# RATIONAL CHOICE THEORY

## CRITICAL CONCEPTS IN THE SOCIAL SCIENCES

Edited by
MICHAEL ALLINGHAM

# RATIONAL CHOICE THEORY

## Critical Concepts in the Social Sciences

*Edited by*
*Michael Allingham*

**Volume V**

Routledge
Taylor & Francis Group
LONDON AND NEW YORK

First published 2006
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon, OX14 4RN

Simultaneously published in the USA and Canada
by Routledge
270 Madison Avenue, New York, NY 10016

*Routledge is an imprint of the Taylor & Francis Group, an informa business*



Typeset in 10/12pt Times by Graphicraft Ltd., Hong Kong
Printed and bound in Great Britain by
MPG Books Ltd., Bodmin, Cornwall



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

*Library of Congress Cataloging in Publication Data*
A catalog record for this book has been requested

ISBN10: 0-415-35751-9 (Set)
ISBN10: 0-415-35756-X (Volume V)
ISBN13: 978-0-415-35751-7 (Set)
ISBN13: 978-0-415-35756-2 (Volume V)

**Publisher's Note**

References within each chapter are as they appear in the original complete work.

# ACKNOWLEDGEMENTS

The publishers would like to thank the following for permission to reprint their material:

MIT Press for permission to reprint H. A. Simon, 'A behavioral model of rational choice', *Quarterly Journal of Economics*, 69, 1955, pp. 99–118.

The American Economic Association and William Baumol for permission to reprint W. J. Baumol and R. E. Quandt, 'Rules of thumb and optimally imperfect decisions', *American Economic Review*, 54, 1964, pp. 23–46.

Springer for permission to reprint James G. March, 'Bounded rationality, ambiguity, and the engineering of choice', *Bell Journal of Economics*, 9, 1978, pp. 587–608.

The American Economic Association for permission to reprint Herbert A. Simon, 'Rationality as process and as product of thought', *American Economic Review*, 68, 1978, pp. 1–16.

The University of Chicago Press for permission to reprint Howard Margolis, 'A new model of rational choice', *Ethics*, 91, 1981, pp. 265–79. Copyright © 1981 by the University of Chicago.

R. Selten, 'Bounded rationality', *Journal of Institutional and Theoretical Economics*, 146, 1990, pp. 649–58.

The Econometric Society for permission to reprint Barton L. Lipman, 'How to decide how to decide how to . . . : modeling limited rationality', *Econometrica*, 59, 1991, pp. 1105–25.

Elsevier for permission to reprint Seppo Honkapohja, 'Adaptive learning and bounded rationality: an introduction to basic concepts', reprinted from *European Economic Review*, Vol. 37, 1993, pp. 587–94. Copyright © 1991, with permission from Elsevier.

The American Economic Association and W. Brian Arthur for permission to reprint W. Brian Arthur, 'Inductive reasoning and bounded rationality', *American Economic Review*, 84, 1994, pp. 406–11.

Blackwell Publishing for permission to reprint Barton L. Lipman, 'Information processing and bounded rationality: a survey', *Canadian Journal of Economics*, 28, 1995, pp. 42–67.

The American Economic Association and John Conlisk for permission to reprint J. Conlisk, 'Why bounded rationality?', *Journal of Economic Literature*, 34, 1996, pp. 669–700.

Elsevier for permission to reprint Larry Samuelson, 'Bounded rationality and game theory', reprinted from *Quarterly Review of Economics and Finance*, 36, 1996, pp. 17–35. Copyright © 1996, with permission from Elsevier.

Elsevier for permission to reprint Robert J. Aumann, 'Rationality and bounded rationality', reprinted from *Games and Economic Behavior*, 21, 1997, pp. 2–14. Copyright © 1997, with permission from Elsevier.

Blackwell Publishing for permission to reprint Hitoshi Matsushima, 'Bounded rationality in economics: a game theorist's view', *Japanese Economic Review*, 48, 1997, pp. 293–306.

The American Economic Association for permission to reprint R. H. Thaler, 'From Homo Economicus to Homo Sapiens', *Journal of Economic Perspectives*, 14, 2000, pp. 133–41.

D. Dequech, 'Bounded rationality, institutions, and uncertainty', *Journal of Economic Issues*, 35, 2001, pp. 911–29. Reprinted by special permission of the copyright holder, the Association for Evolutionary Economics.

The American Economic Association and Daniel Kahneman for permission to reprint D. Kahneman, 'Maps of bounded rationality: psychology for behavioral economics', *American Economic Review*, 93, 2003, pp. 1449–75.

## Disclaimer

# CONTENTS

# 55

# A BEHAVIORAL MODEL OF RATIONAL CHOICE

*Herbert A. Simon*

Source: *Quarterly Journal of Economics* 69 (1955): 99–118.

Traditional economic theory postulates an "economic man," who, in the course of being "economic" is also "rational." This man is assumed to have knowledge of the relevant aspects of his environment which, if not absolutely complete, is at least impressively clear and voluminous. He is assumed also to have a well-organized and stable system of preferences, and a skill in computation that enables him to calculate, for the alternative courses of action that are available to him, which of these will permit him to reach the highest attainable point on his preference scale.

Recent developments in economics, and particularly in the theory of the business firm, have raised great doubts as to whether this schematized model of economic man provides a suitable foundation on which to erect a theory — whether it be a theory of how firms *do* behave, or of how they "should" rationally behave. It is not the purpose of this paper to discuss these doubts, or to determine whether they are justified. Rather, I shall assume that the concept of "economic man" (and, I might add, of his brother "administrative man") is in need of fairly drastic revision, and shall put forth some suggestions as to the direction the revision might take.

Broadly stated, the task is to replace the global rationality of economic man with a kind of rational behavior that is compatible with the access to information and the computational capacities that are actually possessed by organisms, including man, in the kinds of environments in which such organisms exist. One is tempted to turn to the literature of psychology for the answer. Psychologists have certainly been concerned with rational behavior, particularly in their interest in learning phenomena. But the distance is so great between our present psychological knowledge of the learning and choice processes and the kinds of knowledge needed for economic and administrative theory that a marking stone placed halfway between might help travellers from both directions to keep to their courses.

Lacking the kinds of empirical knowledge of the decisional processes that will be required for a definitive theory, the hard facts of the actual world can, at the present stage, enter the theory only in a relatively unsystematic and unrigorous way. But none of us is completely innocent of acquaintance with the gross characteristics of human choice, or of the broad features of the environment in which this choice takes place. I shall feel free to call on this common experience as a source of the hypotheses needed for the theory about the nature of man and his world.

The problem can be approached initially either by inquiring into the properties of the choosing organism, or by inquiring into the environment of choice. In this paper, I shall take the former approach. I propose, in a sequel, to deal with the characteristics of the environment and the interrelations of environment and organism.

The present paper, then, attempts to include explicitly some of the properties of the choosing organism as elements in defining what is meant by rational behavior in specific situations and in selecting a rational behavior in terms of such a definition. In part, this involves making more explicit what is already implicit in some of the recent work on the problem — that the state of information may as well be regarded as a characteristic of the decision-maker as a characteristic of his environment. In part, it involves some new considerations — in particular taking into account the simplifications the choosing organism may deliberately introduce into its model of the situation in order to bring the model within the range of its computing capacity.

## I. Some general features of rational choice

The "flavor" of various models of rational choice stems primarily from the specific kinds of assumptions that are introduced as to the "givens" or constraints within which rational adaptation must take place. Among the common constraints — which are not themselves the objects of rational calculation — are (1) the set of alternatives open to choice, (2) the relationships that determine the pay-offs ("satisfactions," "goal attainment") as a function of the alternative that is chosen, and (3) the preference-orderings among pay-offs. The selection of particular constraints and the rejection of others for incorporation in the model of rational behavior involves implicit assumptions as to what variables the rational organism "controls" — and hence can "optimize" as a means to rational adaptation — and what variables it must take as fixed. It also involves assumptions as to the character of the variables that are fixed. For example, by making different assumptions about the amount of information the organism has with respect to the relations between alternatives and pay-offs, optimization might involve selection of a certain maximum, of an expected value, or a minimax.

Another way of characterizing the givens and the behavior variables is to say that the latter refer to the organism itself, the former to its environment. But if we adopt this viewpoint, we must be prepared to accept the possibility that what we call "the environment" may lie, in part, within the skin of the biological organism. That is, some of the constraints that must be taken as givens in an optimization problem may be physiological and psychological limitations of the organism (biologically defined) itself. For example, the maximum speed at which an organism can move establishes a boundary on the set of its available behavior alternatives. Similarly, limits on computational capacity may be important constraints entering into the definition of rational choice under particular circumstances. We shall explore possible ways of formulating the process of rational choice in situations where we wish to take explicit account of the "internal" as well as the "external" constraints that define the problem of rationality for the organism.

Whether our interests lie in the normative or in the descriptive aspects of rational choice, the construction of models of this kind should prove instructive. Because of the psychological limits of the organism (particularly with respect to computational and predictive ability), actual human rationality-striving can at best be an extremely crude and simplified approximation to the kind of global rationality that is implied, for example, by game-theoretical models. While the approximations that organisms employ may not be the best — even at the levels of computational complexity they are able to handle — it is probable that a great deal can be learned about possible mechanisms from an examination of the schemes of approximation that are actually employed by human and other organisms.

In describing the proposed model, we shall begin with elements it has in common with the more global models, and then proceed to introduce simplifying assumptions and (what is the same thing) approximating procedures.

### *1.1 Primitive terms and definitions*

Models of rational behavior — both the global kinds usually constructed, and the more limited kinds to be discussed here — generally require some or all of the following elements:

1. A set of *behavior alternatives* (alternatives of choice or decision). In a mathematical model, these can be represented by a point set, $A$.
2. The subset of *behavior alternatives that the organism "considers"* or "perceives." That is, the organism may make its choice within a set of alternatives more limited than the whole range objectively available to

it. The "considered" subset can be represented by a point set $\mathring{A}$, with $\mathring{A}$ included in $A$ ($\mathring{A} \subset A$).

3 *The possible future states of affairs*, or outcomes of choice, represented by a point set, $S$. (For the moment it is not necessary to distinguish between actual and perceived outcomes.)

4 A "*pay-off*" function, representing the "value" or "utility" placed by the organism upon each of the possible outcomes of choice. The pay-off may be represented by a real function, $V(s)$ defined for all elements, $s$, of $S$. For many purposes there is needed only an ordering relation on pairs of elements of $S$ — i.e., a relation that states that $s_1$ is preferred to $s_2$ or vice versa — but to avoid unnecessary complications in the present discussion, we will assume that a cardinal utility, $V(s)$, has been defined.

5 *Information as to which outcomes in S will actually occur* if a particular alternative, $a$, in $A$ (or in $\mathring{A}$) is chosen. This information may be incomplete — that is, there may be more than one possible outcome, $s$, for each behavior alternative, $a$. We represent the information, then, by a mapping of each element, $a$, in $A$ upon a subset, $S_a$ — the set of outcomes that may ensue if $a$ is the chosen behavior alternative.

6 *Information as to the probability that a particular outcome will ensue* if a particular behavior alternative is chosen. This is a more precise kind of information than that postulated in (5), for it associates with each element, $s$, in the set $S_a$, a probability, $P_a(s)$ — the probability that $s$ will occur if $a$ is chosen. The probability $P_a(s)$ is a real, non-negative function with $\sum_{S_a} P_a(s) = 1$.

Attention is directed to the threefold distinction drawn by the definitions among the set of behavior alternatives, $A$, the set of outcomes or future states of affairs, $S$, and the pay-off, $V$. In the ordinary representation of a game, in reduced form, by its pay-off matrix, the set $S$ corresponds to the cells of the matrix, the set $A$ to the strategies of the first player, and the function $V$ to the values in the cells. The set $S_a$ is then the set of cells in the $a$th row. By keeping in mind this interpretation, the reader may compare the present formulation with "classical" game theory.

### *1.2 "Classical" concepts of rationality*

With these elements, we can define procedures of rational choice corresponding to the ordinary game-theoretical and probabilistic models.[1]

A. *Max-min Rule*. Assume that whatever alternative is chosen, the worst possible outcome will ensue — the smallest $V(s)$ for $s$ in $S_a$ will be realized. Then select that alternative, $a$, for which this worst pay-off is as large as possible.

$$\hat{V}(\hat{a}) = \min_{s \in S_{\hat{a}}} V(s) = \max_{a \in A} \min_{s \in S_a} V(s)$$

Instead of the maximum with respect to the set, $A$, of actual alternatives, we can substitute the maximum with respect to the set, $\mathring{A}$, of "considered" alternatives. The probability distribution of outcomes, (6) does not play any role in the max-min rule.

B. *Probabilistic Rule.* Maximize the expected value of $V(s)$ for the (assumed known) probability distribution, $P_a(s)$.

$$\hat{V}(\hat{a}) = \sum_{s \in S_{\hat{a}}} V(s) P_{\hat{a}}(s) = \max_{a \in A} \sum_{s \in S_a} V(s) P_a(s)$$

C. *Certainty Rule.* Given the information that each $a$ in $A$ (or in $\mathring{A}$) maps upon a specified $s_a$ in $S$, select the behavior alternative whose outcome has the largest pay-off.

$$\hat{V}(\hat{a}) = V(S_{\hat{a}}) = \max_{a \in A} V(S_a)$$

## II. The essential simplifications

If we examine closely the "classical" concepts of rationality outlined above, we see immediately what severe demands they make upon the choosing organism. The organism must be able to attach definite pay-offs (or at least a definite range of pay-offs) to each possible outcome. This, of course, involves also the ability to specify the exact nature of the outcomes — there is no room in the scheme for "unanticipated consequences." The pay-offs must be completely ordered — it must always be possible to specify, in a consistent way, that one outcome is better than, as good as, or worse than any other. And, if the certainty or probabilistic rules are employed, either the outcomes of particular alternatives must be known with certainty, or at least it must be possible to attach definite probabilities to outcomes.

My first empirical proposition is that there is a complete lack of evidence that, in actual human choice situations of any complexity, these computations can be, or are in fact, performed. The introspective evidence is certainly clear enough, but we cannot, of course, rule out the possibility that the unconscious is a better decision-maker than the conscious. Nevertheless, in the absence of evidence that the classical concepts do describe the decision-making process, it seems reasonable to examine the possibility that the actual process is quite different from the ones the rules describe.

Our procedure will be to introduce some modifications that appear (on the basis of casual empiricism) to correspond to observed behavior processes in humans, and that lead to substantial computational simplifications in the making of a choice. There is no implication that human beings use all of these modifications and simplifications all the time. Nor is this the place to attempt the formidable empirical task of determining the extent to which, and the circumstances under which humans actually employ these simplifications. The point is rather that these are procedures which appear often to be employed by human beings in complex choice situations to find an approximate model of manageable proportions.

### *2.1 "Simple" pay-off functions*

One route to simplification is to assume that $V(s)$ necessarily assumes one of two values, (1, 0), or of three values, (1, 0, −1), for all $s$ in $S$. Depending on the circumstances, we might want to interpret these values, as (a) (satisfactory or unsatisfactory), or (b) (win, draw or lose).

As an example of (b), let $S$ represent the possible positions in a chess game at White's 20th move. Then a (+1) position is one in which White possesses a strategy leading to a win whatever Black does. A (0) position is one in which White can enforce a draw, but not a win. A (−1) position is one in which Black can force a win.

As an example of (a) let $S$ represent possible prices for a house an individual is selling. He may regard $15,000 as an "acceptable" price, anything over this amount as "satisfactory," anything less as "unsatisfactory." In psychological theory we would fix the boundary at the "aspiration level"; in economic theory we would fix the boundary at the price which evokes indifference between selling and not selling (an opportunity cost concept).

The objection may be raised that, although $16,000 and $25,000 are both "very satisfactory" prices for the house, a rational individual would prefer to sell at the higher price, and hence, that the simple pay-off function is an inadequate representation of the choice situation. The objection may be answered in several different ways, each answer corresponding to a class of situations in which the simple function might be appropriate.

First, the individual may not be confronted simultaneously with a number of buyers offering to purchase the house at different prices, but may receive a sequence of offers, and may have to decide to accept or reject each one before he receives the next. (Or, more generally, he may receive a sequence of pairs or triplets or $n$-tuples of offers, and may have to decide whether to accept the highest of an $n$-tuple before the next $n$-tuple is received.) Then, if the elements $S$ correspond to $n$-tuples of offers, $V(s)$ would be 1 whenever the highest offer in the $n$-tuple exceeded the "acceptance price" the seller had determined upon at that time. We can then raise the further question of what would be a rational process for determining the acceptance price.[2]

Second, even if there were a more general pay-off function, $W(s)$, capable of assuming more than two different values, the simplified $V(s)$ might be

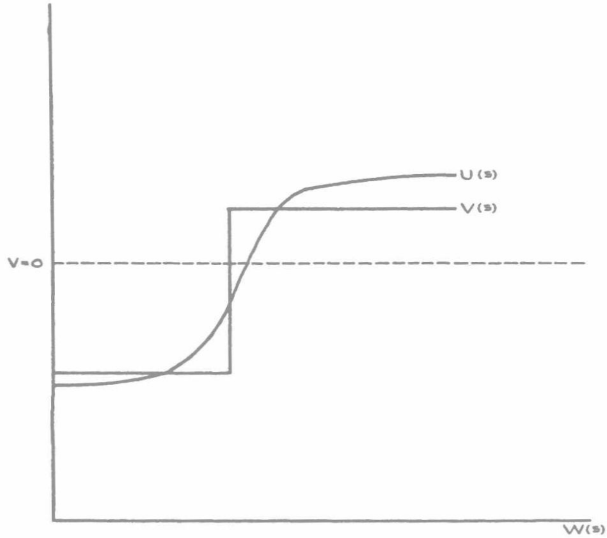


*Figure I*

a satisfactory approximation to $W(s)$. Suppose, for example, that there were some way of introducing a cardinal utility function, defined over $S$, say $U(s)$. Suppose further that $U(W)$ is a monotonic increasing function with a strongly negative second derivative (decreasing marginal utility). Then $V(s) = V\{W(s)\}$ might be the approximation as shown later.

When a simple $V(s)$, assuming only the values (+1, 0) is admissible, under the circumstances just discussed or under other circumstances, then a (fourth) rational decision-process could be defined as follows:

D. (i) Search for a set of possible outcomes (a subset, $S'$ in $S$) such that the pay-off is satisfactory ($V(s) = 1$) for all these possible outcomes (for all $s$ in $S'$).

(ii) Search for a behavior alternative (an $a$ in $\mathring{A}$) whose possible outcomes all are in $S'$ (such that $a$ maps upon a set, $S_a$, that is contained in $S'$).

If a behavior alternative can be found by this procedure, then a satisfactory outcome is assured. The procedure does not, of course, guarantee the existence or uniqueness of an $a$ with the desired properties.

### 2.2 *Information gathering*

One element of realism we may wish to introduce is that, while $V(s)$ may be known in advance, the mapping of $A$ on subsets of $S$ may not. In the extreme case, at the outset each element, $a$, may be mapped on the whole set, $S$. We may then introduce into the decision-making process information-gathering steps that produce a more precise mapping of the various elements of $A$ on nonidentical subsets of $S$. If the information-gathering process is not costless, then one element in the decision will be the determination of how far the mapping is to be refined.

Now in the case of the simple pay-off functions, (+1, 0), the information-gathering process can be streamlined in an important respect. First, we suppose that the individual has initially a very coarse mapping of $A$ on $S$. Second, he looks for an $S'$ in $S$ such that $V(s) = 1$ for $s$ in $S'$. Third, he gathers information to refine that part of the mapping of $A$ on $S$ in which elements of $S'$ are involved. Fourth, having refined the mapping, he looks for an $a$ that maps on to a subset of $S'$.

Under favorable circumstances, this procedure may require the individual to gather only a small amount of information — an insignificant part of the whole mapping of elements of $A$ on individual elements of $S$. If the search for an $a$ having the desirable properties is successful, he is certain that he cannot better his choice by securing additional information.[3]

It appears that the decision process just described is one of the important means employed by chess players to select a move in the middle and end game. Let $A$ be the set of moves available to White on his 20th move. Let $S$ be a set of positions that might be reached, say, by the 30th move. Let $S'$ be some subset of $S$ that consists of clearly "won" positions. From a very rough knowledge of the mapping of $A$ on $S$, White tentatively selects a move, $a$, that (if Black plays in a certain way) maps on $S'$. By then considering alternative replies for Black, White "explores" the whole mapping of $a$. His exploration may lead to points, $s$, that are not in $S'$, but which are now recognized also as winning positions. These can be adjoined to $S'$. On the other hand, a sequence may be discovered that permits Black to bring about a position that is clearly not "won" for White. Then White may reject the original point, $a$, and try another.

Whether this procedure leads to any essential simplification of the computation depends on certain empirical facts about the game. Clearly all positions can be categorized as "won," "lost," or "drawn" in an objective sense. But from the standpoint of the player, positions may be categorized as "clearly won," "clearly lost," "clearly drawn," "won or drawn," "drawn or lost," and so forth — depending on the adequacy of this mapping. If the "clearly won" positions represent a significant subset of the objectively "won" positions, then the combinatorics involved in seeing whether a position can be transformed into a clearly won position, for all possible replies by Black, may not be unmanageable.[4] The advantage of this procedure over the more common notion (which may, however, be applicable in the opening) of a general valuation function for positions, taking on values from −1 to 1, is that it implies much less complex and subtle evaluation criteria. All that is required is that the evaluation function be reasonably sensitive in detecting when a position in one of the three states — won, lost, or drawn — has been transformed into a position in another state. The player, instead of seeking for a "best" move, needs only to look for a "good" move.

We see that, by the introduction of a simple pay-off function and of a process for gradually improving the mapping of behavior alternatives upon possible outcomes, the process of reaching a rational decision may be drastically simplified from a computational standpoint. In the theory and practice of linear programming, the distinction is commonly drawn between computations to determine the *feasibility* of a program, and computations to discover the *optimal* program. Feasibility testing consists in determining whether a program satisfies certain linear inequalities that are given at the outset. For example, a mobilization plan may take as given the maximum work force and the steel-making capacity of the economy. Then a feasible program is one that does not require a work force or steel-making facilities exceeding the given limits.

An optimal program is that one of the feasible programs which maximizes a given pay-off function. If, instead of requiring that the pay-off be maximized, we require only that the pay-off exceed some given amount, then we can find a program that satisfies this requirement by the usual methods of feasibility testing. The pay-off requirement is represented simply by an additional linear inequality that must be satisfied. Once this requirement is met, it is not necessary to determine whether there exists an alternative plan with a still higher pay-off.

For all practical purposes, this procedure may represent a sufficient approach to optimization, provided the minimum required pay-off can be set "reasonably." In later sections of this paper we will discuss how this might be done, and we shall show also how the scheme can be extended to vector pay-off functions with multiple components (Optimization requires, of course, a complete ordering of pay-offs).

### *2.3 Partial ordering of pay-offs*

The classical theory does not tolerate the incomparability of oranges and apples. It requires a scalar pay-off function, that is, a complete ordering of pay-offs. Instead of a scalar pay-off function, $V(s)$, we might have a vector function, $V(s)$; where V has the components $V_1$, $V_2$, . . . A vector pay-off function may be introduced to handle a number of situations:

1 In the case of a decision to be made by a *group of persons*, components may represent the pay-off functions of the individual members of the group. What is preferred by one may not be preferred by the others.

2 In the case of an individual, he may be trying to implement a number of *values that do not have a common denominator* — e.g., he compares two jobs in terms of salary, climate, pleasantness of work, prestige, etc.;

3 Where each behavior alternative, *a*, maps on a set of *n possible consequences*, $S_a$, we may replace the model by one in which each alternative maps on a single consequence, but each consequence has as its pay-off the *n*-dimensional vector whose components are the pay-offs of the elements of $S_a$.

This representation exhibits a striking similarity among these three important cases where the traditional maximizing model breaks down for lack of a complete ordering of the pay-offs. The first case has never been satisfactorily treated — the theory of the *n*-person game is the most ambitious attempt to deal with it, and the so-called "weak welfare principles" of economic theory are attempts to avoid it. The second case is usually handled by superimposing a complete ordering on the points in the vector space ("indifference curves"). The third case has been handled by introducing probabilities as weights for summing the vector components, or by using principles like minimaxing satisfaction or regret.

An extension of the notion of a simplified pay-off function permits us to treat all three cases in much the same fashion. Suppose we regard a pay-off as *satisfactory* provided that $V_i \geq k_i$ for all *i*. Then a reasonable decision rule is the following:

E. Search for a subset $S'$ in $S$ such that $V(s)$ is satisfactory for all $s$ in $S'$ (i.e., $\underset{s \in S'}{V(s)} \geq k$).

Then search for an $a$ in $A$ such that $S_a$ lies in $S'$.

Again existence and uniqueness of solutions are not guaranteed. Rule E is illustrated in Figure II for the case of a 2-component pay-off vector.

In the first of the three cases mentioned above, the satisfactory pay-off corresponds to what I have called a *viable* solution in "A Formal Theory of the Employment Relation" and "A Comparison of Organization Theories."[5] In the second case, the components of V define the *aspiration levels* with respect to several components of pay-off. In the third case (in this case it is most plausible to assume that all the components of $k$ are equal), $k_i$ may be interpreted as the *minimum guaranteed pay-off* — also an aspiration level concept.

## III. Existence and uniqueness of solutions

Throughout our discussion we have admitted decision procedures that do not guarantee the existence or uniqueness of solutions. This was done in order to construct a model that parallels as nearly as possible the decision procedures that appear to be used by humans in complex decision-making settings. We now proceed to add supplementary rules to fill this gap.