This book represents a major contribution to game theory. It offers a new conception of equilibrium in games: strategic equilibrium. This new conception arises from a study of expected utility decision principles, which must be revised to take account of the evidence a choice provides concerning its outcome. The argument for the new principles distinguishes reasons for action from incentives, and draws on contemporary analyses of counterfactual conditionals. The book also includes a procedure for identifying strategic equilibria in ideal normal-form games.

In synthesizing decision theory and game theory in a powerful new way this book will be of particular interest to all philosophers concerned with decision theory and game theory as well as economists and other social scientists.

Equilibrium and Rationality

Cambridge Studies in Probability, Induction, and Decision Theory

Ellery Eells, *Probabilistic Causality*
Richard Jeffrey, *Probability and the Art of Judgment*
Robert C. Koons, *Paradoxes of Belief and Strategic Rationality*
Cristina Bicchieri and Maria Luisa Dalla Chiara (eds.), *Knowledge, Belief, and Strategic Interactions*
Patrick Maher, *Betting on Theories*
Cristina Bicchieri, *Rationality and Coordination*
J. Howard Sobel, *Taking Chances*
Jan von Plato, *Creating Modern Probability: Its Mathematics, Physics, and Philosophy in Historical Perspective*
Ellery Eells and Brian Skyrms (eds.), *Probability and Conditionals*
Cristina Bicchieri, Richard Jeffrey, and Brian Skyrms (eds.), *The Dynamics of Norms*
Patrick Suppes and Mario Zanotti, *Foundations of Probability with Applications*

Equilibrium and Rationality

Game Theory Revised by Decision Rules

Paul Weirich
University of Missouri – Columbia

CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE UNIVERSITY PRESS
Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo

Cambridge University Press
The Edinburgh Building, Cambridge CB2 8RU, UK

Published in the United States of America by Cambridge University Press, New York

www.cambridge.org
Information on this title: www.cambridge.org/9780521593526

First published 1998
This digitally printed version 2007

A catalogue record for this publication is available from the British Library

Library of Congress Cataloguing in Publication data
Weirich, Paul, 1946–
Equilibrium and rationality : game theory revised by decision
rules / Paul Weirich.
p. cm. – (Cambridge studies in probability, induction, and
decision theory)
Includes bibliographical references and index.
ISBN 0-521-59352-2 (hb)
1. Game theory. 2. Statistical decision. I. Title. II. Series.
QA269.W46 1998
519.3 – dc21 97-16555
CIP

ISBN 978-0-521-59352-6 hardback
ISBN 978-0-521-03802-7 paperback

For my parents

Contents

Preface

Game theory is a rich field for the philosopher. It formulates principles of rational action and so makes immediate contributions to the branch of logic that studies practical reasoning. Its principles of strategic reasoning also provide the undergirding for the social contract theory and the study of social and linguistic convention. Revisions of central game-theoretic tenets reverberate throughout philosophy.

This book makes some assumptions about rationality and equilibrium in games and traces out their consequences. These are the main assumptions: (1) There are no dilemmas of rationality; an agent has a rational choice in every choice situation. (2) Every ideal game has a solution, that is, a profile of jointly rational strategies. (3) A rational choice is self-supporting. (4) An equilibrium is a profile of strategies that are jointly self-supporting. These assumptions suggest replacing Nash equilibrium, which fails to exist in some ideal games, with a new type of equilibrium called *strategic equilibrium*. Strategic equilibrium is introduced via a study of self-support. Taking an equilibrium to be a profile of jointly self-supporting strategies, I use a new account of self-support to obtain the new type of equilibrium. I show that a strategic equilibrium exists in every ideal normal-form game meeting certain mild restrictions and provide procedures for finding strategic equilibria in those games.

This project arose from reflections on cooperative games, where individuals may form coalitions that adopt joint strategies and otherwise act as agents. I took being an equilibrium as a necessary condition for being a solution to a cooperative game, and this led to a problem. The most widely accepted account of equilibrium for cooperative games takes a game's outcome to be an equilibrium just in case it is in the game's "core," and so is an outcome such that no coalition can increase its payoff by unilaterally changing its strategy. But many cooperative games have empty cores, and so no solutions according to that account of equilibrium. To address this problem, Aumann and Maschler (1964) advance a type of equilibrium known as a "bargaining point" that may exist outside a game's core. Their main idea is to take some incentives to be defeated by objections to their pursuit. A similar idea also leads Bernheim et al. (1987) to introduce for games with certain cooperative elements a type of equilibrium called "coalition-proof Nash equilibrium."

To investigate defeated incentives in cooperative games, I applied the decision principle of weak ratifiability introduced in Weirich (1988) to cooperative games with an infinite number of salient outcomes, for instance, an infinite number of divisions of some resource. The conclusions about solutions that emerged, however, are clearest for noncooperative games. Complications concerning the treatment of coalitions as agents obscure them in cooperative games. To throw these conclusions into high relief, this book focuses on noncooperative games. It uses the general idea of defeaters for incentives to define equilibrium for noncooperative games. In certain cases where an agent's pursuing an incentive to switch strategy generates incentives for other agents to switch strategy, I say that the agent's incentive is not a sufficient reason to switch strategy – it is defeated. Thus a strategy need not be incentive-proof to be self-supporting and part of an equilibrium. Although I define equilibrium in a detailed way for noncooperative games only, I envisage extending the definition to cooperative games as well.

My goal is to use principles of rational decision, in particular, principles of self-support such as the principle of weak ratification, to explicate equilibrium, taken as a necessary condition for a solution. I originally planned to derive the realization of a Nash equilibrium from principles of self-support, using idealizations about the rationality and common knowledge of the players. But that project required many restrictions on the type of game treated. To gain generality, I introduce strategic equilibrium, a new, weaker type of equilibrium. And to simplify, I focus on its existence rather than its realization and adopt an idealization about the players' foreknowledge of the game's outcome rather than derive that idealization from more basic idealizations about the players' common knowledge.

The book has three main parts. First, it presents the problem that motivates its proposals – the absence of Nash equilibria in some ideal games – and explains the nature of solutions and equilibria in games. Second, it advances a new decision principle of self-support, applies the principle to noncooperative games to obtain a new type of equilibrium, and shows that the new type of equilibrium exists in every ideal normal-form game meeting certain restrictions. Third, it formulates a procedure for finding equilibria, applies it to examples, and compares the new equilibrium standard for solutions to other familiar standards.

Chapter 1 presents the problem of missing Nash equilibria in a preliminary way and explains our conception of games and solutions. Chapter 2 discusses the idealizations we adopt for games. Chapter 3 shows why the absence of Nash equilibrium in ideal games is damaging for that conception of equilibrium and in general lays out the reasons for a revised conception of equilibrium. In particular, it examines problems with the allied standard of incentive-proofness for solutions. This standard says that a solution must

be such that no agent has an incentive to switch from his part in the solution given the parts of other agents. In a noncooperative game in which agents choose independently, every switch in strategy is unilateral. Thus the standard requires that no agent have an incentive to switch from a solution. The standard of incentive-proofness for solutions supports Nash equilibrium as a necessary condition for solutions in ideal noncooperative games. Because of the problems with the standard, we reject it and the companion standard of Nash equilibrium.

Chapter 4 argues for the decision principle of self-support used to flesh out the new equilibrium standard for solutions and rejects competing principles of dominance, ratification, and expected utility maximization. Its arguments undercut decision-theoretic support for Nash equilibria and open the door to a new, less restrictive type of equilibrium. Chapter 5 presents the new type of equilibrium and demonstrates its existence in all ideal normal-form games meeting certain simplifying assumptions.

Chapter 6 formulates a procedure for finding the new type of equilibrium. Chapter 7 applies the new equilibrium standard to games of special interest. Chapter 8, the final chapter, compares the new equilibrium standard with other standards, including the standard of Nash equilibrium and the standard of nondomination, the two most widely endorsed standards for solutions.

The book's main foundational point is the argument for the new principle of self-support, and its main technical points are the demonstrations of the existence of the new type of equilibrium in ideal normal-form games and of the adequacy of the search procedure for equilibria of the new type. Other important auxiliary claims are the following: (1) Decision principles should be adjusted to recognize the futility of pursuing certain types of incentive. It is a mistake to require that all incentives be pursued. The principle of incentive-proofness makes sense only where pursuit of incentives is not futile. (2) An agent's choice should use his knowledge of his pursuit of incentives just as it should use his knowledge of other agents' pursuit of incentives. In particular, it should use his knowledge of his responses to other agents, just as it should use his knowledge of other agents' responses to him. (3) Equilibrium in a noncooperative game may be defined in a way that makes equilibria depend upon features of a game not represented by its payoff matrix. For example, equilibria may depend upon the way ties between strategies are broken.

The book's main limitations are the following: (1) Only ideal normal-form games, in which agents are prescient concerning other agents' choices, are examined in depth. (2) The idealization of prescience is not derived from more basic idealizations concerning common knowledge and the like. (3) Only the existence, not the realization, of equilibria is demonstrated.

Although technical material is self-contained, some general background reading will help those unacquainted with decision theory and game theory.

See, for example, Luce and Raiffa (1957), Jeffrey (1983), and Resnik (1987).

I gratefully acknowledge support during 1994–5 by the University of Missouri Research Board and the University of Missouri-Columbia Research Council. For helpful comments, I am indebted to Peter Markie, Wlodek Rabinowicz, Howard Sobel, Xinghe Wang, and two anonymous readers. Louise Calabro and Susan Thornton provided invaluable editorial assistance.