

This book is published as part of the joint publishing agreement established in 1977 between the Fondation de la Maison des sciences de l'homme and the Press Syndicate of the University of Cambridge. Titles published under this arrangement may appear in any European language or, in the case of volumes of collected essays, in several languages.

New books will appear either as individual titles or in one of the series which the Maison des sciences de l'homme and the Cambridge University Press have jointly agreed to publish. All books published jointly by the Maison des sciences de l'homme and the Cambridge University Press will be distributed by the Press throughout the world.

Cet ouvrage est publié dans le cadre de l'accord de co-édition passé en 1977 entre la Fondation de la Maison des sciences de l'homme et le Press Syndicate of the University of Cambridge. Toutes les langues européennes sont admises pour les titres couverts par cet accord, et les ouvrages collectifs peuvent paraître en plusieurs langues.

Les ouvrages paraissent soit isolément, soit dans l'une des séries que la Maison des sciences de l'homme et Cambridge University Press ont convenu de publier ensemble. La distribution dans le monde entier des titres ainsi publiés conjointement par les deux établissements est assurée par Cambridge University Press.

SOUR GRAPES

Studies in the subversion of rationality

JON ELSTER

*Associate Professor, Department of History
University of Oslo*

CAMBRIDGE UNIVERSITY PRESS

Cambridge

London New York Melbourne

EDITIONS DE LA MAISON DES SCIENCES DE
L'HOMME

Paris

Published by the Press Syndicate of the University of Cambridge
The Pitt Building, Trumpington Street, Cambridge CB2 1RP
32 East 57th Street, New York, NY 10022, USA
296 Beaconsfield Parade, Middle Park, Melbourne 3206, Australia
and Editions de la Maison des Sciences de l'Homme
54 Boulevard Raspail, 75270 Paris Cedex 06

© Maison des Sciences de l'Homme and Cambridge University Press 1983

First published 1983

Printed in Great Britain at the University Press, Cambridge

Library of Congress catalogue card number: 82-22034

British Library cataloguing in publication data

Elster, Jon
Sour grapes.
I. Rationalism
I. Title
153.4'3 BF441

ISBN 0 521 25230 X hard covers
ISBN 2 7351 0031 6 (France only)

WT

B
824.2
E45
1983

CONTENTS

<i>Preface and Acknowledgements</i>	vii
I RATIONALITY	1
1 Introduction	1
2 Individual rationality: the thin theory	2
3 Individual rationality: the broad theory	15
4 Collective rationality: the thin theory	26
5 Collective rationality: the broad theory	33
II STATES THAT ARE ESSENTIALLY BY-PRODUCTS	43
1 Introduction	43
2 Willing what cannot be willed	44
3 Technologies for self-management	53
4 Commands	60
5 Trying to impress	66
6 Faking	71
7 Choice and intention in art	77
8 The impotence of power	86
9 Self-defeating political theories	91
10 The obsessional search for meaning	101
III SOUR GRAPES	109
1 Introduction	109
2 A conceptual map	111
3 Power, freedom and welfare	125
4 Sour grapes and social choice	133

211101-1

stress that even if actions may sometimes be explained as *attempts* to maximize utility in this *ex ante* sense, we would not be justified in thinking that the attempt would succeed; rather the contrary.¹⁹ On the other hand, as I observed earlier, when the utility-maximizing consequences of behaviour *can* be invoked to explain it, they do so by providing a causal explanation of the preferences. Pleasurable inner states enter importantly into the explanation of behaviour, but not as the conscious goal of behaviour.

Secondly, we may usefully contrast *rational man* with *economic man*. The first involves – in the thin sense which we are discussing now – nothing but consistent preferences and (to anticipate) consistent plans. The second is a much better-endowed creature, with preferences that are not only consistent, but also complete, continuous and *selfish*. To be sure, economists have constructed a large variety of models involving non-selfish preferences,²⁰ but their reflex is nevertheless to attempt to derive all apparently non-selfish behaviour from selfish preferences.²¹ This may perhaps be a good research strategy: when setting out to explain a given piece of behaviour, assume first that it is selfish; if not, then at least rational; if not, then at least intentional. But there can be no way of justifying the substantive assumption that all forms of altruism, solidarity and sacrifice really are ultra-subtle forms of self-interest, except by the trivializing gambit of arguing that people have concern for others because they want to avoid being distressed by their distress. And even this gambit, as Allan Gibbard has pointed out, is open to the objection that rational distress-minimizers could often use more efficient means than helping others.²²

¹⁹ As emphasized in van Parijs (1981) and Elster (1982a), one should distinguish between explanation in terms of intended and in terms of actual consequences of behaviour, although there is of course no general presumption that the intended consequences will fail to materialize – except for the class of cases that form the subject of Ch. II below.

²⁰ See the useful survey and discussion in Kolm (1981a).

²¹ See in particular the important synthesis of biological and game-theoretic considerations in Axelrod and Hamilton (1981). They use a model of sequential Prisoner's Dilemmas to show (i) that genuinely altruistic motivation can arise out of natural selection by purely selfish criteria and (ii) that some cases of apparently altruistic motivation can be explained by assuming no more than selfish rationality. In other words, if people behave altruistically, it is either because they have been programmed to feel concern for others or because they have calculated that it pays to fake concern for others. The first explanation, while in a sense reductionist, allows rational resistance to the economic reductionism embodied in the second. Yet there probably are cases that are resistant also to biological reductionism, unless one postulates that fitness-reducing altruism can be explained by the fact that 'it is not worth burdening the germ plasm with the information necessary to realize such an adjustment' (Williams 1966, p. 206).

²² Gibbard (forthcoming).

I now turn to plans and their consistency criteria. To ask for such criteria is to presuppose that an action can be intentional and yet not be rational. I do indeed claim that there are such actions, and much of Ch. II deals with an important subclass of them. Before I go on to discuss the consistency criteria for plans, let me distinguish this claim from a different one, that a person could act intentionally and yet be on the whole irrational. Once again we must agree with Davidson that global rationality is a precondition for imputing intentions to a person, be they irrational. We must be able to make sense of a person on the whole, if we are to be able to say that some of his plans do not make sense.

A rational plan must fulfil two criteria. First, the end state in terms of which it is defined must be a logically coherent one. If it is true, as claimed by Sartre, that we all fundamentally want to be simultaneously *en-soi* and *pour-soi*, resting in ourselves like a thing and yet at a distance from ourselves that allows us to enjoy this, then we are indeed striving for an end that is logically or conceptually incoherent. Acting on this desire will be as self-defeating as the attempt to turn around, very swiftly, to catch one's own shadow. Similarly, the desire for a unilateral recognition – to be recognized by another whom you do not yourself recognize – is a desire to bring about a state that could not possibly obtain, since a condition for recognition is that it is reciprocal.²³ So, on this first criterion, a necessary condition for the consistency of a plan is that there should be a possible world in which it is realized.

As in the case of beliefs, however, we need a second criterion: there must be a possible world in which the plan is realized deliberately, i.e. in which one finds both the plan and its fulfilment. Take the plan to behave spontaneously. There is nothing incoherent in the end state which defines that plan, since people often do behave spontaneously. Yet trying to be spontaneous is a self-defeating plan, since the very act of trying will interfere with the goal. There is a possible world in which I behave spontaneously, but none in which I plan to do so and succeed. Plans that violate the first of these criteria are logically or conceptually contradictory; those that violate the second but not the first are pragmatically contradictory. In Ch. II below I am concerned almost exclusively with the latter kind of contradictory plans.

I conclude this section with some remarks on the ambiguity in the notion of rational behaviour, related to the distinction between an

²³ See also Elster (1976; 1978a, pp. 70ff.) for a further analysis of the sense in which unilateral recognition is an incoherent idea.