

Introduction

Sergiu Hart^a and Andreu Mas-Colell^b

^a The Hebrew University of Jerusalem

^b Harvard University and Universitat Pompeu Fabra, Barcelona

This book constitutes a systematic exposition of the various game theoretic approaches to the issue of cooperation.

Game Theory is the study of decision making in multi-person situations, where the outcome depends on everyone's choice. The goal of each participant is to maximize his own utility, while taking into account that the other participants are doing the same. In such interactive situations, *cooperation* between the agents may lead to results that are better, for everyone, than the non-cooperative outcomes. A simple - but extensively studied - example is the so-called "Prisoners' Dilemma": Assume each one of the two players can ask a generous donor either to give him 1 million dollars, or to give 4 million dollars to the other player, the donor will carry out the instructions of both players (thus, for example, if player 1 asks for \$1M to himself and player 2 asks for \$4M to the other, then player 1 gets \$5M and player 2 gets nothing). Clearly, whatever the other player does, it is strictly better for each player to ask for \$1M to himself (more precisely, it will always lead to an additional "1M"). This yields \$1M for each; cooperation, whereby each one asks for \$4M to the other, would have yielded each \$4M instead! The Prisoner's Dilemma is by no means an artificial example. The economic competition between firms exhibits similar phenomena: keeping a commodity in short supply may be to the advantage of all producers; at the same time, it may be better for any single producer to unilaterally increase his own production.

The problems that need to be addressed are, first, whether cooperation can be reached at all; second, by what procedures are agreements reached; and third, which ones will be indeed attained. This volume will survey some of the contributions of game theory to these questions, from its early traditional theories to its current approaches.

Game theoretical approaches are usually classified as either "cooperative" or "non-cooperative". This should not be viewed as an exclusive division; these are

two ways of looking at the same problem. The Introductory Remarks of R. J. Aumann that follow this Introduction address this point in some detail.

Part A, which opens this volume, surveys the classical cooperative approach. This starts by assuming that binding agreements are possible, and it abstracts away from the detailed bargaining procedures. The selection of the appropriate cooperative outcome is usually based on a set of desired postulates or axioms, which, when applied to a class of problems, characterize one or another solution concept. Chapters 1 and 2 by W. Thomson cover the pure bargaining problems, where only the grand coalition of all players can reach a beneficial agreement; Chapter 1 deals with the classical approaches that originate with Nash's 1950 seminal paper, and Chapter 2 deals with recent axiomatizations based on internal consistency properties (the "reduced game property"). Chapters 3 and 4, by S. Hart, survey the general n -person problems where subcoalitions of players can reach agreements as well, and this of course influences the final outcome. The classical cooperative solution concepts that arise are grouped into "core-like" notions and "value-like" notions. The former include the core, the stable sets of von Neumann and Morgenstern, the bargaining set, the kernel and the nucleolus; the latter include the Nash bargaining solution, the Shapley value and their many extensions and generalizations. Chapter 5 by B. Allen deals with games of incomplete information, i.e., games where some of the participants may possess private information not known to the others. Here, the questions of cooperation are further complicated by the need to address the informational issues; how to ensure that the players have incentive to reveal the appropriate information.

Part B is devoted to non-cooperative approaches, namely, non-cooperative models that lead to cooperative solutions. One may start from a non-cooperative bargaining model, like the Ståhl-Rubinstein "alternating offers" procedure, characterize its strategic equilibria, and relate the resulting outcomes to various cooperative solutions. Or, one may start from a cooperative solution, and construct games whose equilibria yield precisely this given solution. Either way, one establishes connections between non-cooperative and cooperative setups, that further strengthen and reinforce one another. In the literature, all this is usually referred as "bargaining procedures", "non-cooperative foundations", or "implementation". The distinctions are not always clear, in particular since some of the recent implementation literature is concerned with "natural" and "simple" games. Chapter 6 by A. Mas-Colell covers bargaining procedures that lead to value-like cooperative solutions, and the second part of Chapter 7 by P. Reny and Chapter 8 by B. Allen, for the case of complete information and incomplete information, respectively. Chapter 9 by R. Vohra discusses coalitional non-cooperative approaches -i.e., models where not only individuals, but also coalitions may act strategically. Chapter 10 by J. Greenberg surveys the theory of "social situations", which looks for a stable standards of behavior in general coalitional interactions.

Part C deals with dynamic models, that is models of long-term interactions between the participants. Returning, for example, to the Prisoners' Dilemma, it seems clear that if the same participants play it again and again, then cooperation may indeed be attained. However, this is by no means always so; for instance, in a fixed finite-horizon repetition, it is very difficult to escape the non-cooperative outcome of \$1M each. There is by now a large and deep literature on "repeated games" -starting with the so-called "Folk Theorem" - that shows the extent to which cooperation may arise. The complete information case is covered in Chapter 11 by S. Sorin¹. Chapter 12, also by S. Sorin, then goes on to survey models of communication; namely, one examines the effect of the players being able to communicate among themselves before the game is played, and also, in the case of a multi-stage game, during the play. This leads to correlation and cooperation. Another important issue in multi-stage interactions is that they require, by their very nature, extremely complex strategic considerations. This suggests considering models where the assumption that players are fully rational - i.e., that they are restricted in one way or another in their choices. Chapter 13 by R.J. Aumann discusses some of the underlying ideas and approaches of this kind. The case where strategies are implemented by automata of bounded complexity is then studied in Chapter 14 by A. Neyman. Chapter 15 by V. Krishna and T. Sjöström is devoted to a simple but interesting learning model, known as the "fictitious play": players assume that the past behavior of their opponents is, in a certain sense, and appropriate predictor of their future behavior. Chapter 16, also by V. Krishna and T. Sjöström, studies another type of bounded rationality models: the "evolutionary models". These are based on the biological paradigm of natural selection and evolution, where there is no conscious optimization at all; instead, it is the dynamics of the evolution of the population that leads ultimately to equilibria and stable outcomes.

Part D, that concludes this volume, is concerned with "descriptive" results. One looks at the actual behavior of participants in various interactive situations. The question is not "what should rational players do", but rather "what do they do" in specific experiments. Chapter 17 by R. Selten surveys some of the large literature on experimental game theory, in particular relating to issues of cooperation. Since the outcomes are at times at odds with those predicted by the various theories of rational behavior, there is much need to understand what exactly are the principles leading to the different behaviors.

¹ The incomplete information case was also covered in the lectures. The reader is referred to the *Handbook of Game Theory with Economic Applications* (edited by R. J. Aumann and S. Hart, North-Holland, volume I: 1992, volumeii: 1994, volume III: forthcoming), for surveys of this topic (see Chapters 5 and 6 volume I), as well as of many other related topics.