Cambridge University Press
0521416744 - Knowledge, Belief, and Strategic Interaction
Edited by Cristina Bicchieri Maria Luisa Dalla Chiara
Frontmatter
More information

In recent years there has been a great deal of interaction among game theorists, philosophers, and logicians with respect to certain foundational problems concerning rationality, the formalization of knowledge and practical reasoning, and models of learning and deliberation.

This unique volume brings together the work of some of the preeminent figures in their respective disciplines, all of whom are engaged in research at the forefront of their fields. Together they offer a conspectus of the interaction of game theory, logic, and epistemology in the formal models of knowledge, belief, deliberation, and learning and in the relationship between Bayesian decision theory and game theory, as well as between bounded rationality and computational complexity.

Knowledge, belief, and strategic interaction

Cambridge Studies in Probability, Induction, and Decision Theory

General editor: *Brian Skyrms*

This new series is intended to be a forum for the most innovative and challenging work in the theory of rational decision. It focuses on contemporary developments at the interface between philosophy, psychology, economics, and statistics. The series addresses foundational theoretical issues, often quite technical ones, and therefore assumes a distinctly philosophical character.

Other titles in the series
Ellery Eells, *Probabilistic Causality*
Richard Jeffrey, *Probability and the Art of Judgment*
Robert Koons, *Paradoxes of Strategic Rationality*

Forthcoming
J. Howard Sobel, *Taking Chances*
Patrick Maher, *Betting on Theories*
Patrick Suppes and Mario Zanotti, *Foundations of Probability with Applications*
Cristina Bicchieri, *Rationality and Coordination*
Clark Glymour and Kevin Kelly (eds.), *Logic, Confirmation, and Discovery*

Knowledge, belief, and strategic interaction

Edited by

Cristina Bicchieri
Carnegie Mellon University

Maria Luisa Dalla Chiara
University of Florence

CAMBRIDGE
UNIVERSITY PRESS

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
40 West 20th Street, New York, NY 10011-4211, USA
10 Stamford Road, Oakleigh, Victoria 3166, Australia

First published 1992

Library of Congress Cataloging-in-Publication Data

Knowledge, belief, and strategic interaction / [edited by] Cristina Bicchieri, Maria Luisa Dalla Chiara.

p. cm. – (Cambridge studies in probability, induction, and decision theory)

Papers presented at the workshop knowledge, belief, and strategic interaction which took place in Italy in June 1989.

ISBN 0-521-41674-4

1. Reasoning. 2. Belief and doubt. 3. Decision-making. 4. Rationalism. 5. Game theory. I. Bicchieri, Cristina. II. Dalla Chiara, Maria Luisa. III. Series.
BC177.K58 1992
121–dc20 91-38415
CIP

A catalog record for this book is available from the British Library.

ISBN 0-521-41674-4 hardback

Transferred to digital printing 2004

Contents

Preface

This book collects the papers presented at the workshop on "Knowledge, Belief, and Strategic Interaction," which took place in Castiglioncello, Italy, in June 1989. The workshop was a first attempt at an exchange between philosophers and game theorists, as in recent years game theorists have shown a growing interest in foundational problems and there are several areas in which their concerns overlap with those of philosophers.

Philosophers have a tradition of inquiring into topics such as rationality, learning, and knowledge, as well as the practical reasoning that results in deliberation and decision making. These issues have been mainly studied in an individual context: The knowing subject is depicted as facing a stationary natural environment of which he tries to explain and predict some features. Game theorists are equally involved in defining rationality, learning, and knowledge, although the environment faced by the individual is not a natural but a social one. In a game-theoretic context, one's knowledge or beliefs are about what other players plan to do, which is in turn determined by what these players know (or believe) about one's choices. This interdependency is captured by the concept of Nash equilibrium, which is the most common solution for noncooperative games. A traditional argument for the predictive significance of Nash equilibria asserts that a virtual process of reflection about what other rational players should expect, and accordingly should do, will converge to an equilibrium, so that rational agents who understand the game and think it through thoroughly before choosing their action should play in this way, even if the game is played only once. However, game theorists have come to acknowledge that, in many nontrivial cases, assuming that the only inputs needed for obtaining a solution are knowledge of the game being played and self-evident principles of rationality that make no use of any knowledge is not sufficient to guarantee that an equilibrium will be attained.

Many game theorists have questioned the adequacy of Bayesian decision theory as a foundation for a theory of strategic interaction, at least insofar as one sees the concept of Nash equilibrium as a central tenet of game theory. One way to overcome the problem is to add several assumptions about what players know about the game and about each other. For example, it is commonly assumed that the structure of the game, as well as their mutual rationality, are common knowledge among the players. Briefly stated, a fact f is said to be common knowledge among a group of people if for every finite n, the statement "everyone knows that everyone knows that . . . everyone knows f" is true, where "everyone knows" is repeated n times. Since the players in a game assign subjective probabilities to all uncertainty, including the actions of other players, and choose on the basis of these beliefs, common knowledge of players' beliefs is usually needed in order to attain a Nash equilibrium.

Some game theorists consider the assumption of common knowledge of beliefs as gratuitous, and their approach to bridging the gap between Bayesian decision theory and game theory typically does away with it. In a unified theory of rational decision, however, other decision makers should be regarded as part of nature, and rational decision will consist in maximizing one's expected payoff relative to one's uncertainty about the state of nature. The relevance of such a Bayesian viewpoint for the theory of games has become more widely appreciated as a result of fundamental papers by Harsanyi, Aumann, Pearce, Bernheim, and Kreps and Wilson. In all these approaches, common knowledge of rationality is preserved and – depending on what amount of a priori information about the setting is allowed – different solutions (such as rationalizability or correlated equilibrium) are appropriate. Recently philosophers have dealt with this issue, and there is a growing body of literature on the Bayesian foundations of game theory. The chapters by Kadane, Levi, and Seidenfeld; Levi; and Kadane and Seidenfeld exemplify this approach. Not all philosophers share this view, though, and Sobel's and McClennen's chapters take a different, critical approach to the possibility of bridging the gap between decision theory and game theory.

The framework of this literature, although Bayesian, is still static. The focus is on a solution that satisfies a set of conditions rather than on the procedures by which the players attempt to arrive at an optimal decision. The importance of the procedural aspect of rationality has long been emphasized by Herbert Simon. In accordance with this point of view, common knowledge of rationality in strategic situations is to be thought of as common knowledge of a rational deliberational procedure. In such a situation, computations must be conceived of as generating new information; as a result, probabilities can change as a result of pure thought. There are discussions of such a possibility under the name of "dynamic probability" in the writings of Good. The chapters by Skyrms and Rabinowicz show that deliberation can be modeled as a dynamic process. Where deliberation generates new information relevant to the decision under consideration, a rational decision maker will feed back that information and reconsider. A firm decision is reached at a fixed point of this process – a deliberational equilibrium. A joint deliberational equilibrium on the part of the players is a Nash equilibrium of the game. Taking this point of view seriously leads to dynamic models of deliberation within which one can embed the theory of noncooperative games.

If one wants to maintain the centrality (and predictive value) of Nash equilibria, the first type of approach seems more adequate. It is, however, plagued with several problems. For one, the issue of belief formation is sidestepped. This deficiency should be remedied by supplementing the analysis with a model of how players learn. This represents a potentially fruitful and challenging application of theories of inductive inference, in that in a strategic environment players' beliefs may never converge to any single hypothesis, or even to any constant posterior distribution over the various hypotheses, and so agents may not converge upon any constant beliefs regarding the best action to take. Studies about how concensus is reached and how opinions can merge are directly relevant to this topic, and philosophers are giving important contributions to an understanding of the conditions under which conditional probability distributions approach each other as available data increase, as is discussed in Vannucci's chapter and in the paper by Shervish and Seidenfeld.* Another problem is that modeling how players learn to predict another player's choice appears to lead to an infinite regress of "I learn that you learn that I learn. . . ." To avoid this problem it may help to introduce elements of bounded rationality and computational complexity into the analysis, as is done in the chapters by Rubinstein and by Binmore and Shin.

A more general issue that encompasses both approaches is that of providing formal theories of how players reason. In particular, it is important to formally model players' information about the game and about other players. Even if one rejects the assumption of common knowledge of beliefs, it is usually the case that common knowledge of mutual rationality and of the structure of the game are assumed. Following Lewis, Schiffer, and Aumann, common knowledge has been studied by relating it to concepts of knowledge and probability. In this context, the logical theory of epistemic operators plays a relevant role, and there have been

*The paper by Shervish and Seidenfeld was presented at the Castiglioncello workshop but does not appear in the present volume. See "A Fair Minimax Theorem for Two-person (zero-sum) Games Involving Finitely Additive Strategies," Technical Report No. 491, Department of Statistics, Carnegie Mellon University.

several attempts at constructing axiomatic theories of the game using epistemic logic. The recent developments in modal and epistemic logics are discussed in the chapters by Dalla Chiara and by Corsi and Ghilardi, and their application to game theory is presented in Walliser's and Ferrante's chapters. Corsi and Ghilardi investigate, among others, a basic problem of modal semantics, the "transworld identity" question: how to exist, at the same time, in different possible worlds? In spite of its metaphysical appearance, the notion of possible worlds represents a useful abstract tool that permits the modeling of different types of situations. As an example, let us think of the following statement referring to a particular chess game: "Had White moved the Queen, he would have won." A semantic analysis of statements of this kind gives rise to the question of how to represent the "alter ego" of White, living in a state of the world that does not actually occur. Corsi and Ghilardi propose an elegant formal analysis of this problem in the framework of a general categorical semantics. The basic intuitive idea of their approach goes back to Leibniz and to the theory of "counterparts" developed by David Lewis.

In spite of the many successful applications to different fields, possible-worlds semantics is not yet capable of providing an adequate analysis for the problems of epistemic logics. The reason for this failure is simple: Possible worlds are usually considered closed under logical consequence, in the sense that all the consequences of a logical truth in a given world must be true in that world. Hence any semantic analysis of the epistemic operators ("know," "believe," . . .) in terms of possible worlds can hardly avoid the unpleasant and unrealistic result – knowing the axioms of a given theory should be sufficient for knowing all the theorems – according to which agents should be logically omniscient. Dalla Chiara's chapter discusses an alternative approach to epistemic semantics, where epistemic operators are analyzed in terms of intensions, and where logical omniscience is avoided.

An even more basic formal analysis of games should include models of action. Segerberg's chapter proposes a general approach to dynamic logic in the framework of a standard set-theoretical setting. Taking as primitive the notion of state space, he analyzes the concepts of *state-of-affairs, event, process,* and *action*. The theory gives rise to a number of natural game-theoretic applications. One can try to characterize formally not only actions but also prescriptions of behavior or more generally moral principles. This aim is pursued in Magari's chapter, where an abstract characterization for the notion of Pascalian ethics is proposed. Roughly, a Pascalian ethics is identified with a function that permits one to maximize the probability of a given situation, determined by an action, multiplied by the value which one attributes to that situation.

Another area where logic can contribute to game theory is that of games in which there are communication stages. Mundici's chapter describes the logic of Ulam's game with lies: a logic where contradictory sentences do not lead to total ignorance. More precisely, an Ulam game where n lies are admitted corresponds to a Lucasiewicz $(n+2)$-valued logic. Lucasiewicz many-valued logics are examples of "paraconsistent" logics, where a certain degree of tolerance toward contradictions is admitted. For example, both the principle of noncontradiction (the negation of a contradiction is always true) and the Scotian principle ("ex absurdo sequitur quodlibet") are violated. From an intuitive viewpoint, these logics may represent a good framework for modeling epistemic situations in which contradictions do not play a totally destructive role, contrary to what happens in classical logic as well as in many alternative logics. Significantly enough, Lucasiewicz logics are also deeply related to AF C^*-algebras that are used in the description of quantum spin systems, as has been shown by Mundici in other papers. As a consequence, the logical interpretation of Ulam games might be useful in the abstract investigation of the physical systems that are studied in quantum mechanics.

Turning to the topic of common knowledge, one persistent problem is that it seems difficult to attain. Bacharach's and Rubinstein's chapters each give arguments why common knowledge cannot be attained in finite games. Moreover, it is not evident that common knowledge needs to be assumed in order to obtain a Nash equilibrium. For example, in finite, extensive-form games of perfect information, assuming the players to have only limited knowledge often provides a better approximation to real-life interactions and allows for solutions that are more plausible and intuitive, as the chapters by Bicchieri and Reny show.

Another important topic of relevance to game theorists and philosophers alike is a definition of what constitutes rationality of beliefs. After Nash, it was commonly assumed that once a strategy combination is a Nash equilibrium, it can be the solution for a noncooperative game. Selten pointed out that some Nash equilibria involve irrational moves and that such equilibria cannot serve as solutions for sequential games played by rational players. Equilibria may involve irrational moves because a player may have irrational beliefs about another player's choice. This happens because the beliefs that support a Nash equilibrium need only be internally consistent and self-fulfilling; no further restriction is imposed on their rationality. In extensive-form games, equilibria may involve irrational moves only at information sets that will never be reached if the players follow their equilibrium strategies. What happens out of equilibrium, however, affects what constitutes a best choice in a Nash equilibrium. In an extensive-form game, it makes sense to ask what another's reaction will be if one deviates from the prescribed equilibrium strategy. For suppose there are two equilibria, one of which involves some threat on the part of one of the players. Then it makes sense for the threatened party to ask whether the opponent, facing a deviation on his part, would fulfill the threat. If the answer is negative then the threat is not credible, and the equilibrium which is based on it is ruled out as not sensible. To rule out irrational equilibria, game theorists generally refer to the extensive form of the game, which conveys more information than the normal form. That normal and extensive form are not equivalent is not, however, an accepted tenet; Harsanyi's chapter is a defense of their basic asymmetry.

The vast literature on refinements of Nash equilibrium, while searching for criteria that restrict the number of possible equilibria, is mainly concerned with defining what it means to be rational at information sets off the equilibrium path. Within the class of refinements of Nash equilibrium, two different approaches can be identified. One solution aims at imposing restrictions on players' beliefs by explicitly allowing for the possibility of error on the part of the players. This approach underlies both Selten's notion of "perfect" equilibrium and Myerson's notion of "proper" equilibrium. The alternative solution is based instead upon an examination of rational beliefs rather than mistakes. The idea is that players form conjectures about other players' choices, and that a conjecture should not be maintained in the face of evidence that refutes it. This approach underlies the notion of "sequential" equilibrium proposed by Kreps and Wilson. All of these solutions aim at imposing restrictions on players' beliefs, so as to obtain a unique rational recommendation as to what to believe about the other players' behavior. This guarantees that rational players will select the unique equilibrium that is supported by these beliefs. Both approaches, however, fail to rule out some equilibria supported by beliefs that, although coherent, are still intuitively implausible. The limit of these approaches is that restrictions are imposed only on equilibrium beliefs, while out-of-equilibrium beliefs are unrestricted: A player will ask whether it is reasonable to believe the other player will play a given equilibrium strategy, but not whether the beliefs supporting the other player's choice are rational. To provide a satisfactory means of discriminating among equilibria, restrictions need to be imposed on all sorts of beliefs, even out-of-equilibrium ones.

A player, that is, must be able to rationally justify to himself every belief, and expect the other players to expect him to adopt such a rational justification. A possible solution lies in combining the heuristic method implicit in the "small mistakes" approach with the analysis of belief rationality characteristic of the sequential equilibrium notion. The "small mistakes" approach stresses the role of anticipated actions off the equilibrium path in sustaining the equilibrium. Because the reference point is an equilibrium of which one tests the stability against deviations, players are modeled as being involved in counterfactual arguments concerning what would happen if they were to deviate from their equilibrium strategy, or were themselves faced with other players' deviations. These arguments involve a change in the original set of beliefs, and for the process of belief change not to be arbitrary, rationality conditions must be imposed on it. Belief rationality, in this case, is a property of beliefs that are revised through a rational procedure. If there were a unique rational process of belief revision, then there would be a unique best theory of deviations that a rational player would be expected to adopt, and common knowledge of rationality would suffice to eliminate all equilibria that are robust only with respect to implausible deviations.

Philosophers have developed two main approaches to modeling epistemic states. One is the foundations theory, which holds that one needs to keep track of the justification for one's beliefs. The other is the coherence theory, which holds that one need not consider where one's beliefs come from. In coherence theory, the focus is on the logical structure of the beliefs; what matters is how a belief coheres with the other beliefs that are accepted in the present state. The foundations and the coherence theories have very different implications for what should count as rational changes of belief systems. According to the foundations theory, belief revision should consist, first, in giving up all beliefs that no longer have a satisfactory justification and, second, in adding new beliefs that have become justified. According to the coherence theory, the objectives are, first, to maintain consistency in the revised epistemic state and, second, to make minimal changes of the old state that guarantee sufficient overall coherence. A satisfactory model of belief change in games must encompass both approaches. Such dynamics of belief revision are extensively explored in Gärdenfors's chapter. Another approach to counterfactual reasoning in games is proposed by Shin, who applies the logic of conditionals developed by Stalnaker and Lewis.

We are greatly indebted to the Centro Fiorentino di Storia e Filosofia della Scienza and to the Centro Interuniversitario per la Teoria dei Giochi e le Applicazioni, which made the workshop possible. We are also grateful to Alessandro Pagnini, who organized a pleasant stay in Castiglioncello for all of us.

Pittsburgh and Firenze
May 1991

Cristina Bicchieri
Maria Luisa Dalla Chiara

Contributors

Michael Bacharach *Christ Church, Oxford*
Cristina Bicchieri *Department of Philosophy, Carnegie Mellon University*
Ken Binmore *Department of Economics, University of Michigan*
Giovanna Corsi *Dipartimento di Filosofia, Università di Firenze*
Maria Luisa Dalla Chiara *Dipartimento di Filosofia, Università di Firenze*
Vittorioemanuele Ferrante *Dipartimento di Economia, Università di Firenze*
Peter Gärdenfors *Filosofiska Institutionen, Lund Universitet*
Silvio Ghilardi *Dipartimento di Matematica, Università di Milano*
John C. Harsanyi *School of Business Administration, University of California, Berkeley*
Joseph B. Kadane *Department of Statistics, Carnegie Mellon University*
Isaac Levi *Department of Philosophy, Columbia University*
Roberto Magari *Dipartimento di Matematica, Università di Siena*
Edward F. McClennen *Department of Philosophy, Bowling Green State University*
Daniele Mundici *Dipartimento di Informatica, Università di Milano*
Wlodzimierz Rabinowicz *Filosofiska Institutionen, Uppsala Universitet*
Philip J. Reny *Department of Economics, University of Western Ontario*
Ariel Rubinstein *Department of Economics, Tel Aviv University*
Krister Segerberg *Filosofiska Institutionen, Uppsala Universitet*
Teddy Seidenfeld *Department of Philosophy, Carnegie Mellon University*
Hyun Song Shin *University College, Oxford*
Brian Skyrms *Department of Philosophy, University of California, Irvine*
Jordan Howard Sobel *Department of Philosophy, Scarborough College, University of Toronto*
Stefano Vannucci *Dipartimento di Economia, Università di Siena*
Bernard Walliser *CERAS, Ecole Nationale Des Ponts et Chaussees, Paris*