# Game Theory

## A critical introduction

Shaun P. Hargreaves Heap
and Yanis Varoufakis

# GAME THEORY

## A Critical Introduction

*Shaun P. Hargreaves Heap and*
*Yanis Varoufakis*


London and New York

First published 1995
by Routledge
11 New Fetter Lane, London EC4P 4EE

Simultaneously published in the USA and Canada
by Routledge
29 West 35th Street, New York, NY 10001

Reprinted 1995 (twice), 1996



Typeset in Garamond by
J&L Composition Ltd, Filey, North Yorkshire

Printed and bound in Great Britain by Intype London Ltd



*British Library Cataloguing in Publication Data*
A catalogue record for this book is available from the British Library.

*Library of Congress Cataloguing in Publication Data*
Heap, Shaun Hargreaves, 1951–
Game theory: a critical introduction / Shaun P. Hargreaves Heap and Yanis Varoufakis.
p. cm.
Includes bibliographical references and index.
1 Social sciences – Mathematics. 2. Game theory. I. Varoufakis, Yanis. II. Title.
H61.25.H4 1994
519.3–dc20 94–22051

ISBN 0–415–09402–X (hbk)
ISBN 0–415–09403–8 (pbk)

# PREFACE

As ever there are people and cats to thank. There is also on this occasion electronic mail. The first draft of this book took shape in various cafeterias in Florence during YV's visit to Europe in 1992 and matured on beaches and in restaurants during SHH's visit to Sydney in 1993. Since then the mail wires between Sydney and Norwich, or wherever they are, have rarely been anything other than warm to hot, and of course we shall claim that this might account for any mistakes.

The genesis of the book goes back much longer. We were colleagues together at the University of East Anglia, where game theory has long been the object of interdisciplinary scrutiny. Both of us have been toying with game theory in an idiosyncratic way (see SHH's 1989 and YV's 1991 books) – it was a matter of time before we did so in an organised manner. The excuse for the book developed out of some joint work which we were undertaking during SHH's visit to Sydney in 1990. During the gestation period colleagues both at Sydney and at UEA exerted their strong influence. Martin Hollis and Bob Sugden, at UEA, were obvious sources of ideas while Don Wright, at Sydney, read the first draft and sprinkled it with liberal doses of the same question: 'Who are you writing this for?' (Ourselves of course Don!) Robin Cubbitt from UEA deserves a special mention for being a constant source of helpful advice throughout the last stages. We are also grateful to the Australian Research Council for grant 24657 which allowed us to carry out the experiments mentioned in Chapter 8.

It is natural to reflect on whether the writing of a book exemplifies its theme. Has the production of this book been a game? In a sense it has. The opportunities for conflict abounded within a two-person interaction which would have not generated this book unless strategic compromise was reached and cooperation prevailed. In another sense, however, this was definitely no game. The point about games is that objectives and rules are known in advance. The writing of a book by two authors is a different type of game, one that game theory does not consider. It not only involves

moving within the rules, but also it requires the ongoing creation of the rules. And if this were not enough, it involves the ever shifting profile of objectives, beliefs and concerns of each author as the writing proceeds. Our one important thought in this book is that game theory will remain deficient until it develops an interest in games like the one we experienced over the last two years. Is it any wonder that this is *A Critical Introduction*?

Lastly, there are the people and the cats: Lucky, Margarita, Pandora, Sue, Thibeau and Tolstoy – thank you.

Shaun P. Hargreaves Heap
Yanis Varoufakis
May 1994

# CONTENTS

# LIST OF BOXES

# 1

# AN OVERVIEW

## 1.1 INTRODUCTION

### 1.1.1 Why study game theory?

Game theory is everywhere these days. After thrilling a whole generation of post–1970 economists, it is spreading like a bushfire through the social sciences. Two prominent game theorists, Robert Aumann and Oliver Hart, explain the attraction in the following way:

> Game Theory may be viewed as a sort of umbrella or 'unified field' theory for the rational side of social science . . . [it] does not use different, ad hoc constructs . . . it develops methodologies that apply in principle to all interactive situations.
>
> (Aumann and Hart, 1992)

Of course, you might say, two practitioners would say that, wouldn't they. But the view is widely held, even among apparently disinterested parties. Jon Elster, for instance, a well-known social theorist with very diverse interests, remarks in a similar fashion:

> if one accepts that interaction is the essence of social life, then . . . game theory provides solid microfoundations for the study of social structure and social change.
>
> (Elster, 1982)

In many respects this enthusiasm is not difficult to understand. Game theory was probably born with the publication of *The Theory of Games and Economic Behaviour* by John von Neumann and Oskar Morgenstern (first published in 1944 with second and third editions in 1947 and 1953). They defined a game as any interaction between agents that is governed by a set of rules specifying the possible moves for each participant and a set of outcomes for each possible combination of moves. One is hard put to find an example of social phenomenon that cannot be so described. Thus a theory of games promises to apply to almost any social interaction where

individuals have some understanding of how the outcome for one is affected not only by his or her own actions but also by the actions of others. This is quite extraordinary. From crossing the road in traffic, to decisions to disarm, raise prices, give to charity, join a union, produce a commodity, have children, and so on, it seems we will now be able to draw on a single mode of analysis: the theory of games.

At the outset, we should make clear that we doubt such a claim is warranted. This is a *critical* guide to game theory. Make no mistake though, we enjoy game theory and have spent many hours pondering its various twists and turns. Indeed it has helped us on many issues. However, we believe that this is predominantly how game theory makes a contribution. It is useful mainly because it helps clarify some fundamental issues and debates in social science, for instance those within and around the political theory of liberal individualism. In this sense, we believe the contribution of game theory to be largely pedagogical. Such contributions are not to be sneezed at.

If game theory does make a further substantial contribution, then we believe that it is a negative one. The contribution comes through demonstrating the limits of a particular form of individualism in social science: one based *exclusively* on the model of persons as preference satisfiers. This model is often regarded as the direct heir of David Hume's (the 18th century philosopher) conceptualisation of human reasoning and motivation. It is principally associated with what is known today as rational choice theory, or with the (neoclassical) economic approach to social life (see Downs, 1957, and Becker, 1976). Our main conclusion on this theme (which we will develop through the book) can be rephrased accordingly: we believe that game theory reveals the limits of 'rational choice' and of the (neoclassical) economic approach to life. In other words, game theory does not actually deliver Jon Elster's 'solid microfoundations' for all social science; and this tells us something about the inadequacy of its chosen 'microfoundations'.

The next section (1.2) sketches the philosophical moorings of game theory, discussing in turn its three key assumptions: **agents are instrumentally rational (section 1.2.1); they have common knowledge of this rationality (section 1.2.2); and they know the rules of the game (section 1.2.3)**. These assumptions set out where game theory stands on the big questions of the sort 'who am I, what am I doing here and how can I know about either?'. The first and third are ontological.[1] They establish what game theory takes as the material of social science: in particular, what it takes to be the essence of individuals and their relation in society. The second raises epistemological issues[2] (and in some games it is not essential for the analysis). It is concerned with what can be inferred about the beliefs which people will hold about how games will be played when they have common knowledge of their rationality.

We spend more time discussing these assumptions than is perhaps usual in texts on game theory because we believe that the assumptions are both controversial and problematic, in their own terms, when cast as general propositions concerning interactions between individuals. This is one respect in which this is a critical introduction. The discussions of instrumental rationality and common knowledge of instrumental rationality (sections 1.2.1 and 1.2.2), in particular, are indispensable for anyone interested in game theory. In comparison section 1.2.3 will appeal more to those who are concerned with where game theory fits in to the wider debates within social science. Likewise, section 1.3 develops this broader interest by focusing on the potential contribution which game theory makes to an evaluation of the political theory of liberal individualism. We hope you will read these later sections, not least because the political theory of liberal individualism is extremely influential. Nevertheless, we recognise that these sections are not central to the exposition of game theory *per se* and they presuppose some familiarity with these wider debates within social science. For this reason some readers may prefer to skip through these sections now and return to them later.

Finally, section 1.4 offers an outline of the rest of the book. It begins by introducing the reader to actual games by means of three classic examples which have fascinated game theorists and which allow us to illustrate some of the ideas from sections 1.2 and 1.3. It concludes with a chapter-by-chapter guide to the book.

### 1.1.2 Why read this book?

In recent years the number of texts on game theory has multiplied. For example, Rasmussen (1989) is a good 'user's manual' with many economic illustrations. Binmore (1990) comprises lengthy, technical but stimulating essays on aspects of the theory. Kreps (1990) is a delightful book and an excellent eclectic introduction to game theory's strengths and problems. More recently, Myerson (1991), Fudenberg and Tirole (1991) and Binmore (1992) have been added to the burgeoning set. Dixit and Nalebuff (1993) contribute a more informal guide while Brams (1993) is a revisionist offering. One of our favourite books, despite its age and the fact that it is not an extensive guide to game theory, is Thomas Schelling's *The Strategy of Conflict*, first published in 1960. It is highly readable and packed with insights few other books can offer. However, *none* of these books locates game theory in the wider debates within social science. This is unfortunate for two reasons.

Firstly, it is liable to encourage further the insouciance among economists with respect to what is happening elsewhere in the social sciences. This is a pity because mainstream economics is actually founded on philosophically controversial premises and game theory is potentially in

rather a good position to reveal some of these foundational difficulties. In other words, what appear as 'puzzles' or 'tricky issues' to many game theorists are actually echoes of fundamental philosophical dispute; and so it would be unfortunate to overlook this invitation to more philosophical reflection.

Secondly, there is a danger that other social sciences will greet game theory as the latest manifestation of economic imperialism, to be championed only by those who prize technique most highly. Again this would be unfortunate because game theory really does speak to some of the fundamental disputes in social science and as such it should be an aid to all social scientists. Indeed, for those who are suspicious of economic imperialism within the social sciences, game theory is, somewhat ironically, a potential ally. Thus it would be a shame for those who feel embattled by the onward march of neoclassical economics if the potential services of an apostate within the very camp of economics itself were to be denied.

This book addresses these worries. It has been written for all social scientists. It does not claim to be an authoritative textbook on game theory. There are some highways and byways in game theory which are not travelled. But it does focus on the central concepts of game theory, and it aims to discuss them critically and simply while remaining faithful to their subtleties. Thus we have trimmed the technicalities to a minimum (you will only need a bit of algebra now and then) and our aim has been to lead with the ideas. We hope thereby to have written a book which will introduce game theory to students of economics and the other social sciences. In addition, we hope that, by connecting game theory to the wider debates within social science, the book will encourage both the interest of non-economists in game theory and the interest of economists to venture beyond their traditional and narrow philosophical basis.

## 1.2 THE ASSUMPTIONS OF GAME THEORY

Imagine you observe people playing with some cards. The activity appears to have some structure and you want to make sense of what is going on; who is doing what and why. It seems natural to break the problem into component parts. First we need to know the rules of the game because these will tell us what actions are permitted at any time. Then we need to know how people select an action from those that are permitted. This is the approach of game theory and the first two assumptions in this section address the last part of the problem: how people select an action. One focuses on what we should assume about what motivates each person (for instance, are they playing to win or are they just mucking about?) and the other is designed to help with the tricky issue of what each thinks the other will do in any set of circumstances.

### 1.2.1 Individual action is instrumentally rational

Individuals who are instrumentally rational have preferences over various 'things', e.g. bread over toast, toast and honey over bread and butter, rock over classical music, etc., and they are deemed rational because they select actions which will best satisfy those preferences. One of the virtues of this model is that very little needs to be assumed about a person's preferences. Rationality is cast in a means–end framework with the task of selecting the most appropriate means for achieving certain ends (i.e. preference satisfaction); and for this purpose, preferences (or 'ends') must be coherent in only a weak sense that we must be able to talk about satisfying them more or less. Technically we must have a 'preference ordering' because it is only when preferences are ordered that we will be able to begin to make judgements about how different actions satisfy our preferences in different degrees. In fact this need entail no more than a simple consistency of the sort that when rock music is preferred to classical and classical is preferred to muzak, then rock should also be preferred to muzak (the interested reader may consult Box 1.1 on this point).[3]

Thus it appears a promisingly general model of action. For instance, it could apply to any type of player of games and not just individuals. So long as the State or the working class or the police have a consistent set of objectives/preferences, then we could assume that it (or they) too act instrumentally so as to achieve those ends. Likewise it does not matter what ends a person pursues: they can be selfish, weird, altruistic or whatever; so long as they consistently motivate then people can still act so as to satisfy them best.

Readers familiar with neoclassical *Homo economicus* will need no further introduction. This is the model found in standard introductory texts, where preferences are represented by indifference curves (or utility functions) and agents are assumed rational because they select the action which attains the highest feasible indifference curve (maximises utility). For readers who have not come across these standard texts or who have forgotten them, it is worth explaining that preferences are sometimes represented mathematically by a utility function. As a result, acting instrumentally to satisfy best one's preferences becomes the equivalent of utility maximising behaviour. In short, the assumption of instrumental rationality cashes in as an assumption of utility maximising behaviour. Since game theory standardly employs the metaphor of utility maximisation in this way, and since this metaphor is open to misunderstanding, it is sensible to expand on this way of modelling instrumentally rational behaviour before we discuss some of its difficulties.

#### *Ordinal utilities, cardinal utilities and expected utilities*

Suppose a person is confronted by a choice between driving to work or catching the train (and they both cost the same). Driving means less waiting

in queues and greater privacy while catching the train allows one to read while on the move and is quicker. Economists assume we have a preference ordering: each one of us, perhaps after spending some time thinking about the dilemma, will rank the two possibilities (in case of indifference an equal ranking is given). The metaphor of utility maximisation then works in the following way. Suppose you prefer driving to catching the train and so choose to drive. We could say equivalently that you derive $X$ utils from driving and $Y$ from travelling on the train and you choose driving because this maximises the utils generated, as $X > Y$.

## Box 1.1

### UTILITY MAXIMISATION AND CONSISTENT CHOICE

Suppose that a person is choosing between different possible alternatives which we label $x_1$, $x_2$, etc. We shall use the following notation to describe the preferences which inform these choices: $x_1 \; x_2$ means that the person 'prefers $x_1$ to $x_2$ **or** is indifferent between them'; $x_1 \geq x_2$ means that he or she 'prefers $x_1$ to $x_2$'; and $x_1 = x_2$ means that he or she is 'indifferent between the two'. A person is deemed *instrumentally rational* if he or she has preferences which satisfy the following conditions:

(1) *Reflexivity*: For any $x_i$, $x_i \geq x_i$
(2) *Completeness*: For any $x_i$, $x_j$, either $x_i \geq x_j$ or $x_i \leq x_j$
(3) *Transitivity*: For any $x_i$, $x_j$, $x_k$, if $x_i \geq x_j$ **and** $x_j \geq x_k$, then $x_i \geq x_k$
(4) *Continuity*: For any $x_i$, $x_j$, $x_k$, if $x_i > x_j > x_k$, then there must exist some 'composite' of $x_i$ and $x_k$, say $y$, which gives the same amount of utility as $x_j$; that is, $y = x_j$ and our individual is indifferent between them.

In the definition of *continuity* above there are more than one way of interpreting the 'composite' alternative denoted by $y$. One is to think of $y$ as a basket containing bits of $x_i$ and bits of $x_j$. For example, if $x_i$ is '5 croissant', $x_j$ is '3 bagels' and $x_k$ is '10 bread rolls', then some combination of croissant and bread rolls (e.g. 2 croissant and 4 bread rolls) must be equally valued as the 3 bagels. Another interpretation of $y$ is probabilistic. Imagine that $y$ is a lottery which gives the individual $x_i$ with probability $p$ ($0 < p < 1$) and $x_k$ with probability $1 - p$. Then the continuity axiom says that there exists some probability $p$ (e.g. 0.3) such that this lottery (that is, alternative $y$) is valued by the individual exactly as much as $x_j$.

When axioms (1), (2) and (3) hold, then the individual has a well-defined preference ordering. When (4) also holds, this preference ordering can be represented by a utility function. (A utility function takes what the individual has, e.g. $x_i$, and translates it into a unique level of utility. Its mathematical representation in this case is $U(x_i)$.) Thus the individual who makes choices with a view to satisfying his or her preference ordering can be conceived as one who is maximising this utility function.

## Box 1.2

### REFLECTIONS ON INSTRUMENTAL RATIONALITY

Instrumental rationality is identified with the capacity to choose actions which best satisfy a person's objectives. Although there is a tradition of instrumental thinking which goes back to the pre-Socratic philosophers, it is David Hume's *Treatise on Human Nature* which provides the clearest philosophical source. He argued that 'passions' motivate a person to act and 'reason' is their servant.

> We speak not strictly and philosophically when we talk of the combat of passion and reason. Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.

Thus reason does not judge or attempt to modify our 'passions', as some might think. This, of course, does not mean that our 'passions' might not be 'good', 'bad', 'wishy-washy' when judged by some light or other. The point is that it is not the role of reason to form such judgements. Reason on this account merely guides action by selecting the best way to satisfy our 'passions'.

This hypothesis has been extremely influential in the social sciences. For instance, the mainstream, neoclassical school of economics has accepted this Humean view with some modification. They have substituted preferences for passions and they have required that these preferences should be consistent. This, in turn, yields a very precise interpretation for how instrumental reason goes to work. It is as if we had various desires or passions which when satisfied yield something in common; call it 'utility'. Thus the fact that different actions are liable to satisfy our different desires in varying degrees (for instance, eating some beans will assuage our desire for nourishment while listening to music will satisfy a desire for entertainment) presents no special problem for instrumental reason. Each action yields the same currency of pleasure ('utils') and so we can decide which action best satisfies our desires by seeing which generates the most 'utility' (see Box 1.1 on consistent choice).

This maximising, calculative view of instrumental reason is common in economics, but it needs careful handling because it is liable to suggest an unwarranted connection with the social philosophy of 'utilitarianism' as presented by Jeremy Bentham and later John Stewart Mill (especially since J.S. Mill is a key figure associated with both the beginnings of neoclassical economics and the social philosophy of utilitarianism). The key difference is that Bentham's social philosophy envisioned a universal currency of happiness for all people. Everything in people's lives either adds to the sum total of utility in society (i.e. it is pleasurable) or subtracts from it (i.e. is painful) and the good society is the one that maximises the sum of those utilities, or average utility (see also Box 4.5 in Chapter 4). This was a radical view at the time because it broke with the tradition of using some external authority (God, the Church, the Monarch) to judge social