AN INCOMPLETENESS PROBLEM IN HARSANYI’S GENERAL THEORY OF GAMES AND CERTAIN RELATED THEORIES OF NON-COOPERATIVE GAMES

ABSTRACT. The theory of games recently proposed by John C. Harsanyi in ‘A General Theory of Rational Behavior in Game Situations’, (Econometrica, Vol. 34, No. 3) has one anomalous feature, viz., that it generates for a special class of non-cooperative games solutions which are not equilibrium points. It is argued that this feature of the theory turns on an argument concerning the instability of weak equilibrium points, and that this argument, in turn, involves appeal to an unrestricted version of a postulate subsequently included in the theory in restricted form. It is then shown that if this line of reasoning is permitted, then one must, by parity of reasoning, permit another instability argument. But, if both of these instability arguments are permitted in the construction of the theory, the resultant theory must be incomplete, in the sense that there will be simple non-cooperative games for which such a theory cannot yield solutions. This result is then generalized and shown to be endemic to all theories which have made the equilibrium condition central to the treatment of non-cooperative games. Some suggestions are then offered concerning how this incompleteness problem can be resolved, and what one might expect concerning the postulate structure and implications of a theory of games which embodies the revisions necessitated by a resolution of this problem.

In a very comprehensive and closely reasoned paper entitled, ‘A General Theory of Rational Behavior in Game Situations’ John C. Harsanyi has sought to integrate nearly twenty-five years of analysis and theory construction with regard to quite different classes of games, and to develop a unified theory of rational decision making for all games. His theory has the threefold merit that (1) it provides a very clear and concise conceptual framework within which one can understand and relate a number of important problems concerning such matters as, e.g., the stability of solutions, joint-efficiency requirements, coalition structures, bargaining criteria, and coordination difficulties; (2) it is based on a relatively simple set of postulates; and (3) it provides determinate solutions for all classes of games. Harsanyi’s theory does contain, however, one anomalous feature: it generates for a special class of noncooperative games solutions which do not constitute equilibrium $n$-tuples of strategy choices for the $n$ players.

I propose to examine in detail the line of reasoning by which Harsanyi
is led to the construction of a theory of games with this feature. The examination will issue in a number of closely related findings. In Sections 3 and 4 I shall argue that Harsanyi's construction turns on an argument concerning the instability of weak equilibrium points. In Sections 5 and 6 I shall show that this argument involves an appeal to an unrestricted version of a postulate which is subsequently included in the theory itself in an explicitly restricted form. In Section 7 I shall argue that if this line of reasoning is permitted in Harsanyi's construction, then, by parity of reasoning, there is another instability argument which must be permitted also. This other argument concerns the instability of non-equilibrium points, and can be shown to turn on an appeal to an unrestricted version of another postulate which is included in the theory in explicitly restricted form. In Section 8 I shall show that if both of these instability arguments are permitted in the construction, Harsanyi's theory must be incomplete, in the sense that there are simple non-cooperative games for which the theory cannot yield solutions. In Section 9 I shall sketch an argument to the effect that the incompleteness problem of Harsanyi's construction can be shown to be endemic to all theories which have made the equilibrium condition central to the treatment of non-cooperatives games. Finally, in Section 10, I shall make some suggestions concerning how this incompleteness problem can be resolved, and what one might expect concerning the postulate structure and implications of a theory of games which embodies the revisions necessitated by a resolution of this problem.

II

Harsanyi's theory is based on eight postulates which he divides into two distinct groups. While I shall be concerned primarily in this paper with A1, A2, A5 and B1, I shall briefly describe them all so as to provide the reader with a general sense of Harsanyi's theory as a whole.

Class A: Postulates of Rational Behavior

A1. Maximin Postulate. This postulate requires that in a game which is unprofitable for a given player, that player should select a maximin strategy.

A game is said to be unprofitable to a given player if it can be shown that no solution to that game can yield that player more than his