficient condition is that the original game C be a.s.c. This condition is not necessary; the necessary and sufficient condition is that our condition (ii) be satisfied, i.e., that the set of ordinary and the set of twisted equilibrium points meet. This is equivalent to the existence of a saddle-point in the sense of (3.1). The reader should note, however, that this only assures invariance of the optimal strategies, not of the payoffs.

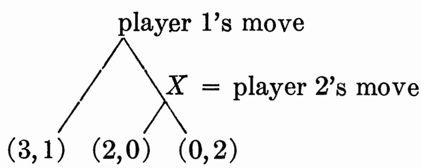
(d). We investigated a number of other possible definitions that are similar in spirit to the one given. For example, motivated by the discussion in the previous subsection, we might consider dropping condition (i) of our definition entirely. The remaining condition (ii) is equivalent to the saddle-point formulation (3.1). Theorem A would fail, as is shown by example (3.2); but we could redefine value to be the unique saddle-point payoff, and the remaining theorems go through (the proof of Theorem D becomes much easier, but on the other hand the conclusion of the theorem would be less significant).

Another possibility is to retain condition (i) and drop condition (ii). Theorem A would remain true, but Theorems B and C would fail. For example, in the game

$$(3.3) \qquad \begin{array}{|c|c|c|} \hline 0,0 & -2,\ -1 & 2,\ -1 \\ \hline 0,0 & 1,\ -2 & 1,\ \ 2 \\ \hline \end{array}$$

player 1 can guarantee himself at least 0 by playing the bottom strategy and can guarantee that player 2 won't obtain more than 0 by playing the top strategy, but there is no strategy that will guarantee both (I am indebted to L. S. Shapley for this example). Theorem D can be made to go through if "good" is appropriately redefined.

Finally, we could strengthen the definition by demanding that the set of ordinary and the set of twisted equilibrium *points* (rather than the payoffs) coincide. In this case Theorem D would fail; in fact the extensive game



does not satisfy this strengthened definition, though G_X and G_D do satisfy it.

4. Proofs.

Proof of Theorem A. Let s and t be equilibrium points. Let t^* be a twisted equilibrium point such that $h(t) = h(t^*)$; by the definitions of equilibrium point and twisted equilibrium point, $h_1(s) \geqq h_1(t_1^*, s_2) \geqq h_1(t^*) = h_1(t)$. Similarly, $h_1(t) \geqq h_1(s)$; hence $h_1(t) = h_1(s)$. The proof is similar for h_2.

Proof of Theorem B. Any component of a twisted equilibrium point will assure (a), and any component of an equilibrium point will assure (b). By condition (ii) of the definition of a.s.c. games, some points are both; this completes the proof.

Proof of Theorem C. The first statement follows at once from the definition of good strategies, and the converse from the proof of Theorem B.

Proof of Theorem D. Let s be an equilibrium point in G. Denote by s^X and s^D the strategy pair s restricted to G_X and G_D respectively; subscripts will denote components.

LEMMA 1. *If X occurs with positive probability when s is played, then s^X is an equilibrium point in G_X.*

Proof. Suppose that player 1 can improve his payoff in G_X by playing t_1^X instead of s_1^X, while player 2 plays s_2^X. Then he can also improve his payoff in G by playing a strategy composed of s_1^D and t_1^X, while player 2 plays s_2.

LEMMA 2. *s^D is an equilibrium point in G_D.*

Proof. Suppose player 1 can improve his payoff in G_D by playing t_1^D instead of s_1^D. Denote by G_D^s the game in which we attach to X the outcome of G_X if the players play s^X (rather than attaching the value of G_X). If X occurs with positive probability when s is played, then s^X is an equilibrium point in G_X, and hence its payoff is the value of G_X; thus $G_D^s = G_D$. Hence by composing t_1^D with s_1^X, player 1 can improve his payoff in G, contrary to the assumption that s is an equilibrium point. If X occurs with probability 0 when s is played, then the payoff to s in G_D is the same as that in G_D^s (since the two games differ at most in their payoff at X). By assumption, player 1 can improve his payoff in G_D; hence if he plays a good strategy in G_X and t_1^D in G_D, he will obtain in G at least what he obtains in G_D by playing t_1^D. But this is more than the payoff to s_1^D in G_D, which is the same as the payoff to s_1^D in G_D^s, which is the same as the payoff to s_1 in G (cf. [1], Theorem 2). So s is not an equilibrium point in G, contrary to assumption.

LEMMA 3. *Let the value of G_D be v. Then every equilibrium point in G has payoff v.*

Proof. Follows at once from Lemmas 1 and 2.

COROLLARY 4. *Condition* (i) *of the definition of a.s.c. games is satisfied by G.*

Proof. We can apply Lemma 3 to the game G^*, obtaining v^* ($= (-v_2, -v_1)$) for payoffs to equilibrium points in G^*, hence twisted equilibrium points in G have payoff v.

LEMMA 5. *The composition of equilibrium points in G_X and G_D yields an equilibrium point in G.*

Proof. Let s^X and s^D be the respective equilibrium points. The payoff to s^X is the value of G_X; hence $G_D^s = G_D$. Hence s^D is an equilibrium point in G_D^s. The result now follows from Theorem 3 of [1].

To complete the proof of Theorem D, apply Lemma 5 to G^*, and deduce a result corresponding to Lemma 5 for twisted equilibrium points. Now let s^X and s^D be strategy pairs in G_X and G_D respectively, that are both ordinary and twisted equilibrium points. Let s be a strategy pair in G that decomposes into s^X and s^D. By Lemma 5 s is an equilibrium point; and by the corresponding result for twisted equilibrium points, s is also a twisted equilibrium point. Hence condition (ii) of the definition of a.c.s. games is satisfied, and hence G is a.s.c. (because of Corollary 4). Finally, suppose good strategies s_1^X and s_1^D to be given; choose good strategies s_2^X and s_2^D. Then s^X and s^D are both ordinary and twisted equilibrium points, and therefore their composition s also is. Hence (by Theorem C) s_1 is good.

Acknowledgments. I am greatly indebted to L. S. Shapley for several highly helpful comments. I also wish to thank H. W. Kuhn, C. E. Langenhop and R. Reichardt for a number of helpful discussions.

REFERENCES

1. H. W. KUHN, *Extensive games and the problem of information*, in [7], pp. 189–216.
2. J. F. NASH, *The bargaining problem*, Econometrica, 18 (1950), pp. 155–162.
3. ———, *Non-cooperative games*, Ann. of Math., 54 (1951), pp. 286–295.
4. ——— *Two-person cooperative games*, Econometrica, 21 (1953), pp. 128–140.
5. H. RAIFFA, *Arbitration schemes for generalized two-person games*, in [7], pp. 361–387.
6. N. N. VOROBYEV, *Equilibrium points in bimatrix games*, Teor. Veroyatnost. i Primenen., 3 (1958), pp. 318–331 (Russian); English translation in Theory of Probability and its Applications, 3 (1958), pp. 297–309.
7. H. W. KUHN AND A. W. TUCKER, editors, *Contributions to the Theory of Games* II, Annals of Mathematics Study No. 28, Princeton University Press, 1953.
8. R. DUNCAN LUCE AND HOWARD RAIFFA, *Games and Decisions*, Wiley, New York, 1957, pp. 117–118.