EPISTEMIC LOGIC AND THE FOUNDATIONS OF GAME THEORY

For two reasons the meeting of epistemic logic and game theory was no doubt inevitable. In the first place, during the 1970s and 1980s game theorists were developing, using their own formal tools, more and more precise treatments of epistemic matters. In the second place, it was becoming clear during these years that the formalisms of measure and probability in which game theory is standardly cast were much less distant than they at first seemed from modal propositional logic, the branch of logic to which the most widely used epistemic logics belong. In this foreword we shall briefly describe these twin developments, to which each of the five papers collected in this special issue of *Theory and Decision* attests in its own way. We shall take as read the main ideas of epistemic logic itself; readers who would like an introduction to the subject may wish to consult Section 2 of Bacharach's paper.

The many attempts which have been made either to justify the use of Nash equilibrium as a solution, or to restrict it by appropriate 'refinements', have enabled game theorists to detach, and explore the game-theoretical consequences of, alternative formal principles governing the subjective reasoning of the players.\(^1\) To give one example, the analysis of Selten's 'subgame-perfect' equilibrium refinement has revealed unsuspected obstacles to classical arguments by 'backward induction'. Game theory has also moved in the opposite direction, seeking to weaken the Nash equilibrium solution concept in various criteria of 'rationalizability' which appeared in the 1980s. These criteria are justified by considerations which are distinctively epistemic. They incorporate a finite or infinite regress of reciprocal beliefs, rooted in simple beliefs in the rationality of the players and the rules of the game. Issues such as these, concerning the relation between solution- hood and the epistemic principles informing the reasoning of players, are central in the papers of Bacharach and Stalnaker.

Bacharach's paper offers an introduction to the basic ideas of modal propositional logic and standard epistemic logic as well as to certain

relevant developments in 'nonmonotonic' logic. He applies the concepts of epistemic logic to define a formal object called a broad theory of a game, intended to make manifest the full epistemic structure of the constructions used by game theorists to describe games. This formal object allows him to make explicit first the 'knowledge base' attributed to each player by the theorist, then the dual principle according to which players know all the logical consequences of the base ('cleverness') and only these consequences ('cloisteredness'). The first half of the principle is an assumption of 'logical omniscience' and raises the question of how we might restrict the deductive consequence relation under which standard epistemic logics assume belief sets to be closed. The second half, which is trickier to capture because it calls for a metalinguistic formulation, raises the opposite question of how we might expand the total collection of beliefs attributable to players; one way Bacharach proposes is the method, characteristic of theories of belief revision and of nonmonotonic logics, which consists in allowing inferences to exceed, in appropriate cases, that which is sanctioned by classical deductions.

Stalnaker analyses alternative concepts of game-theoretic equilibrium by a new method which draws both on expected utility theory and on one of the classic constructions of modal logic, Kripke's (1963) semantics. Under the name model of a game he defines what is, in effect, a Kripke structure enriched by endowing each player with a prior probability measure and a decision function, each defined over the structure's set of possible worlds. Different versions of the notion of rationalizable solution, and Nash equilibrium, can be characterized, extensionally, by appropriate classes of models of the game. These classes correspond to natural epistemic properties of the players. For example, Stalnaker characterizes in this manner the class of Nash equilibria in two-person games thus: Pi knows the beliefs of Pj about Pi's strategy choice, and knows that Pj maximizes expected utility. Stalnaker then uses his semantic method to define and to justify epistemically a new solution concept, 'strong rationalizability'.

Both Bacharach's and Stalnaker's papers concern the theory of games of 'complete information'. A now classic construction in the theory of games of 'incomplete information' illustrates the role played by players' beliefs in another way. In 1967–68 Harsanyi showed that an