

Ŧ	T T	V	C	C E	· C	AN	n '	rы	E	C I	D	T	NI	C
ι	L	. Y	2	S E	٠,	$\mathbf{A} \mathbf{N}$	ע	IН	Ŀ	21	ĸ	Ľ	IN	3

Studies in rationality and irrationality





# ULYSSES AND THE SIRENS

## Studies in rationality and irrationality

Revised edition

JON ELSTER

Associate Professor, Department of History University of Oslo



EDITIONS DE LA MAISON DES SCIENCES DE L'HOMME

Paris



> CAMBRIDGE UNIVERSITY PRESS Cambridge, New York, Melbourne, Madrid, Cape Town, Singapore, São Paulo, Delhi, Tokyo, Mexico City

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/9780521269841

© Maison des Sciences de l'Homme Cambridge University Press 1979, 1984

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 1979 Revised edition 1984 Reprinted 1986, 1988, 1990, 1993, 1996 Re-issued 2013

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

ISBN 978-0-521-22388-1 Hardback ISBN 978-0-521-26984-1 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Information regarding prices, travel timetables, and other factual information given in this work is correct at the time of first printing but Cambridge University Press does not guarantee the accuracy of such information thereafter.



### CONTENTS

Preface to the revised edition Preface and acknowledgements				
I	PERFECT RATIONALITY: BEYOND GRADIENT-			
	CLIMBING	1		
1	Introduction	1		
2	The locally maximizing machine	4		
3	The globally maximizing machine	9		
4	Strategic behaviour in animals and men	18		
5	Functionalist explanation in sociology	28		
ΙΙ	IMPERFECT RATIONALITY: ULYSSES AND			
	THE SIRENS	36		
1	Introduction	36		
2	Towards a definition	37		
3	Pascal	47		
4	Descartes	54		
5	Inconsistent time preferences	65		
6	Endogenous change of preferences	77		
7	Precommitment in animal behaviour	86		
8	Abdication from power	88		
9	Some conclusions and further questions	103		
111	PROBLEMATIC RATIONALITY: SOME			
	UNRESOLVED PROBLEMS IN THE THEORY OF			
	RATIONAL BEHAVIOUR	112		
1	Introduction	112		



vi	Contents	
2	Games without solutions	117
3	Lexicographic preferences	124
4	Subjective probability	128
5	Maximizing, satisficing and natural selection	133
6	Traditional behaviour and random behaviour	137
7	Explaining altruism	141
8	Inconstancy	147
9	Paradox	150
10	And so what?	153
IV	IRRATIONALITY: CONTRADICTIONS OF THE	
	MIND	157
1	Introduction	157
2	Hate	161
3	Love	165
4	Self-deception	172
Ref	erences	180
Ind	l'ex	191



#### PREFACE TO THE REVISED EDITION

The only (but major) change in the present edition is that most of chapter II.5 on inconsistent time preferences has been completely rewritten. My friend Aanund Hylland of the Economics Department of the University of Oslo pointed out a gross mathematical error in the English edition, thus leaving what I thought to be a profound philosophical conclusion without a leg to stand on. Moreover, in a note dated 26 October 1982 he provided a detailed analysis of the correct mathematical structure of the problem. The results are given below. My conclusions have been reformulated accordingly. I should add that as usual his contribution goes much beyond the technical aspect of the problem; in fact the main conceptual arguments below are also due to his suggestions.

I would like to point out to the reader an article by Rebecca Dresser, 'Ulysses and the psychiatrists: A legal and policy analysis of the voluntary commitment contract', *Harvard Civil Rights-Civil Liberties Review* 16 (1982), pp. 777–854. The article cites my brief discussion of this issue (in II.2 below) and then goes on to provide a rich empirical material. Also, Thomas Schelling has discussed the problem of precommitment and strategic behaviour toward self in several recent articles, notably 'The intimate contest for self-command', *The Public Interest* 60 (1980), pp. 94–118.

J. E.



#### PREFACE AND ACKNOWLEDGEMENTS

Many philosophers and social scientists at some time in their lives have wanted to write fiction or poetry, only to find that they didn't have what it takes. Others have chosen philosophy or social science as a second choice when they decided that their first choice of doing mathematics really was not within their abilities. The present work is at the intersection of these two failures. But to fail is always to fail at something, and it leaves you with a knowledge of the kind of thing you unsuccessfully tried to do. In the essays collected here I have tried to exploit this knowledge for an analysis of rational and irrational behaviour.

The essays were written independently of each other, but have been rewritten to avoid redundancies and to incorporate further reflection. It may be useful to explain here how their topics relate to each other. Chapter I sets out the paradigm of individually rational behaviour, which is distinguished both from biological adaptation and from functional adaptation in societies. The main idea defended here is that the specifically human rationality is characterized by the capacity to relate to the future, in contradistinction to the myopic gradient-climbing in natural selection. Chapter II then introduces the notion of imperfectly rational behaviour, the need for which arises because weakness of will may prevent us from using our capacity for perfectly rational behaviour. The notion of binding oneself, as did Ulysses before setting out towards the Sirens, is the crucial concept of the chapter, though the alternative strategy of 'private side bets' is also explored. Chapter III is essentially a list of problems in rationalactor theory, with a view to evaluating the power of this theory compared to norm-oriented or structuralist approaches. I conclude that the rational-actor theory is logically prior to its

viii



#### Preface and acknowledgements

ix

competitors, though not necessarily more successful in each particular case. In chapter IV some of the problems discussed in chapter III are singled out for more intensive discussion. In particular I try to explain how contradictory beliefs and contradictory desires can be understood as meaningful even if irrational. There is, in other words, a descending sequence of perfect rationality, imperfect rationality, problematic rationality and irrationality which, in spite of the very diverse material included, lends a conceptual unity to the essays that justifies their being collected in book form.

Underlying all the essays is a particular view of the philosophy of science which I hope to be able to set out more fully elsewhere. A brief outline may prepare the reader for some of the ideas explored below.

- (i) There are basically three modes of explanation in science: the causal, the functional and the intentional.
  - (ii) All sciences use causal explanation.
- (iii) The physical sciences use only causal explanation, least-time principles and other variational formulations being merely analytical artifacts without explanatory power.
- (iv) There is no place for intentional explanation in biology. This statement is defended in chapter I below.
- (v) There is no place for functional explanation in the social sciences. This statement is defended (and qualified) in I.5 and II.8.
- (vi) In biology a distinction can be made between sub-functional causality (mutations, senescence) and supra-functional causality (beneficial or harmful spill-over effects of individual adaptations). This distinction is briefly touched upon in chapter I.
- (vii) In the social sciences a similar distinction can be made between sub-intentional causality and supra-intentional causality. The former refers to causal processes taking place within the individual, forming or perverting his intentions. This is the subject of much of chapters II and III. The latter refers to causal interaction between individuals. In my Logic and Society, which is in a sense a twin volume to the present book, I discuss this subject at some length.
- (viii) Animal and human behaviour should be studied with the notions of function and of intention as regulative ideas. Not all <sup>1</sup> Elster (1978a).



#### x Preface and acknowledgements

animal behaviour is functional, and not all human behaviour is rational or intentional, but there is a well-grounded *presumption* that this will typically be the case.

Chapter I was originally presented at the Fourth International Congress of the International Organization for the Study of Human Development, Paris, 1977. The present version has benefited from the comments of Roger Masters, Arthur Stinchcombe and George Williams. A very much shorter version of chapter II was first presented at the ECPR Workshop on Political Theory, Louvain, 1976. Finn Tschudi then helped me by pointing out the closely related work of George Ainslie, from whom I later received comments and access to unpublished manuscripts that proved very important for the development of my ideas. I also would like to thank Francis Sejersted, Sissel Reichelt, Dagfinn Føllesdal, John Perry, Michael Bratman, Amélie Rorty, Peter Hammond, Arthur Stinchcombe and Robert Goodin for criticism and advice. Chapter III was presented at the Séminaire International sur l'Economie Sociologique, Paris, 1977. I would like to thank John Harsanyi for stimulating discussion during the gestation period of the paper, and Robert Goodin for constructive criticism. Chapter IV partly overlaps with chapter 4 of my Logic and Society, but there are large differences both in the material itself and in the way in which it is organized. Among those I am in debt to here are Amélie Rorty, Eugene Genovese and Paul Watzlawick, the first for important suggestions and the latter two for confirming that I had indeed understood them aright.